Ancient horse genomes reveal the timing and extent of dispersals across the Bering Land Bridge

Alisa Vershinina, Peter Heintzman, Duane Froese, Grant Zazula, Molly Cassatt-johnstone, Love Dalén, Clio Der Sarkissian, Shelby Dunn, Luca Ermini, Cristina Gamba, et al.

Ancient horse genomes reveal the timing and extent of dispersals across the Bering Land Bridge

Alisa O. Vershinina[1], Peter D. Heintzman[2], Duane G. Froese[3], Grant Zazula[4,5], Molly Cassatt-Johnstone[1], Love Dalén[6,7], Clio Der Sarkissian[8], Shelby G. Dunn[1], Luca Ermini[9], Cristina Gamba[9], Pamela Groves[10], Joshua D. Kapp[1], Daniel H. Mann[10], Andaine Seguin-Orlando[8], John Southon[11], Mathias Stiller[1,12], Matthew J. Wooller[13,14], Gennady Baryshnikov[15], Dmitry Gimranov[16,17], Eric Scott[18], Elizabeth Hall[5], Susan Hewitson[5], Irina Kirillova[19], Pavel Kosintsev[16], Fedor Shidlovsky[20], Hao-Wen Tong[21,22], Mikhail P. Tiunov[23], Sergey Vartanyan[24], Ludovic Orlando[8], Russell Corbett-Detig[25], Ross D. MacPhee[26], Beth Shapiro[1,27]

[1]Ecology and Evolutionary Biology, University Of California Santa Cruz, Santa Cruz, United States

[2]The Arctic University Museum of Norway, UiT - The Arctic University of Norway, Tromsø, Norway

[3]Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, Canada

[4]Collections and Research, Canadian Museum of Nature, Station D, Ottawa, Canada

[5]Government of Yukon, Department of Tourism and Culture, Palaeontology Program, Whitehorse, Whitehorse, Yukon Territory Y1A 2C6 Canada

[6]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, 18 SE-10405 Stockholm, Sweden

[7]Centre for Palaeogenetics, Svante Arrhenius väg 20C, 10691 Stockholm, Sweden

[8]Centre d'Anthropobiologie et de Génomique de Toulouse UMR5288, Faculté de Médecine Purpan, Université Paul Sabatier, Toulouse, France

[9]Lundbeck Foundation GeoGenetics Center, University of Copenhagen, 1350K Copenhagen, Denmark

[10]Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, United States

[11]Keck-CCAMS Group, Earth System Science Department, University of California, Irvine, Irvine, United States

[12]Division Molecular Pathology, Institute of Pathology, University Hospital Leipzig, Leipzig, Germany

[13]Alaska Stable Isotope Facility, Water and Environmental Research Center, Institute of Northern Engineering, University of Alaska Fairbanks, Fairbanks, United States

[14]Department of Marine Biology, College of Fisheries and Ocean Sciences, University of Alaska Fairbanks, Fairbanks, United States

[15]Laboratory of Theriology, Zoological Institute of the Russian Academy of Sciences, St. Petersburg, Russia

[16]Institute of Plant & Animal Ecology of the Russian Academy of Sciences, Ural Branch, Ekaterinburg, Russia

[17]Ural Federal University named after the first President of Russia B. N. Yeltsin, Ekaterinburg, Russia

[18]California State University, San Bernardino, United States

[19]Institute of Geography, Russian Academy of Sciences, Staromonetny Lane, Bldg 29, Moscow, 119017 Russia

[20]National Alliance of Shidlovskiy "Ice Age", Moscow 129223, Russia

[21]Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, 100044 Beijing, China

[22]CAS Center for Excellence in Life and Paleoenvironment, Chinese Academy of Sciences, 100044 Beijing, China

[23]Federal Scientific Center of the East Asia Terrestrial Biodiversity, Far Eastern Branch of Russian Academy of Sciences, Vladivostok, Russia

[24]North-East Interdisciplinary Scientific Research Institute N.A. Shilo, Far East Branch, Russian Academy of Sciences, Magadan, Russia

[25]Biomolecular Engineering, University Of California Santa Cruz, Santa Cruz, United States

[26]American Museum of Natural History, New York, United States

[27]Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz CA 95064, USA

Keywords: *Equus ferus*, Bering Land Bridge, paleogenomics, population structure, horses

Corresponding authors: Alisa Vershinina (avershin@ucsc.edu), Beth Shapiro (bashapir@ucsc.edu)

Abstract

The Bering Land Bridge (BLB) last connected Eurasia and North America during the Late Pleistocene. Although the BLB would have enabled transfers of terrestrial biota in both directions, it also acted as an ecological filter whose permeability varied considerably over time. Here we explore the possible impacts of this ecological corridor on genetic diversity within, and connectivity among, populations of a once wide-ranging group, the caballine horses (*Equus* spp.). Using a panel of 187 mitochondrial and eight nuclear genomes recovered from present-day and extinct caballine horses sampled across the Holarctic, we found that Eurasian horse populations initially diverged from those in North America, their ancestral continent, around 1.0-0.8 million years ago. Subsequent to this split our mitochondrial DNA analysis identified two bi-directional long-range dispersals across the BLB ~875-625 and ~200-50 thousand years ago, during the Middle and Late Pleistocene. Whole genome analysis indicated low levels of gene flow between North American and Eurasian horse populations, which likely occurred as a result of these inferred dispersals. Nonetheless, mitochondrial and nuclear diversity of caballine horse populations retained strong phylogeographic structuring. Our results suggest that barriers to gene flow, currently unidentified but possibly related to habitat distribution across Beringia or ongoing evolutionary divergence, played an important role in shaping the early genetic history of caballine horses, including the ancestors of living horses within *Equus ferus*.

Introduction

The physiographic region of Beringia extends from the Lena River in Russia to the Mackenzie River in Canada and consists of both terrestrial and marine components (Figure 1a) (David M. Hopkins, 1959). Beringia's land area varied during the Pleistocene (2.6 million, Ma, to 11.7 thousand, ka, years ago) along with the extent of continental ice and its effect on sea level (Elias & Brigham-Grette, 2013; David M. Hopkins, 1973). Times of maximum subaerial exposure created a nearly continuous plain between Western (Asian) and Eastern (North American) Beringia, referred to as the Bering Land Bridge (BLB). At other times these regions were separated by the rejuvenated marine passage (Sher, 1999). Beringia therefore could have acted as either a conduit or a barrier for dispersing marine and terrestrial species (Alter et al., 2015; Guthrie, 1982; D. M. Hopkins, 1959; McKeon et al., 2016; Wooller et al., 2018; Zazula et al., 2014).

Barriers like the BLB can influence species' evolutionary trajectories by impeding or eliminating gene flow. Barriers may, for example, lead to allopatric speciation via genetic drift and local adaptation (Marques, Lucek, Sousa, Excoffier, & Seehausen, 2019; Schield et al., 2019), or create opportunities for evolutionary innovation via subsequent admixture after barriers are removed (Coyne & Orr, 1998; Nosil & Feder, 2012; Payseur & Rieseberg, 2016). The impact of barriers on species' evolution will, however, depend on their nature and permanency. In Beringia, the periodic inundation of the BLB created temporarily isolated continental populations of many wide-ranging terrestrial animals, including bison (*Bison* sp.), mammoths (*Mammuthus* sp.), and wolves (*Canis* sp.). Analyses of ancient mitochondrial DNA (mtDNA) from *Mammuthus, Bison,* and *Canis* have revealed strong phylogeographic contrasts between the continents, presumably a consequence of the barrier imposed by intermittent inundation of the BLB (Chang et al., 2017; Froese et al., 2017; Loog et al., 2019; Lorenzen et al., 2011). Consistent with this structuring, all

three mitochondrial data sets include evidence of multiple dispersals across the BLB. These data imply that gene flow may have regularly occurred between continental populations of different species at least periodically during the Pleistocene, a hypothesis amenable to further testing using nuclear genomic data sets.

The chronology of BLB availability as a terrestrial corridor for dispersal between Eurasia and North America remains poorly resolved. During the Pleistocene, ice volume and sea level estimates indicate that the BLB was largely exposed during Marine Isotope Stage (MIS) 6, ~190-130 ka; MIS 4, ~70-60 ka; and from middle MIS 3 to early MIS 1, ~45-11 ka (Hu et al., 2010; Jakobsson et al., 2017). The timing of the BLB exposure prior to MIS 6 is obscure (see (O'Regan et al., 2010)). However, exposure of the BLB does not necessarily imply that the region was viable as an ecological corridor for megafauna. Marine transgression would have disrupted the continuity of the Mammoth Steppe grassland and fragmented habitat available to large-bodied grazers (Guthrie, 2001; Lorenzen et al., 2011; Zimov, Zimov, Tikhonov, & Chapin, 2012). Habitat re-establishment following re-emergence of the BLB was likely impacted by variable sea levels, local climates, and regional topography, all of which varied across Beringia (Bond, 2019; Elias & Crocker, 2008; Knebel & Creager, 1973). The patchiness of habitat availability on the BLB would have facilitated some species' migrations while hampering others (Guthrie, 1982; Mann et al., 2015).

Reconstructions of first appearance dates (FADs) for different mammals dispersing across the BLB support the conclusion that optimal conditions for dispersal were intermittent and varied by taxon (Salis et al., 2020). The current North American FAD for mammoths, for example, is Early Pleistocene, ~1.5 Ma (Lister & Sher, 2015), while bison first entered North America during MIS 6, ~195-135 ka, and then again at ~45-21 ka (Froese et al., 2017). If humans crossed over the

BLB, as opposed to sailing along its southern coast, they may have done so at this later time as well (Moreno-Mayar et al., 2018). Moose (*Alces alces*) and wapiti (*Cervus elaphus*) expanded across the BLB into North America later, ~15 ka (Guthrie, 2006; Meiri, Lister, Kosintsev, Zazula, & Barnes, 2020; Meiri et al., 2014). By contrast, there is no evidence that the woolly rhinoceros (*Coelodonta antiquitatis*) dispersed into North America at any time (Guthrie, 2001; Stuart & Lister, 2012).

Successful dispersals in the opposite direction were fewer, but an outstanding case is that of caballine horses (within Quaternary equid diversity, the horses most closely related to the extant domestic horse, *Equus caballus*). Caballine horses diverged from their sister group in North America ~4.5-4.0 Ma (Orlando et al., 2013), and their fossil record suggests that they had already dispersed into Eurasia by ~0.9-0.8 Ma (Azzaroli, 1989; Forstén, 1991). This timing is in good agreement with the most recent molecular estimate of the Eurasian FAD for caballines based on nuclear genomes (~1.06 Ma; (Orlando et al., 2013)). Although the North American caballine horse population would eventually become extinct during the early Holocene (Haile et al., 2009), caballines on either side of the BLB became widespread after the initial dispersal (MacFadden, 2005) and, in the case of one or more Eurasian populations, were eventually domesticated (Fages et al., 2019; Gaunitz et al., 2018; Orlando, 2020; Schubert et al., 2014).

Despite their extensive fossil record, much less is known compared to other megafaunal species about how often and in which direction caballine horses dispersed across the BLB. Previous phylogenetic analyses of part of the mitochondrial control region suggest that caballine horses dispersed back into North America from Asia at least once during the Pleistocene (Barrón-Ortiz et al., 2017; Weinstock et al., 2005). If this happened, this could provide an excellent

opportunity to examine how alternating periods of isolation and connectivity might have affected the population biology of ancient horse populations on both sides of the BLB.

Here, we use palaeogenomic data from caballine horse remains collected in Eastern and Western Beringia as well as in temperate northeastern Asia and the mid-continental United States to explore the history of caballine horse dispersal across the BLB. We sequence and assemble two new caballine horse palaeogenomes from Yukon Territory, Canada, which represent the first high-coverage nuclear sequences for North American caballines, and 78 new mitochondrial genomes from caballine horses across Eurasia and North America. We then use these and previously published data (7 nuclear and 112 mitochondrial genomes) to explore the timing and nature of caballine horse dispersal across the BLB and the potential for bidirectional gene flow among horse populations across the transient corridor.

Methods

Sample collection, screening for DNA preservation, and dating

We collected samples of 262 horse bones and teeth from Eurasia (Ural Mountains and Siberia, Russia; eastern China) and North America (Yukon Territory, Canada; Alaska, USA; continental USA) (Figure 1a, Table S1). Most horse remains analysed in this study have been attributed as *Equus* sp.; however, some have been assigned on morphological grounds to a variety of likely nominal species, including *Equus ferus*, *E. lambei*, *E. uralensis,* and *E. przewalskii* (see Table S1). As species-level systematics of Holarctic equids is a subject of considerable debate (see (Barron-Ortiz et al., 2019) for discussion), we avoid official species designations here.

We processed specimens following standard protocols for working with degraded DNA (Fulton & Shapiro, 2019) in dedicated ancient DNA facilities at the University of California Santa

Cruz Paleogenomics Laboratory (PGL) or Center for GeoGenetics, University of Copenhagen, Denmark (CGG). For each specimen, we cleaned the surface to remove potential surface contaminants and then sampled ~1 g of bone powder for radiocarbon dating and genetic analyses (Table S1).

To extract DNA, we applied a sodium hypochlorite pretreatment to ~0.1g of bone powder to remove external DNA contamination (Korlević & Meyer, 2019) and then followed one of several extraction methods optimized for ancient DNA (Dabney & Meyer, 2019; Rohland et al., 2010). We used several library preparation protocols and indexing protocols, including single- and double-stranded DNA library preparation (Kapp, Green, & Shapiro, 2020; Meyer & Kircher, 2010) and single and dual-indexing strategies (Kircher, Sawyer, & Meyer, 2012). Sample-specific details are provided in Table S1. We sequenced each library on an Illumina MiSeq (2x75 cycles) or HiSeq (2x150 cycles) to a depth of 0.2-8.0 million reads, and screened samples by mapping recovered data to the EquCab2 reference genome as described below. Out of 262 samples, 78 had >1% endogenous DNA and a maximum of 1-30% clonality, and we selected these for genetic analyses.

We sent 71 of the 78 well-preserved samples (7 were dated previously) for Accelerator Mass Spectrometry (AMS) radiocarbon dating at either the Lawrence Livermore National Laboratory or the UC Irvine KECK AMS facilities (Appendix S1). All dates with the references for the published samples are provided in Tables 1 and S1.

Mitochondrial genome assembly

Twenty of the 78 well-preserved samples yielded >4-fold coverage mitochondrial genomes from the screening data (Table S1). For the remaining 58, we performed RNA-bait hybridization-based target enrichment using the horse “myBaits Expert Mito" kit (Daicel Arbor Biosciences,

previously MYcroarray) and following manufacturer's protocols v. 1 to v. 4, with hybridization at 65°C for 36 hours as described by Vershinina et al. (2019). We sequenced the enriched libraries on an Illumina MiSeq (2x75 cycles). To assemble mitochondrial genomes, we used a bash script (mito_assembly_pipeline.sh) available from https://github.com/Paleogenomics/DNA-Post-Processing/. This pipeline trims adapters, merges reads overlapping by ≥15 bp, discards reads shorter than 27 bp with SeqPrep2 (https://github.com/jeizenga/SeqPrep2), and removes reads with low sequence complexity using PRINSEQ-lite v0.20.4 (Schmieder & Edwards, 2011) with the flag 'dust=7'. Merged and unmerged fastq files are concatenated and converted to fasta with the FASTX-toolkit v0.0.13.2 (Gordon, Hannon, & Others, 2010). We then assembled mitochondrial genomes using Mapping Iterative Assembler v1.0 (MIA; https://github.com/mpieva/mapping-iterative-assembler) with the domestic horse mitochondrial genome (NC_001640; Xu & Arnason, 1994) as the seed sequence. From the MIA assembly, we called bases at sites covered by at least three independent reads and for which we observed >67% consensus among reads. The 78 new mitochondrial genomes resulting from this pipeline are available at GenBank with accessions MW846090-MW846167.

We augmented our data set with 110 previously published ancient and present-day horse mitogenomes from across their Holarctic Pleistocene range (Fages et al., 2019; Librado et al., 2015; Lippold, Matzke, Reissmann, & Hofreiter, 2011); see Table S2 for the complete list of references). Mitochondrial genomes from Fages et al. (2019) were published as BAM alignments. We converted these to fastq with bedtools v2.25.0 bamtofastq (Quinlan & Hall, 2010) and re-assembled mitochondrial genomes following the pipeline above.

We added the donkey (*E. asinus*, GenBank: NC_001788; (Xu, Gullberg, & Arnason, 1996)), Ovodov's horse and zebra (*E. ovodovi*: NC_018783.1; *E. zebra*: NC_020476.1; both Vilstrup et al. (2013)) mitochondrial genomes to our data set and aligned all mitogenomes using Muscle v3.8.425 (Edgar, 2004). We then manually inspected the alignment to correct misaligned gaps. Our final alignment comprised 190 equid mitochondrial genomes, including three outgroup sequences (donkey, zebra, Ovodov's horse) and 187 caballine horses sampled from across the Holarctic and ranging in age from early-Middle Pleistocene (*E.* cf. *scotti*, (Orlando et al., 2013)) to the present (Figures 1a,b,c; Tables S1, S2).

Mitochondrial phylogenetic analysis

We estimated mitochondrial phylogenies using three approaches. First, we used the Maximum Likelihood (ML) approach implemented in RAxML v8.2.4 (Stamatakis, 2014) and the Bayesian time-calibrated approach implemented in BEAST v1.10.1 (Suchard et al., 2018). In RAxML, we considered the alignment as one partition and ran three instances of RAxML assuming the GTRGAMMAI model of nucleotide substitution (Figure S1).

Next, we estimated the timing of divergence between caballine horse lineages using BEAST. We calibrated the molecular clock using median calibrated radiocarbon dates or mean stratigraphic age for each sample (Tables S1, S2), excluding samples with non-finite or no age estimates, and a prior at the root of the tree of 4.0–4.5 Ma (normal prior, mean 4.25 Ma, stdev: 0.15 Ma) based on (Orlando et al., 2013). We placed an additional normal prior on the age of the early-Middle Pleistocene Thistle Creek horse YG148.20/TC21 (GenBank: KT757763) with a mean of 670 ka and stdev of 56.4 ka following the proposed range of this sample's age (560–780 ka BP; Orlando et al., 2013). We assumed the GTR+G nucleotide model with 4 rate categories and

an uncorrelated log-normally distributed relaxed clock model with exponential prior 4.68E-8 (mean, substitutions/site/million years) (Schubert et al., 2014), and the Bayesian SkyGrid (Gill et al., 2013; Hill & Baele, 2019) coalescent prior with 150 groups. We ran two MCMC chains for 100 million iterations each, sampling parameters and trees every 10,000 states, discarded the first 25% as burn-in, combined the remainder using LogCombiner v1.10.4, and then analyzed parameter convergence in Tracer v1.7.1 (Rambaut, Drummond, Xie, Baele, & Suchard, 2018) (Table S4). We combined the posterior trees and calculated the maximum clade credibility (MCC) tree in TreeAnnotator v1.10.4 (Rambaut & Drummond, 2010) (Figure S2).

Third, we used the continuous-time Markov chain (CTMC) (Minin & Suchard, 2008a, 2008b) to infer the timing and number of migrations across the Bering Land Bridge, using only dated caballine horses (Figure 1b, c). It has been previously shown that placement of various outgroups, or the lack thereof, affects phylogenetic position of horses from Far Eastern Asia and New Siberian Islands (Fages et al., 2019; Heintzman et al., 2017; Yuan et al., 2020). The deeply-divergent clade likely forms a root polytomy, but appears to be more closely related to specimens from the rest of Eurasia. Through initial CTMC runs, we found that the polytomy problem was further exacerbated when the archaic Middle Pleistocene YG148.20/TC21 was included in the analysis. To account for the confounding effects of the polytomy on the unrooted CTMC BEAST run, we enforced two monophylies. One monophyly united samples from Far Eastern Asia and New Siberian Islands with the remaining Eurasian haplogroup horses. The second monophyly united samples with the North American haplogroup. We selected samples for both haplogroups based on our RAxML and BEAST analyses, which included non-caballine horses, and on previous studies (Yuan et al., 2020). We placed a normal prior of 850 ka (stdev: 200 ka) on the root height of the tree based on (Orlando et al., 2013).

For the CTMC analysis, we subdivided the mitochondrial alignment into four partitions: protein-coding genes (CDS), tRNA, rRNA, and control region (CR), and removed the variable portion of the CR spanning 16,167 to 16,475 bp that cannot be assembled accurately from short reads (e.g. Heintzman et al. 2017). Following results of jModeltest v2.1.10 (Darriba, Taboada, Doallo, & Posada, 2012), we assigned GTR+I+G model to the CDS, CR, and rRNA partitions and HKY+I+G to the tRNA partition, and assumed a strict molecular clock. We ran and analysed two MCMC chains and posterior trees with the same number of steps, burn-in and sampling parameters as in BEAST that included the outgroups as described above. We computed the density distribution of the geographic parameter with R v4.0.2 (R Core Team, 2014). We used R packages ggtree (Yu, Smith, Zhu, Guan, & Lam, 2017), treeio (Wang et al., 2020), and ggplot2 (Wickham, 2016) to visualize all phylogenetic trees presented in the current study (Figure 1b; Figures S1-S3). We report the log-combined trace file results for all BEAST runs in the Table S4.

Nuclear genome reconstruction, alignment, and filtering

We selected two particularly well-preserved samples from Yukon for whole genome shotgun sequencing: YG303.325 and YG188.42, both dated to the Late Pleistocene (Table 1). For YG303.325, we extracted DNA from two ~0.12 g aliquots of bone powder and converted these extracts into six double-stranded, single-indexed DNA libraries (Kircher et al., 2012; Meyer & Kircher, 2010) that we then sequenced on dedicated lanes of an Illumina HiSeq-X (2x150 cycles) at the SciLifeLab (Stockholm, Sweden) and the HiSeq-2500 (2x100 cycles) at the UC San Diego Institute for Genomic Medicine. For YG188.42 (YT03-40), we extracted DNA from three ~0.15 g aliquots of bone powder following the D1, D2, S1, S2, Y1, and Y2 methods (Gamba et al., 2016), and converted these extracts into nine double-stranded, single-indexed DNA libraries (Librado et

al., 2017) which we sequenced at the Danish National DNA Sequencing Centre on dedicated lanes of the Illumina HiSeq-2500 instrument (1x101 cycles).

For both samples, we removed adapter sequences, merged paired-end reads that overlapped by at least 10 nucleotides, and removed reads shorter than 30 bp using SeqPrep2. We mapped retained reads (both merged and unmerged) to the horse reference genome assembly EquCab2.0 (GenBank: GCA_000002305.1; (Wade et al., 2009), using the BWA v0.7.12 backtrack algorithm (Li & Durbin, 2009) with seed disabled as suggested by Schubert et al. (2012), converting SAM alignments into BAM alignments using BWA 'sampe' for unmerged and 'samse' for merged reads. We filtered out bases with a mapping quality <20 using samtools v0.1.19 (Li et al., 2009). We removed PCR-duplicated reads using samtools rmdup, and realigned reads around indels using the GATK v3.7 IndelRealigner (McKenna et al., 2010). We then explored patterns of DNA damage with mapDamage2 (Jónsson, Ginolhac, Schubert, Johnson, & Orlando, 2013) and used this program to rescale quality scores taking into account the damage estimates, by using the --rescale and --fix-nicks flags (Figures S4-S6). We calculated genome-wide sequencing and type-specific error rates following Orlando et al., 2013 (Appendix S2, Table S5, Figure S7).

We added our two new palaeogenomes to a data set of seven previously published equid genomes (Table 1): Batagai (Librado et al., 2015); CGG10022 and CGG10023 (both (Schubert et al., 2014)); Middle Pleistocene *E.* cf. *scotti* (TC21), present-day Przewalski's horse, Thoroughbred horse (Twilight), and *E. asinus* (donkey Willy; version mapped to horse) (all four from (Orlando et al., 2013)).

Evaluating population structure

We detected population structure in our horse autosomal data set using ANGSD v0.918 (Korneliussen, Albrechtsen, & Nielsen, 2014) and genotype likelihoods (GL) (Nielsen, Paul, Albrechtsen, & Song, 2011), which incorporate genotyping uncertainty that results from low coverage. We generated GLs for eight horses with ANGSD options -GL 2 -doGlf 2 -minQ 30 -minMapQ 30 -C 50 -rmTrans 1, and generated a GL covariance matrix with PCAngsd (Meisner & Albrechtsen, 2018). We used the R packages 'prcomp' and 'Factoextra' (Kassambara & Mundt, 2017) to run and visualize the principal component analysis (PCA) of the GL covariance matrix (Figure 2a).

Estimating admixture

We ran an admixture analysis using the '-admix' flag in PCAngsd for K=2 to K=5 (Figure 2b; Figure S8), and visualized the results with the R packages RcppCNPy (Eddelbuettel & Wu, 2016) and ggplot2 (Wickham, 2016). We next calculated D-statistics (ABBA/BABA test) using ANGSD to explore post-divergence admixture between sampled horse lineages (Durand, Patterson, Reich, & Slatkin, 2011; Green et al., 2010). We excluded the Middle Pleistocene TC21 genome from this and all subsequent analyses due to its low coverage and age, which is close to the divergence of the caballine horse populations ~1.0-0.8 Ma ago. We polarized alleles using the donkey genome and calculated the number of shared derived sites between one of the two closely related lineages (P1 and P2) and a possible introgressor (P3). As hybridization history can differ by sex (Cahill et al., 2015), we performed analyses separately for the autosomes and the X chromosome (Figure 2d; Figure S6). We used ANGSD '-doAbbababa 1' in 5 Mb chromosome blocks and computed the number of ABBA\BABA sites with ANGSD quality filters '-minQ 30 -

minMapQ 30' and '-trim 5' to account for DNA damage at the ends of the reads and sequencing error rate difference between the genomes. Using this set of filters we run D-statistic analysis on two datasets - with all genomic sites and with transition sites removed (Figure 2c,d). We considered the results to be significant if $|Z| > 3$ and moderately significant if $|Z| > 2$ (Barlow et al., 2018; Zheng & Janke, 2018) (Figure 2c,d; Figure S9).

To explore gene flow patterns further, we calculated the number of admixed non-overlapping windows in our autosomal dataset using D_{FOIL} v2017-011-25 (Pease & Hahn, 2015). The D_{FOIL} approach is similar to the D-statistic tests, but it expands the model to five-population symmetric tree topology which allowed us to test the directionality of gene flow. Following the phylogenetic position of the whole genome samples, we used Yukon horses as P1 and P2, Eurasian horses as P3 and P4, and donkey as an outgroup. We paired Eurasian samples only if they had close error rates to decrease the impact of false positive allele sharing (Appendix S2, Table S5, see Rasmussen et al., 2011). To run D_{FOIL}, we created pseudohaploid fasta files that excluded transitions to account for aDNA damage. To avoid reference bias, we called a random base at each site and filtered the data using the same criteria as the D-statistic analysis described above. We then split the genomes into 200 kb non-overlapping windows with bedtools makewindows tool. We converted each fasta file to D_{FOIL} counts and tested the treeness of a five-population model with native D_{FOIL} package scripts (Table S6, Figure S10).

Estimating demographic history with G-PhoCS and PSMC

To infer demographic history and connectivity between horse populations in Eurasia and North America, we used the Bayesian coalescent approach implemented in G-PhoCS v1.3 (Gronau, Hubisz, Gulko, Danko, & Siepel, 2011), which reconstructs the coalescent process of

multiple neutral loci assuming negligible recombination and selection. We excluded the low coverage CGG10023 and TC21 horses and converted the base quality score recalibrated BAM alignments of other genomes to FASTA format using ANGSD '-doFasta 4 -doCounts 1 -minQ 25 -minMapQ 25 -uniqueOnly -setMinDepth 5 -setMaxDepth 100' (Korneliussen et al., 2014). We used the flag '-doFasta 4' in ANGSD to incorporate IUPAC ambiguity codes for variable sites, which allows G-PhoCS to iterate over all possible diploid phases (Gronau et al., 2011). To ascertain putatively neutral autosomal loci, we excluded sites with low and extremely high coverage, sites with low base quality, sites adjacent to indels, sites in clusters of apparent SNPs or those in regions of the genome putatively undergoing selection as described below.

To gather coordinates of regions putatively affected by selection, we used annotations of the EquCab2 genome reference (Wade et al., 2009). We downloaded bed-files with exon data from Ensembl (Build 92) (Yates et al., 2020), tandem repeat tracks and genome assembly gaps from UCSC Genome Browser http://genome.ucsc.edu/ (Karolchik et al., 2004), duplicated genes track from the Duplicated Genes database (Ouedraogo et al., 2012), simple and tandem repeats from the RepeatMasker database (Smit, Hubley, & Green, 2013-2015), and coordinates of CpG islands from http://www.haowulab.org/software/makeCGI/ (Irizarry, Wu, & Feinberg, 2009; Wu, Caffo, Jaffee, Irizarry, & Feinberg, 2010).

To ascertain a high-quality set of variable sites (SNPs) among our genomes, we used bcftools v1.1 (Li, 2011), AntCaller v1.1 (Zhou et al., 2017), and GATK HaplotypeCaller to identify variants. We intersected the three call sets with VCFtools v0.1.16 vcf-isec (Danecek et al., 2011), and used only variants called by all three programs for downstream analysis. We filtered the final data set by removing SNPs with base call quality <20, covered by fewer than 5 reads, and located within 5 bp of indels. To limit the impact of multi-copy genes, we also removed SNPs with

depth of coverage greater than the 0.995 quantiles of coverage distribution for each genome. Using bedtools, we created a "filter" file with genome coordinates for exons, genome assembly gaps, tandem and simple repeats, paralogs, CpG islands, and clustered SNPs (genomic intervals with two or more clustered SNPs within 5, 10, and 50 bp windows), all of which we excluded from further analysis (see Figure S11 for the schematic). We extended each excluded region by 500 bp on both ends to account for selection and possible nearby genome mis-assemblies (Vitti, Grossman, & Sabeti, 2013). Using a custom python script, we ascertained a random set of 1 kb loci that would satisfy three criteria: (1) no overlap with the "filter" file coordinates; (2) a minimum of 30 kb interlocus distance to avoid linkage disequilibrium; and (3) a maximum of 10% missing nucleotides in each locus. This provided 4,215 putatively neutral 1-kb loci across all autosomes, separated by a median distance of 282 kb. To verify that the ascertained neutral loci data set followed the expected phylogenetic relationships between the samples, we concatenated all loci and used this alignment to reconstruct a phylogeny in Geneious v2020.0.5 (https://www.geneious.com) using the RAxML v8.2.11 plugin (Stamatakis, 2014) (Figure S12).

We ran G-PhoCS on our ascertained data set of putatively neutral 1 kb loci. Since G-PhoCS is computationally intensive, we limited each run to only four samples: an outgroup (donkey) and one each of a present-day, ancient Eurasian, and ancient North American horse genome (Figure 3a). This provided eight possible combinations. In each, we modeled three bi-directional migration bands, or intervals during which gene flow could have occurred between populations: m1 between ancestral North American and ancestral Eurasian populations; m2 between ancient North American and present-day horses; and m3 between ancient North American and ancient Eurasian populations. Migration bands are bounded by the time when the populations connected by gene flow existed; for example, the m1 migration band spans the entire time between $T_{\text{North America-Eurasia}}$

$_{\text{divergence}}$ and $T_{\text{Within Eurasia divergence}}$ (Figure 3a). We incorporated the following priors in the form of a gamma-distribution with the scale parameter β = 20,000: ages (τ) of the present-day samples were set to 0, ages of the ancient samples were set to their calibrated radiocarbon ages, the age of the root was set to the MRCA of *Equus* 4.5 Ma (Orlando et al., 2013), the MRCA of caballine horses was set to 1.0 Ma (Heintzman et al., 2017; Orlando et al., 2013), and the split between Eurasian populations was set to the younger boundary of 0.127 Ma, following (Schubert et al., 2014). Posterior distributions of each parameter were re-calculated to divergence times in units of years (T_{div}) and effective population sizes (N_e) by scaling with a per-generation mutation rate (μ) of 7.24E-9 and a generation time (g) of 8 years (Orlando et al., 2013). We scaled the parameters using $N_e=\theta/4\mu$, $T_{div}= \tau g/\mu$, where θ is Watterson's estimator denoting population mutation rate (Figure S13).

The G-PhoCS migration band is modeled as a migration rate between two lineages over the time period when both of these lineages existed (Figure 3). Thus we calculated the total migration rate m_{total} by scaling the per generation migration rate $m_{A\rightarrow B}$ by the duration of the corresponding migration bands $\tau_{A\rightarrow B}$ (the length of the entire branch when both A and B populations existed). The m_{total} estimate is free of the assumption about the mutation rate (Gronau et al., 2011). Estimates of m_{total} approximate the probability that a locus sampled in a target population originated in the source population. To account for uncertainty in μ and g, we randomly subsampled these parameters from their gamma distributions while calibrating the G-PhoCS results: for μ*E-9 we used Gamma(α=784, λ=112); for g - Gamma(α=21, λ=2.6) (Figure S14).

We ran each model through the MCMC process three times to ensure a sufficient effective sampling size and parameter convergence, which we monitored in Tracer v1.7.1 (Rambaut et al.,

2018). We ran each MCMC chain for 10 million steps sampled every 10 iterations, removing the first 5% as burn-in, and allowed automatic fine-tuning of the parameters for the first 1,000 steps.

Next, we estimated changes in effective population size from each genome using the pairwise sequentially Markovian coalescent (PSMC) model (Li & Durbin, 2011). We generated a pseudohaploidized consensus sequence of the autosomes with samtools mpileup using -EA options and -C50, followed by fq2psmcfa -q20 (Li et al., 2009). We ran PSMC with parameters -N30 -t15 -r5 -p "4+25*2+4+6" (Li & Durbin, 2011) (Figure S15).

Results

Mitochondrial and nuclear palaeogenomes of Pleistocene horses

We extracted ancient DNA from a total of 262 horse skeletal and dental remains, from which we selected 78 with sufficient preserved DNA for further processing. When reads were aligned to the EquCab2 reference genome, all 78 libraries exhibited characteristic post-mortem ancient DNA damage and signal of depurination-driven DNA fragmentation: excesses of C>T and G>A transitions at the ends of the molecules, excess of purines in sites preceding read start positions, and short DNA fragment length (<100 bp) (Table S5, Figure S4-S5, S7). Using a combination of shotgun sequencing and hybridization-based enrichment, we assembled 19 new ancient Eurasian and 59 ancient North American horse mitochondrial genomes that ranged in coverage from 4.2-fold to 2,287-fold (median 22.8-fold) (Table S1), and two nuclear genomes of North American horses to an average coverage of 25.1-fold and 25.9-fold (Table 1, Figure S6). We complemented these data with previously published data to assemble a final data set of 190 mitochondrial and nine nuclear genomes (Figure 1a, Table 1, Tables S1, S2). Below, we refer to horse populations on alternate sides of the Bering Land Bridge as “North American” or “Eurasian,”

and to the ancient horse nuclear genomes by their geographically precise locations or by their museum catalogue numbers.

Mitochondrial phylogeny of caballine horses

Horse mitochondrial genomes are strongly phylogeographically structured within Beringia, with two major clades corresponding to North America and Eurasia (Figures 1a,b; Figure S1-S3; Tables S1-S4), to which we assign clade designations following Weinstock et al. (2005). On the tip-calibrated BEAST analysis that included three outgroup sequences, we estimated that Clade A and B diverged 859-589 ka (95% HPD; median: 722 ka) (Figure S2, Table S4). This date is similar to that recovered with no outgroup (940-759 ka 95% HPD; median: 845 ka) and consistent with previous estimates (Heintzman et al., 2017; Orlando et al., 2013; Yuan et al., 2020). This timing coincides with the earliest of two dispersal events across the BLB identified in the CTMC model, which occurred around 875-625 ka in an east to west direction (Figure 1c).

We estimate a second dispersal across the BLB ~200-50 ka, predominantly in a west-to-east direction (Figure 1c). This signal is driven by two clusters of North American horses within clade A, which we designate A1 and A2 (Figures 1b). Clade A1 originated around 160-117 ka (95% HPD; median: 136 ka) and Clade A2 around 144-93 ka (95% HPD; median: 117 ka). Both A1 and A2 clades dispersed eastward at the same time, suggesting a single synchronous migration. Horses from clades A and B overlap geographically and temporally (Figure 1b; Figures S1-S3; Tables S1-S2). Western Beringian horses of the Clade B, in turn, are not monophyletic, in accordance with their maternal lineage redistribution during domestication (Fages et al., 2019; Gaunitz et al., 2018).

We found an unclear signal in the branching order of Clade C (Figure 1b). In our maximum likelihood analysis, samples from present-day China, Russian Far East and New Siberian Islands (Clade C) cluster with Clade B, but with weak bootstrap support (61, see Figure S1). Previous studies, however, suggest monophyly of Clade C with the Eurasian Clade A (Yuan et al., 2020). We found that rooting of the tree with outgroup sequences impacts the phylogenetic position of this clade, in agreement with previously published results (Fages et al., 2019; Heintzman et al., 2017; Yuan et al., 2020). Thus Clade C may represent a deeply-divergent population of Eurasian caballine horses that have not been sampled sufficiently neither in mtDNA nor in whole genome data to resolve its polytomy at the root of the caballine horse tree.

Population structure

The genetic relationships between present-day and ancient caballine horses reflect geography and time. On the PCA, the first principal component separates Eurasian and North American horses, while the second highlights variation between ancient and present-day Eurasian horses (Figure 2a). In the admixture analysis, K=2 separates samples into two groups corresponding to the present-day continental division between Eurasia and North America (Figure 2b; Figure S8). We found a small proportion of North American ancestry in CGG10022 (0.70%) and Przewalski's horses (0.65%). The early-Middle Pleistocene horse TC21, which lived close to the time of the initial divergence between Eurasian and North American horse populations, exhibits a mixed ancestry. Consistent with the antiquity of its genome, it falls between all other sampled individuals in both analyses (Figure 1b and (Orlando et al., 2013)).

D-statistic and D_{FOIL}

We found a significant excess of alleles shared among ancient Siberian, ancient Yukon, and present-day horses with the D-statistic test and D_{FOIL} analysis (Figures 2c,d. S8). Out of 11,225 200kb non-overlapping genomic windows, ~10% (n=1156) were admixed between Yukon, YG188.42 and YG303.325, and ancient Taymyr, CGG10022 and CGG10023 (Table S6, Figure S10). For the Batagai, Przewalski's, and Twilight horses we found ~7.7% admixed windows (n=869.7 on average across three D_{FOIL} tests) (Figure S10, Table S6). We observed similar allele sharing patterns on both autosomes and the X-chromosome regardless of whether nucleotide transitions, which are more strongly impacted by post-mortem DNA damage, were included or excluded (Figures 2d, S9). For each five-population topology tested with D_{FOIL}, we found bi-directional gene flow between Yukon genomes and either of the Eurasian individuals (Figure S10).

We found a contradicting pattern of gene flow between Batagai and Yukon horses. In the D-statistic test, Batagai does not have a statistically significant excess of allele sharing with either of the North American horses (Figure 2d). However, in the D_{FOIL} analysis, we found 7-8% of genomic windows admixed between Batagai and Yukon horses in either direction (Table S6, Figure S10). We note that the sequencing error rate in the Batagai genome is lower than error rates in CGG10022 and CGG10023 (Appendix S2, Table S5), which can inflate sensitivity of the D-statistic when these genomes are tested as P1 and P2, but not as P3 ((Rasmussen et al., 2011). Nevertheless, in the D-statistic analysis, Batagai did not share excess alleles with either of the Yukon horses when tested as P3 or with a present-day horse when tested as P1. For the D_{FOIL} analysis, we only tested Batagai with present-day Eurasian horses that had similar error rates (Figure S10). Batagai's contradicting admixture pattern held regardless of the four and five-taxon test configuration.

Gene flow and demographic history

To further explore the timing and extent of admixture between Eurasian and North American horse populations, we fitted a demographic model using G-PhoCS (Figures 3a,b,c, S13) (Gronau et al., 2011). We estimated the total migration rate within three bidirectional migration bands connecting horse lineages corresponding to mitochondrial clades A and B (denoted m1-3, Figure 3a,b), as well as the divergence time among them (Figure 3c). Assuming a mutation rate centered around 7.242E-9 per site per generation (Orlando et al 2013), caballine horses diverged from their sister species *E. asinus* 4.4±1 Ma (Figure 3c). Eurasian and North American caballine horse populations diverged 0.818±0.187 Ma, which is slightly younger than a previous estimate of 1.062 Ma (Orlando et al., 2013) but consistent with the 1.0-0.7 Ma estimate of their mitochondrial divergence (Heintzman et al., 2017) and our time-calibrated mitochondrial phylogeny (Figure 1b). In analyses in which the Przewalski's horse was in the position of the present-day population, we estimated a slightly older divergence between the ancestors of extinct and present-day Eurasian caballine horses (416±0.9 ka) compared to 346±0.8 ka with the Twilight horse in this position (note that these estimates do not account for post-divergence gene flow between extinct Eurasian horses and the ancestors of domestic horses). These estimates are similar to the 384 ka BP estimate from (Schubert et al., 2014), but younger than a ~625-500 ka estimate (Orlando et al., 2013) that included the Middle Pleistocene TC21 genome. Because lineages diverge gradually in the presence of gene flow, these estimates represent upper boundaries (Edwards & Beerli, 2000).

During the period that followed the initial divergence between Eurasian and North American populations (m1), we estimated a total migration rate of 4.0±0.8% across the Bering Land Bridge (Figure 3b). The total migration rate is determined by multiplying the number of migrants per generation with the time span of the migration band, which in this case corresponds

to 3-4% migration probability during the Middle Pleistocene. After this initial divergence, the migration rate declined to ~2%. While the migration rate in all three bands differs from zero, we do not have statistical power given our sample size to determine directionality. Finally, in contrast to the D-statistic tests but similar to the D_{FOIL} result, the Batagai genome displays signatures of gene flow from either of the Yukon horses.

G-PhoCS estimates suggest that the effective population size (N_e) of the North American horse population was smaller than that of the Pleistocene Taymyr horses ($N_e \approx 30{,}000-32{,}000$ and $N_e \approx 48{,}000$, accordingly, Figures S13). This smaller N_e is also reflected in PSMC plots from each horse (Figure S15). The estimated N_e for the ~5.5 ka old Batagai horse is $\approx 18{,}000$, which is consistent with the sharp decline in its PSMC profile beginning ~10,000 years prior to its death (Figure S15) and previous reports of a pre-domestication bottleneck (Librado et al., 2015). Estimates of N_e are similar for the Przewalski ($N_e \approx 19{,}000$) and Thoroughbred ($N_e \approx 18{,}000$) horses.

Discussion

Although the Bering Land Bridge (BLB) was a barrier for dispersal for much of the last million years of horse evolutionary history, horses used the land bridge when it was viable. Caballine horse populations on the two sides of the corridor maintained low levels of gene flow and biological connectivity until at least the time of the Last Glacial Maximum, 26-20 ka (Figure 4). Signatures of bi-directional dispersal across the BLB are present in both our mitochondrial and nuclear data sets. The latter provided an estimate of continuous gene flow equivalent to 2-4% of the population migrating across the BLB during the Pleistocene and ~0.6-0.7% of North American

ancestry in ancient and present-day horses in Eurasia (Figures 2b,c, 3b, S9, S10). This suggests that caballine horses in Beringia functioned as a metapopulation, at least until the local extinction of horses in North America and flooding of the BLB in the early Holocene (Figure 4).

Our mitochondrial phylogeny reveals at least two intercontinental dispersal events across the BLB during the Pleistocene (Figures 1, 4)**.** The first, which we estimate to have been predominantly in the east-to-west direction, occurred between ~0.95 and ~0.45 Ma (Figures 1c, 4). The start of the dispersal event is consistent with the age of the earliest recognized caballine horses in Eurasia ~0.9-0.8 Ma (Azzaroli, 1983, 1989; Forsten, 1996), and therefore may reflect the first wave of caballine horse dispersal into Eurasia from North America. The second dispersal occurred between ~0.2 and ~0.05 Ma, a time interval corresponding to an open BLB during the lowered sea level associated with MIS 6 glaciation. This second dispersal was bi-directional but dominated by west-to-east movement from Eurasia into North America. Horses dispersing into North America during this interval migrated east as far as Alaska's North Slope and Northern Yukon, where members of the Eurasian-origin Clade A1 co-occur temporally with the established North American Clade B, providing opportunity for horses of these two mitochondrial lineages to interbreed.

The temporal windows identified in the mitochondrial phylogeny coincide with periods when the BLB would have been open and available to dispersing megafauna, but are otherwise limited by our sampling (few samples from Middle and early-Late Pleistocene, and no North American samples that lived more recently than ~13 ka BP). Our data may therefore fail to capture other intervals during which horses could have dispersed across the BLB (e.g., between ~800 and 50 ka, and after 13 ka). Geological and palaeoclimatic data indicate that the BLB was flooded after ca. 13 ka (Elias, Short, Hans Nelson, & Birks, 1996; England & Furze, 2008; Hu et al., 2012) and

perhaps as late as ca. 11 ka (Jakobsson et al., 2017), which would have provided opportunity for gene flow until the early Holocene. Data from the Middle Pleistocene and late surviving North American horses, may reveal additional dispersal events across the BLB.

Although mitochondrial data can only identify dispersal events across the BLB, our nuclear genomic analyses confirm that dispersal events led to bi-directional gene flow between horse populations in Eurasia and North America (Figures 2b,c,d 3a,b, 4, Table S6, Figure S10). Immediately following the initial colonization of Eurasia by horses ~0.8 Ma, gene flow between the continents was maintained at a rate of around 3-4%. This migration rate declined after the Middle Pleistocene to around 2%, suggesting that, while gene flow was maintained across the BLB until at least the LGM, relatively fewer individuals were dispersing, or fewer dispersing individuals were integrating into local populations, during the Late Pleistocene (Figure 3c). The decline may be due in part to decreasing availability of the BLB during ca. 130-60 ka, and perhaps as late as 40 ka, as the Bering Sea level hovers close to the transgression depth at that time (Hu et al., 2012, 2010). This higher sea level may have enhanced an ecological barrier between western and eastern Beringia if the Mammoth Steppe was less permeable (e.g. Guthrie, 2001). Other factors also may have contributed to the overall reduction in gene flow, such as challenges to immigrants integrating into established populations, as yet unknown ecological barriers limiting local dispersals, or postzygotic isolation (Coughlan & Matute, 2020; H. Allen Orr, Masly, & Presgraves, 2004; H. A. Orr & Presgraves, 2000; Savolainen, Lascoux, & Merilä, 2013; Turelli & Orr, 2000). Teasing apart the relative roles of these processes will require denser sampling of horses from across Eastern Beringia.

The decline in gene flow over time is also supported by D-statistic tests (Figures 2c,d, S9). CGG10022 is dated to ~43 ka BP, and appears to be the most admixed in the panel. CGG10023,

which is dated to ~16 ka BP, shares fewer alleles than the older Siberian horse does with either Yukon horse, and the ~5 ka BP Batagai horse does not appear to share an excess of derived alleles with either of the Yukon horses (Figures 2b,c). The decreasing admixture signal may be due to selection against North American ancestry, perhaps similar to observed selection against deleterious Neanderthal alleles in humans (Juric, Aeschbacher, & Coop, 2016), or may result from subsequent gene flow with an unsampled lineage without North American ancestry (Figure 4). While we prefer the latter explanation, given that Western and Eastern Beringian horses were adapted to such similar environments, additional genomic data will be necessary to test these alternate hypotheses. Whole genome data, such as high coverage genomes or targeted enrichment of ancestry informative SNPs, from clade A1 and A2 horses has the potential to provide additional insight into gene flow patterns between Western and Eastern Beringian horses.

Additional horse genomes will also continue to refine understanding of how genetic diversity was partitioned within continental populations. For example, we found evidence that an as-yet unsampled horse population contributed ancestry to the Batagai horse (Figures 2c, 3b, 4). While we observed post-divergence gene flow between the Siberian, domestic, and Yukon horses, D-statistics did not indicate shared ancestry between Batagai and either Yukon horse. However, G-PhoCS and D_{FOIL} suggested gene flow across the BLB when Batagai was included (Figure 3b). These contrasting patterns may be due to ancestral population structure within Eurasia. Thereby, North American horses contributed alleles into some populations but not others, and an as-yet unsampled Eurasian population contributed ancestry into the Batagai genome but not to the more ancient horse from Taymyr (Figure 4). Possibly supporting this hypothesis, a previous study suggested that yet unsampled population contributed ancestry to some ancient Siberian horse populations (Fages et al., 2019; Orlando, 2020). The evolutionary history of uniparental markers

differs in Eurasian horses, with Batagai falling either within (per mtDNA) or outside (per Y-chromosome) the diversity of other Eurasian horses analyzed here (Fages et al., 2019), further hinting at an unsampled source of ancestry. A potential candidate for this population with unclear nuclear genome ancestry is the East Asian mitochondrial Clade C (Figures 1b, 4), assigned the name *Equus dalianensis* by Yuan et al (2020). *E. dalianensis* had a range that extended westward to the region of present-day Kazakhstan, the putative center of horse domestication (Librado & Orlando, 2020; Orlando, 2020), and northward to Chukotka, where the Batagai horse was recovered (Librado et al., 2015), but not as far northward as Taymyr, where CGG10022 was found (Orlando et al., 2013). Although additional genetic data from *E. dalianensis* will be necessary to test this hypothesis directly, our results suggest a more complex model of horse population structure and dispersals than is reflected in existing data. While the maternally inherited non-recombining mitochondrial locus provides limited information, nuclear genome data enables direct estimates of gene flow and population isolation in cases of complex evolutionary histories, such as those of caballine horses. Thus, whole-genome sequencing of even a few samples is a more promising approach for resolving this question.

Our results have implications for horse taxonomy and palaeontology. We observed that present-day Przewalski's and domestic horses have a proportion of genomic ancestry that derives from relatively recent gene flow from extinct North American horses. From a systematic perspective, fossils of Late Quaternary caballine equids from Western and Eastern Beringia have been attributed on morphological grounds to a variety of nominally separate species, including *Equus ferus*, *E. lambei*, *E. alaskae, E. lenensis, E. dalianensis*, and *E. scotti* among many others (Azzaroli, 1989; Azzaroli & Voorhies, 1993; Barron-Ortiz et al., 2019; Barrón-Ortiz et al., 2017). Some of these are surely *nomina nuda* (empty names), with no real standing in biology even if

their nomenclature is correctly formed. Although recent genomic investigations have shown that evolutionary diversification has occurred in caballine populations during the Pleistocene (e.g., Weinstock et al., 2005; Barron-Ortiz et al., 2017; Heintzman et al., 2017), the number of lineages sufficiently distinct to warrant formal taxonomic recognition is small. In the case of the Pleistocene Beringian caballines considered here, it is reasonable to view them as a bicontinental metapopulation, component parts of which were in intermittent biological contact via the BLB. How this should be reflected systematically is, however, not straightforward. For example, although complete reproductive isolation of caballines on opposite sides of the Bering Strait was evidently never achieved, the comparatively low hybridization rate detected in our samples indicates that meta-population mixing was limited even when the BLB was available. Thus a different method of discriminating population boundaries is needed.

One way of recognizing continuities as well as differences among closely related populations within a widely-distributed species is to distinguish subspecies. Ideally, formally defined subspecies ought to have congruent phenotypic and genotypic identities and live in circumscribed geographical spaces, so that they can be distinguished from other clusters from which they are spatially, but not reproductively, isolated (Patten, 2015). This is a hard standard to meet when dealing with fossil populations, which can only be partially characterized for these parameters. Indeed, because of the blurring effects of convergence and hybridization within *Equus ferus*, the only generally recognized horse subspecies is extant *E. f. przewalskii,* previously considered a different species (Gaunitz et al., 2018). As Barrón-Ortiz and colleagues (2017) have pointed out, efforts to distinguish extinct populations of *E. ferus* from the Western Interior of North America as subspecies (e.g., *E. f. lambei*, *E. f. scotti*) does not clarify their evolutionary position because morphological and mtDNA analysis of these otherwise distinct populations demonstrates

their extensive similarities. Furthermore, Eurasian populations of *Equus* sp. have markedly different evolutionary histories when their mitochondrial, Y-chromosome, and autosomal phylogenies are compared (Fages et al., 2019) and this study, Figure 4). Defining populations using individual and/or uniparentally-inherited genetic markers is as equally challenging as using size-related morphological differences as a basis for taxonomic classification (see (Saarinen, Cirilli, Strani, Meshida, & Bernor, 2021). Such maneuvers may be theoretically justifiable, but, in the case of caballine horses, risk to provide little insight into this taxon's true evolutionary history and systematic status. Until a strong basis for formalizing biological distinctions among Quaternary Holarctic caballine horses can be established, it is preferable to avoid formal trinomials or similar devices, and to refer instead to "Eurasian clades" and "North American clades" as we have done here.

Finally, our study demonstrates the advantage of genomics in identifying the role of the Bering Land Bridge not only as a biogeographical corridor, but as a critical contact zone where significant evolutionary processes unfolded for Holarctic cold adapted taxa (Figure 4). Our data show that horses dispersed back into North America from Eurasia around the same time as the initial expansion of bison, brown bears, and lions through the BLB (Froese et al., 2017; Salis et al., 2020). When present, the BLB clearly played a key role as a bi-continental dispersal corridor for many taxa, including ones that used it more than once (Debruyne et al., 2008; Elias & Crocker, 2008; Froese et al., 2017; Meiri et al., 2020, 2014). This Ice Age fauna thrived on the Mammoth Steppe, the widespread steppe-tundra biome present in Beringia during cold periods of the Pleistocene (Guthrie, 1982; Zimov et al., 2012). With the onset of dramatic climate change and ultimate disappearance of the BLB at the end of the Pleistocene, the biogeographical importance of this ecological corridor fundamentally changed for terrestrial taxa.

Acknowledgments

We are grateful to the placer gold mining community, the Tr'ondëk Hwëch'in First Nation, and the Vuntut Gwitchin First Nation for their collaboration and support with our research in Yukon. Tamara Pico, Sarah Crump, and Paul Koch provided advice interpreting the palaeoclimate of Beringia. Data generation was supported with funds from NSF ARC-1417036, the Gordon & Betty Moore Foundation (#3804) and the American Wild Horse Campaign. AOV was additionally supported by a UCSC Chancellor's Dissertation Year Fellowship and the CANA Foundation. MT and DG were supported by Russian Foundation for Basic Research (project 18–04–00327), and LE by a Marie-Curie Intra-European fellowship (FP7, IEF-302617). GB is supported by the Federal theme of Zoological Institute of the Russian Academy of Sciences no. AAAA-A19-119032590102-7. SV was funded by Russian Foundation for Basic Research, Grant 19-05-00477. LD acknowledges support from FORMAS (project 2018-01640). DF was supported by the Natural Science and Engineering Research Council. PG and DM are supported by the Bureau of Land Management. We acknowledge support from Science for Life Laboratory, the Knut and Alice Wallenberg Foundation, the National Genomics Infrastructure funded by the Swedish Research Council, and Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. LO received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 681605).

References

Alter, S. E., Meyer, M., Post, K., Czechowski, P., Gravlund, P., Gaines, C., … Hofreiter, M. (2015). Climate impacts on transocean dispersal and habitat in gray whales from the Pleistocene to 2100. *Molecular Ecology*, *24*(7), 1510–1522.

Azzaroli, A. (1983). Quaternary mammals and the "end-Villafranchian" dispersal event—a turning point in the history of Eurasia. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *44*(1-2), 117–139.

Azzaroli, A. (1989). The Genus Equus in Europe. In E. H. Lindsay, V. Fahlbusch, & P. Mein (Eds.), *European Neogene Mammal Chronology* (pp. 339–356). Boston, MA: Springer US.

Azzaroli, A., & Voorhies, M. R. (1993). The genus Equus in North America. The Blancan species. *Palaeontographia Italica*, *80*, 175–198.

Barlow, A., Cahill, J. A., Hartmann, S., Theunert, C., Xenikoudakis, G., Fortes, G. G., … Hofreiter, M. (2018). Partial genomic survival of cave bears in living brown bears. *Nature Ecology & Evolution*, *2*(10), 1563–1570.

Barron-Ortiz, C., Avilla, L., Jass, C., Bravo-Cuevas, V., Machado, H., & Mothé, D. (2019). What is Equus? Reconciling taxonomy and phylogenetic analyses. *Frontiers in Ecology and Evolution*, *7*, 343.

Barrón-Ortiz, C. I., Rodrigues, A. T., Theodor, J. M., Kooyman, B. P., Yang, D. Y., & Speller, C. F. (2017). Cheek tooth morphology and ancient mitochondrial DNA of late Pleistocene horses from the western interior of North America: Implications for the taxonomy of North American Late Pleistocene Equus. *PloS One*, *12*(8), e0183045.

Bond, J. D. (2019). Paleodrainage map of Beringia. Retrieved November 20, 2020, from Yukon

Geological Survey website: http://data.geology.gov.yk.ca/Reference/81642#InfoTab

Cahill, J. A., Stirling, I., Kistler, L., Salamzade, R., Ersmark, E., Fulton, T. L., … Shapiro, B. (2015). Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Molecular Ecology*, *24*(6), 1205–1217.

Chang, D., Knapp, M., Enk, J., Lippold, S., Kircher, M., Lister, A., … Shapiro, B. (2017). The evolutionary and phylogeographic history of woolly mammoths: a comprehensive mitogenomic analysis. *Scientific Reports*, *7*, 44585.

Coughlan, J. M., & Matute, D. R. (2020). The importance of intrinsic postzygotic barriers throughout the speciation process. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *375*(1806), 20190533.

Coyne, J. A., & Orr, H. A. (1998). The evolutionary genetics of speciation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *353*(1366), 287–305.

Dabney, J., & Meyer, M. (2019). Extraction of Highly Degraded DNA from Ancient Bones and Teeth. In B. Shapiro, A. Barlow, P. D. Heintzman, M. Hofreiter, J. L. A. Paijmans, & A. E. R. Soares (Eds.), *Ancient DNA: Methods and Protocols* (pp. 25–29). New York, NY: Springer New York.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics* , *27*(15), 2156–2158.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772.

Debruyne, R., Chu, G., King, C. E., Bos, K., Kuch, M., Schwarz, C., … Poinar, H. N. (2008).

Out of America: ancient DNA evidence for a new world origin of late quaternary woolly mammoths. *Current Biology: CB*, *18*(17), 1320–1326.

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, *28*(8), 2239–2252.

Dyke, A. S. (2004). An outline of North American deglaciation with emphasis on central and northern Canada. *Quaternary Glaciations-Extent and Chronology - Part II: North America*, pp. 373–424. doi: 10.1016/s1571-0866(04)80209-4

Eddelbuettel, D., & Wu, W. (2016). RcppCNPy: Read-Write Support for NumPy Files in R. *Journal of Open Source Software*, *1*(5), 55.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.

Edwards, S. V., & Beerli, P. (2000). Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution; International Journal of Organic Evolution*, *54*(6), 1839–1854.

Elias, S. A., & Brigham-Grette, J. (2013). Late Pleistocene Glacial Events in Beringia. In S. A. Elias & C. J. Mock (Eds.), *Encyclopedia of Quaternary Science (Second Edition)* (pp. 191–201). Amsterdam: Elsevier.

Elias, S. A., & Crocker, B. (2008). The Bering Land Bridge: a moisture barrier to the dispersal of steppe–tundra biota? *Quaternary Science Reviews*, *27*(27), 2473–2483.

Elias, S. A., Short, S. K., Hans Nelson, C., & Birks, H. H. (1996). Life and times of the Bering land bridge. *Nature*, Vol. 382, pp. 60–63. doi: 10.1038/382060a0

England, J. H., & Furze, M. F. A. (2008). New evidence from the western Canadian Arctic Archipelago for the resubmergence of Bering Strait. *Quaternary Research*, *70*(1), 60–67.

Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., … Orlando, L. (2019). Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. *Cell*, *177*(6), 1419–1435.e31.

Forstén, A. (1991). Mitochondrial-DNA time-table and the evolution of Equus: comparison of molecular and paleontological evidence. *Annales Zoologici Fennici*, *28*(3/4), 301–309.

Forsten, A. (1996). Climate and the evolution of Equus [Perissodactyla, Equidae] in the Plio-Pleistocene of Eurasia. *Acta Zoologica Cracoviensia*, *1*(39). Retrieved from https://www.infona.pl/resource/bwmeta1.element.agro-article-373d902b-114a-4ab2-9de1-0103dcdb0f82

Froese, D., Stiller, M., Heintzman, P. D., Reyes, A. V., Zazula, G. D., Soares, A. E. R., … Shapiro, B. (2017). Fossil and genomic evidence constrains the timing of bison arrival in North America. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(13), 3457–3462.

Fulton, T. L., & Shapiro, B. (2019). Setting Up an Ancient DNA Laboratory. In B. Shapiro, A. Barlow, P. D. Heintzman, M. Hofreiter, J. L. A. Paijmans, & A. E. R. Soares (Eds.), *Ancient DNA: Methods and Protocols* (pp. 1–13). New York, NY: Springer New York.

Gamba, C., Hanghøj, K., Gaunitz, C., Alfarhan, A. H., Alquraishi, S. A., Al-Rasheid, K. A. S., … Orlando, L. (2016). Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Molecular Ecology Resources*, *16*(2), 459–469.

Gaunitz, C., Fages, A., Hanghøj, K., Albrechtsen, A., Khan, N., Schubert, M., … Orlando, L. (2018). Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science*, *360*(6384), 111–114.

Gill, M. S., Lemey, P., Faria, N. R., Rambaut, A., Shapiro, B., & Suchard, M. A. (2013).

Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Molecular Biology and Evolution*, *30*(3), 713–724.

Gordon, A., Hannon, G. J., & Others. (2010). Fastx-toolkit. *FASTQ/A Short-Reads Preprocessing Tools (unpublished) Http://hannonlab. Cshl. Edu/fastx_toolkit*, *5*.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., … Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*, *328*(5979), 710–722.

Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., & Siepel, A. (2011). Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics*, *43*(10), 1031–1034.

Guthrie, R. D. (1982). Mammals of the mammoth steppe as paleoenvironmental indicators. In D. M. Hopkins, J. V. Matthews, C. E. Schweger, & S. B. Young (Eds.), *Paleoecology of Beringia* (pp. 307–326). Academic Press.

Guthrie, R. D. (2001). Origin and causes of the mammoth steppe: a story of cloud cover, woolly mammal tooth pits, buckles, and inside-out Beringia. *Quaternary Science Reviews*, *20*(1), 549–574.

Guthrie, R. D. (2006). New carbon dates link climatic change with human colonization and Pleistocene extinctions. *Nature*, *441*(7090), 207–209.

Haile, J., Froese, D. G., Macphee, R. D. E., Roberts, R. G., Arnold, L. J., Reyes, A. V., … Willerslev, E. (2009). Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(52), 22352–22357.

Heintzman, P. D., Zazula, G. D., MacPhee, R. D. E., Scott, E., Cahill, J. A., McHorse, B. K., … Shapiro, B. (2017). A new genus of horse from Pleistocene North America. *eLife*, *6*. doi:

10.7554/eLife.29944

Hill, V., & Baele, G. (2019). Bayesian estimation of past population dynamics in BEAST 1.10 using the Skygrid coalescent model. *Molecular Biology and Evolution*. doi: 10.1093/molbev/msz172

Hopkins, D. M. (1959). Cenozoic History of the Bering Land Bridge. *Science*, *129*(3362), 1519–1528.

Hopkins, D. M. (1959). Cenozoic History of the Bering Land Bridge: The seaway between the Pacific and Arctic basins has often been a land route between Siberia and Alaska. *Science*, Vol. 129, pp. 1519–1528. doi: 10.1126/science.129.3362.1519

Hopkins, D. M. (1973). Sea Level History in Beringia During the Past 250,000 Years1. *Quaternary Research*, *3*(4), 520–540.

Hu, A., Meehl, G. A., Han, W., Timmermann, A., Otto-Bliesner, B., Liu, Z., … Wu, B. (2012). Role of the Bering Strait on the hysteresis of the ocean conveyor belt circulation and glacial climate stability. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(17), 6417–6422.

Hu, A., Meehl, G. A., Otto-Bliesner, B. L., Waelbroeck, C., Han, W., Loutre, M.-F., … Rosenbloom, N. (2010). Influence of Bering Strait flow and North Atlantic circulation on glacial sea-level changes. *Nature Geoscience*, *3*(2), 118–121.

Irizarry, R. A., Wu, H., & Feinberg, A. P. (2009). A species-generalized probabilistic model-based definition of CpG islands. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*, *20*(9-10), 674–680.

Jakobsson, M., Pearce, C., Cronin, T. M., Backman, J., Anderson, L. G., Barrientos, N., … O'Regan, M. (2017). Post-glacial flooding of the Bering Land Bridge dated to 11 cal ka BP

based on new geophysical and sediment records. *Climate of the Past*, *13*(8), 991–1005.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* , *29*(13), 1682–1684.

Juric, I., Aeschbacher, S., & Coop, G. (2016). The Strength of Selection against Neanderthal Introgression. *PLoS Genetics*, *12*(11), e1006340.

Kapp, J. D., Green, R. E., & Shapiro, B. (2020). A fast and efficient single-stranded genomic library preparation method optimized for ancient DNA. *Manuscript Submitted for Publication*.

Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., & Kent, W. J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Research*, *32*(Database issue), D493–D496.

Kassambara, A., & Mundt, F. (2017). Factoextra: extract and visualize the results of multivariate data analyses. *R Package Version*, *1*(4), 2017.

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, *40*(1), e3.

Knebel, H. J., & Creager, J. S. (1973). Yukon River: Evidence for Extensive Migration during the Holocene Transgression. *Science*, *179*(4079), 1230–1232.

Korlević, P., & Meyer, M. (2019). Pretreatment: Removing DNA Contamination from Ancient Bones and Teeth Using Sodium Hypochlorite and Phosphate. In B. Shapiro, A. Barlow, P. D. Heintzman, M. Hofreiter, J. L. A. Paijmans, & A. E. R. Soares (Eds.), *Ancient DNA: Methods and Protocols* (pp. 15–19). New York, NY: Springer New York.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next

Generation Sequencing Data. *BMC Bioinformatics*, *15*, 356.

Librado, P., Der Sarkissian, C., Ermini, L., Schubert, M., Jónsson, H., Albrechtsen, A., … Orlando, L. (2015). Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(50), E6889–E6897.

Librado, P., Gamba, C., Gaunitz, C., Der Sarkissian, C., Pruvost, M., Albrechtsen, A., … Orlando, L. (2017). Ancient genomic changes associated with domestication of the horse. *Science*, *356*(6336), 442–445.

Librado, P., & Orlando, L. (2020). Genomics and the Evolutionary History of Equids. *Annual Review of Animal Biosciences*. doi: 10.1146/annurev-animal-061220-023118

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* , *27*(21), 2987–2993.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* , *25*(14), 1754–1760.

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–496.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , *25*(16), 2078–2079.

Lippold, S., Matzke, N. J., Reissmann, M., & Hofreiter, M. (2011). Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. *BMC Evolutionary Biology*, *11*, 328.

Lisiecki, L. E., & Raymo, M. E. (2005). A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. *Paleoceanography*, *20*(1). doi: 10.1029/2004PA001071

Lister, A. M., & Sher, A. V. (2015). Evolution and dispersal of mammoths across the Northern Hemisphere. *Science*, *350*(6262), 805–809.

Loog, L., Thalmann, O., Sinding, M.-H. S., Schuenemann, V. J., Perri, A., Germonpré, M., … Manica, A. (2019). Ancient DNA suggests modern wolves trace their origin to a Late Pleistocene expansion from Beringia. *Molecular Ecology*. doi: 10.1111/mec.15329

Lorenzen, E. D., Nogués-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K. A., … Willerslev, E. (2011). Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, *479*(7373), 359–364.

MacFadden, B. J. (2005). Fossil horses - evidence for evolution. *Science*, *307*(5716), 1728–1730.

Mann, D. H., Groves, P., Reanier, R. E., Gaglioti, B. V., Kunz, M. L., & Shapiro, B. (2015). Life and extinction of megafauna in the ice-age Arctic. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(46), 14301–14306.

Marques, D. A., Lucek, K., Sousa, V. C., Excoffier, L., & Seehausen, O. (2019). Admixture between old lineages facilitated contemporary ecological speciation in Lake Constance stickleback. *Nature Communications*, *10*(1), 4240.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303.

McKeon, C. S., Weber, M. X., Alter, S. E., Seavy, N. E., Crandall, E. D., Barshis, D. J., … Oleson, K. L. L. (2016). Melting barriers to faunal exchange across ocean basins. *Global Change Biology*, *22*(2), 465–473.

Meiri, M., Lister, A., Kosintsev, P., Zazula, G., & Barnes, I. (2020). Population dynamics and range shifts of moose (Alces alces) during the Late Quaternary. *Journal of Biogeography*, *47*(10), 2223–2234.

Meiri, M., Lister, A. M., Collins, M. J., Tuross, N., Goebel, T., Blockley, S., … Barnes, I. (2014). Faunal record identifies Bering isthmus conditions as constraint to end-Pleistocene migration to the New World. *Proceedings. Biological Sciences / The Royal Society*, *281*(1776), 20132167.

Meisner, J., & Albrechtsen, A. (2018). Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics*, *210*(2), 719–731.

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, *2010*(6), db.prot5448.

Minin, V. N., & Suchard, M. A. (2008a). Counting labeled transitions in continuous-time Markov models of evolution. *Journal of Mathematical Biology*, *56*(3), 391–412.

Minin, V. N., & Suchard, M. A. (2008b). Fast, accurate and simulation-free stochastic mapping. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1512), 3985–3995.

Moreno-Mayar, J. V., Potter, B. A., Vinner, L., Steinrücken, M., Rasmussen, S., Terhorst, J., … Willerslev, E. (2018). Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature*, *553*(7687), 203–207.

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews. Genetics*, *12*(6), 443–451.

Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society of London. Series B,*

Biological Sciences, *367*(1587), 332–342.

O'Regan, M., John, K. S., Moran, K., Backman, J., King, J., Haley, B. A., … Röhl, U. (2010). Plio-Pleistocene trends in ice rafted debris on the Lomonosov Ridge. *Quaternary International: The Journal of the International Union for Quaternary Research*, *219*(1), 168–176.

Orlando, L. (2020). Ancient Genomes Reveal Unexpected Horse Domestication and Management Dynamics. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *42*(1), e1900164.

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., … Willerslev, E. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, *499*(7456), 74–78.

Orr, H. A., Masly, J. P., & Presgraves, D. C. (2004). Speciation genes. *Current Opinion in Genetics & Development*, *14*(6), 675–679.

Orr, H. A., & Presgraves, D. C. (2000). Speciation by postzygotic isolation: forces, genes and molecules. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *22*(12), 1085–1094.

Ouedraogo, M., Bettembourg, C., Bretaudeau, A., Sallou, O., Diot, C., Demeure, O., & Lecerf, F. (2012). The duplicated genes database: identification and functional annotation of co-localised duplicated genes across genomes. *PloS One*, *7*(11), e50653.

Patten, M. A. (2015). Subspecies and the philosophy of science. *The Auk: Ornithological Advances*, *132*(2), 481–485.

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*(11), 2337–2360.

Pease, J. B., & Hahn, M. W. (2015). Detection and Polarization of Introgression in a Five-Taxon Phylogeny. *Systematic Biology*, *64*(4), 651–662.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* , *26*(6), 841–842.

Rambaut, A., & Drummond, A. J. (2010). TreeAnnotator version 1.6. 1. *University of Edinburgh, Edinburgh, UK*. Retrieved from https://scholar.google.ca/scholar?cluster=16034997356612446945&hl=en&as_sdt=0,5&sciodt=0,5

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904.

Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K. E., Rasmussen, S., Albrechtsen, A., … Willerslev, E. (2011). An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science*, *334*(6052), 94–98.

R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rohland, N., Reich, D., Mallick, S., Meyer, M., Green, R. E., Georgiadis, N. J., … Hofreiter, M. (2010). Genomic DNA sequences from mastodon and woolly mammoth reveal deep speciation of forest and savanna elephants. *PLoS Biology*, *8*(12), e1000564.

Saarinen, J., Cirilli, O., Strani, F., Meshida, K., & Bernor, R. L. (2021). Testing Equid Body Mass Estimate Equations on Modern Zebras—With Implications to Understanding the Relationship of Body Size, Diet, and Habitats of Equus in the Pleistocene of Europe. *Frontiers in Ecology and Evolution*, *9*, 90.

Salis, A. T., Bray, S. C. E., Lee, M. S. Y., Heiniger, H., Barnett, R., Burns, J. A., … Mitchell, K. J. (2020). *Lions and brown bears colonized North America in multiple synchronous waves of dispersal across the Bering Land Bridge* (p. 2020.09.03.279117). doi: 10.1101/2020.09.03.279117

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews. Genetics*, *14*(11), 807–820.

Schield, D. R., Perry, B. W., Adams, R. H., Card, D. C., Jezkova, T., Pasquesi, G. I. M., … Castoe, T. A. (2019). Allopatric divergence and secondary contact with gene flow: a recurring theme in rattlesnake speciation. *Biological Journal of the Linnean Society. Linnean Society of London*, *128*(1), 149–169.

Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* , *27*(6), 863–864.

Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J. F., Al-Rasheid, K. A. S., Willerslev, E., … Orlando, L. (2012). Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics*, *13*, 178.

Schubert, M., Jónsson, H., Chang, D., Der Sarkissian, C., Ermini, L., Ginolhac, A., … Orlando, L. (2014). Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(52), E5661–E5669.

Sher, A. (1999). Traffic lights at the Beringian crossroads. *Nature*, *397*(6715), 103–104.

Smit, A. F. A., Hubley, R., & Green, P. (2013-2015). RepeatMasker Open-4.0. Retrieved from http://www.repeatmasker.org

South, A. (2011, June). rworldmap: A New R package for Mapping Global Data. *The R Journal*,

Vol. 3, pp. 35–43. Retrieved from http://journal.r-project.org/archive/2011-1/RJournal_2011-1_South.pdf

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* , *30*(9), 1312–1313.

Stuart, A. J., & Lister, A. M. (2012). Extinction chronology of the woolly rhinoceros Coelodonta antiquitatis in the context of late Quaternary megafaunal extinctions in northern Eurasia. *Quaternary Science Reviews*, *51*, 1–17.

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, *4*(1), vey016.

Turelli, M., & Orr, H. A. (2000). Dominance, epistasis and the genetics of postzygotic isolation. *Genetics*, *154*(4), 1663–1679.

Vershinina, A. O., Kapp, J. D., Baryshnikov, G. F., & Shapiro, B. (2019). The case of an arctic wild ass highlights the utility of ancient DNA for validating problematic identifications in museum collections. *Molecular Ecology Resources*. doi: 10.1111/1755-0998.13130

Vilstrup, J. T., Seguin-Orlando, A., Stiller, M., Ginolhac, A., Raghavan, M., Nielsen, S. C. A., … Orlando, L. (2013). Mitochondrial phylogenomics of modern and ancient equids. *PloS One*, *8*(2), e55950.

Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting natural selection in genomic data. *Annual Review of Genetics*, *47*, 97–120.

Wade, C. M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., … Lindblad-Toh, K. (2009). Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science*, *326*(5954), 865–867.

Wang, L.-G., Lam, T. T.-Y., Xu, S., Dai, Z., Zhou, L., Feng, T., … Yu, G. (2020). Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. *Molecular Biology and Evolution*, *37*(2), 599–603.

Weinstock, J., Willerslev, E., Sher, A., Tong, W., Ho, S. Y. W., Rubenstein, D., … Cooper, A. (2005). Evolution, systematics, and phylogeography of pleistocene horses in the new world: a molecular perspective. *PLoS Biology*, *3*(8), e241.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Wooller, M. J., Saulnier-Talbot, É., Potter, B. A., Belmecheri, S., Bigelow, N., Choy, K., … Williams, J. W. (2018). A new terrestrial palaeoenvironmental record from the Bering Land Bridge and context for human dispersal. *Royal Society Open Science*, *5*(6), 180145.

Wu, H., Caffo, B., Jaffee, H. A., Irizarry, R. A., & Feinberg, A. P. (2010). Redefining CpG islands using hidden Markov models. *Biostatistics* , *11*(3), 499–514.

Xu, X., & Arnason, U. (1994). The complete mitochondrial DNA sequence of the horse, Equus caballus: extensive heteroplasmy of the control region. *Gene*, *148*(2), 357–362.

Xu, X., Gullberg, A., & Arnason, U. (1996). The complete mitochondrial DNA (mtDNA) of the donkey and mtDNA comparisons among four closely related mammalian species-pairs. *Journal of Molecular Evolution*, *43*(5), 438–446.

Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., … Flicek, P. (2020). Ensembl 2020. *Nucleic Acids Research*, *48*(D1), D682–D688.

Yuan, J., Sheng, G., Preick, M., Sun, B., Hou, X., Chen, S., … Hofreiter, M. (2020). Mitochondrial genomes of Late Pleistocene caballine horses from China belong to a separate clade. *Quaternary Science Reviews*, *250*, 106691.

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution / British Ecological Society*, *8*(1), 28–36.

Zazula, G. D., MacPhee, R. D. E., Metcalfe, J. Z., Reyes, A. V., Brock, F., Druckenmiller, P. S., … Southon, J. R. (2014). American mastodon extirpation in the Arctic and Subarctic predates human colonization and terminal Pleistocene climate change. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(52), 18460–18465.

Zheng, Y., & Janke, A. (2018). Gene flow analysis method, the D-statistic, is robust in a wide parameter space. *BMC Bioinformatics*, *19*(1), 10.

Zhou, B., Wen, S., Wang, L., Jin, L., Li, H., & Zhang, H. (2017). AntCaller: an accurate variant caller incorporating ancient DNA damage. *Molecular Genetics and Genomics: MGG*, *292*(6), 1419–1430.

Zimov, S. A., Zimov, N. S., Tikhonov, A. N., & Chapin, F. S. (2012). Mammoth steppe: a high-productivity phenomenon. *Quaternary Science Reviews*, *57*, 26–45.

Data Accessibility

Radiocarbon dates are uploaded to the Arctic Data Center (doi:10.18739/A28C9R40R). Raw sequencing data generated from YG303.325 and YG188.42 are available in NCBI BioProject PRJNA727160. All newly generated mitochondrial genomes are uploaded to GenBank with ID numbers MW846090-MW846167. BEAST files and the final G-PhoCS'filter' file are available on Data Dryad https://doi.org/10.7291/D18W9G. All scripts, as well as filtering criteria used to ascertain the set of putatively neutral loci, are published on github https://github.com/avershinina/BLB_horses.

Author contributions

BS, PG, DHM, LO, and RDM conceived the study. PG, DHM, BS, GZ, GB, DG, EH, SH, IK, PK, FS, HT, MT, ES, SV, LD performed field work and provided the samples. AOV, PDH, MCJ, CDS, SGD, LE, LD, CG, JDK, ASO, MS generated data. JS generated radiocarbon data. AOV and PDH analysed the data with contributions from RCD, BS, and LO. BS, RDM, AOV, PDH, RCD, GZ, DF, PG, DM, and LO interpreted the data. AOV wrote the initial manuscript draft. BS, RDM and PDH revised the manuscript with AOV, with contributions from LO, DF, RCD, GZ, PG, DHM, MJW and IK.

Tables and Figures (with captions)

Table 1. Ancient and present-day nuclear genomes analyzed in this study.

Coverage is reported based on alignment of whole genome data to the EquCab2.0 reference (Wade et al., 2009). Median and 1 sigma values are reported for dates calibrated with IntCal13 curve as implemented in OxCal 4.3 (Ramsey, 2009; Reimer et al., 2013). Previously published radiocarbon and stratigraphic dates: (*) (Orlando et al., 2013); (**) (Lorenzen et al., 2011).

Specimen ID	Nuclear Genome Coverage (fold)	Uncalibrated radiocarbon date	Calibrated age (years BP)	Radiocarbon facility ID	Reference
Batagai	18.3	4,450 ± 35	5,104 ± 109	Gr-50842	(Librado et al., 2015)
CGG10022	24.3	38,565 ± 602	42,620 ± 440	UBA-16478	(Schubert et al., 2014)*
CGG10023	7.4	13,389 ± 52	16,112 ± 96	UBA-16479	(Schubert et al., 2014)*
YG303.325	25.9	26,020 ± 140	30,318 ± 245	CAMS-157454	This study
YG188.42/ YT03-40	25.1	23,920 ± 100	27,947 ± 130	OxA-17686	This study**
YG148.20/ TC21	0.7	-	ca. 560,000-780,000	-	(Orlando et al., 2013)*
Thoroughbred (Twilight)	21.0	-	present-day	-	(Orlando et al., 2013)
Przewalski (MK822)	9.1	-	present-day	-	(Orlando et al., 2013)
Donkey (Willy)	11.8	-	present-day	-	(Orlando et al., 2013)

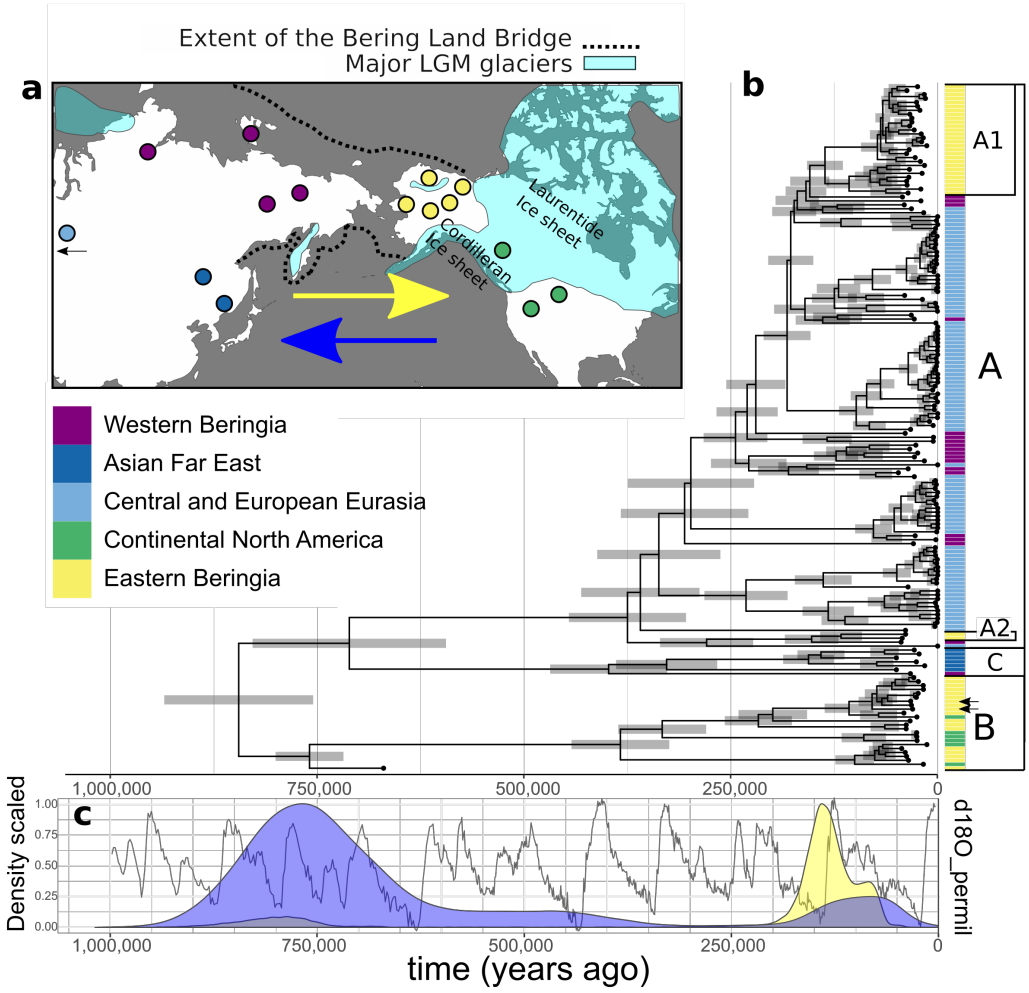

Figure 1. Sampling locations and phylogeographic analysis of ancient and present-day caballine horse mitochondrial genomes. a) Geographic distribution of the samples with palaeogeographic reconstruction of Beringia during the Last Glacial Maximum (LGM) ~25-23 ka years ago (redrawn from Dyke (2004)). Arrows designate direction of dispersal modelled in the CTMC Markov jumps analysis. **b)** Time-calibrated mitochondrial genealogy. Each sample was assigned to either Eurasian or North American continent based on their sampling location relative to the present-day Bering Strait. Gray bars represent 95% highest probability density (HPD) interval of node heights. Arrows show positions of YG188.42 and YG303.325. We calibrated radiocarbon ages of the Late Pleistocene samples reported in Table S1 using the IntCal13 curve (Reimer et al., 2013) and OxCal v4.2 (Ramsey, 2009), and assigned the median calibrated age of each sample as prior information. If a sample was dated using stratigraphic information (Fages et al., 2019), we assigned the mean of the tentative age interval as a date prior (Table S2). Samples without age information were excluded from the analysis. The version of the genealogy with tip labels is presented in the Figure S3. **c)** Density plots (scaled to 1) representing the timing of movements between Western and Eastern Beringia with $\delta^{18}O$ values plotted on the second Y-axis. Density colors indicate the direction of movement, following the arrows in panel (a). $\delta^{18}O$ curve is a proxy for global temperature and ice volume (Lisiecki & Raymo, 2005).

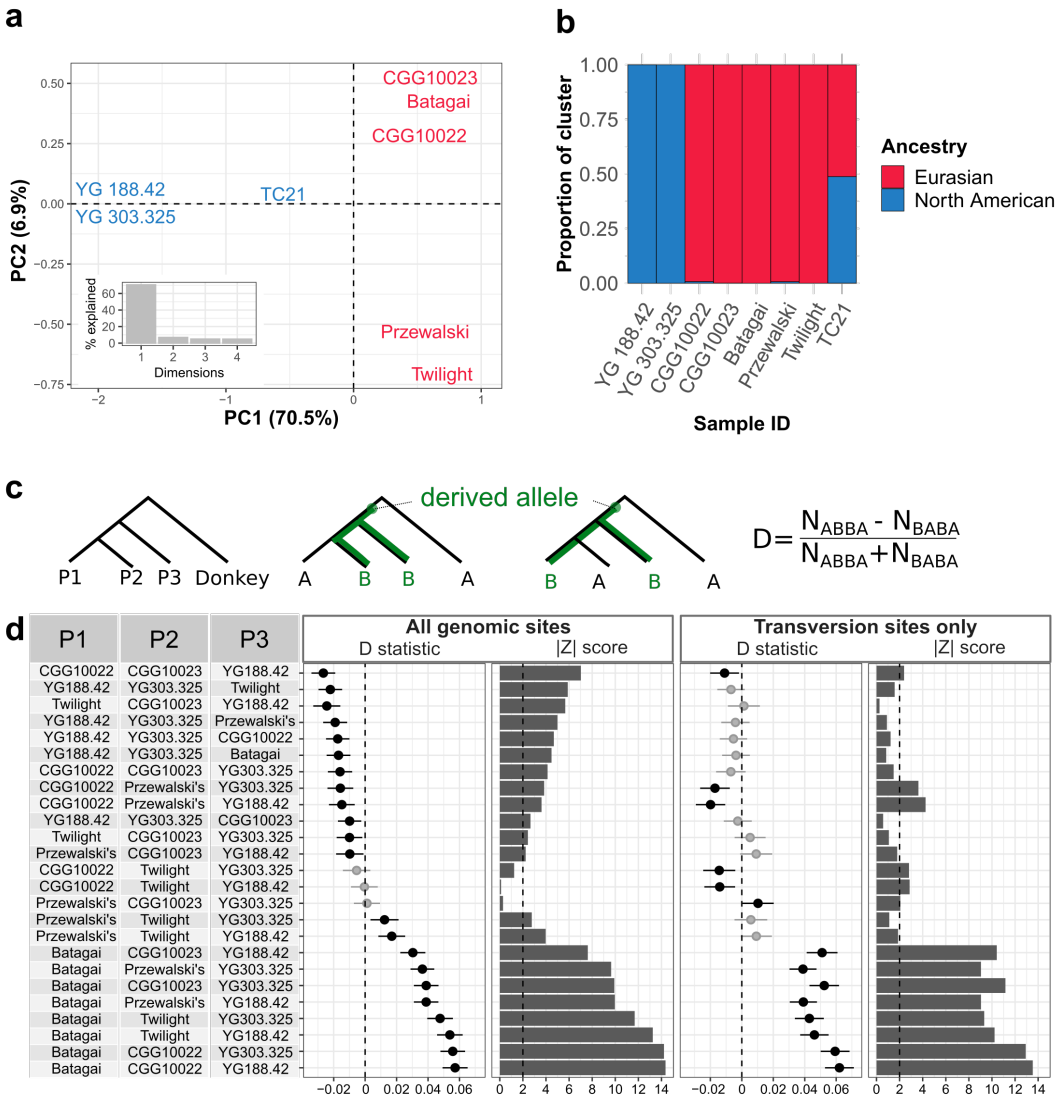

Figure 2. Geographic structure and gene flow between present-day and ancient caballine horses. a) Principal component (PC) analysis on the genotype likelihood covariance matrix. Percent of variation explained in the first four PCs is in the inset. **b)** Admixture clustering for K=2. Analyses for K=3 to K=5 are presented in Figure S5. Both (a) and (b) are estimated based on 6.7 million autosomal variable sites across eight horse genomes. In both (a) and (b) we excluded transitions to account for artifacts caused by post-mortem DNA damage. **c)** Schematic of the D-statistic test for the patterns of derived allele sharing. D<0 suggests gene flow between P1 and P3, while D>0 -- between P2 and P3. **d)** D-statistic tests on all autosomal genomic sites (0.82 million ABBA+BABA sites on average per P1, P2, P3 combination) and sites that exclude transitions (0.24 million ABBA+BABA sites on average per combination). Black and gray dots represent statistically significant and non-significant D-statistic values accordingly (estimated with Z-score using the jackKnife.R script from the ANGSD package). We excluded the low coverage TC21 horse from the D-statistic analysis.

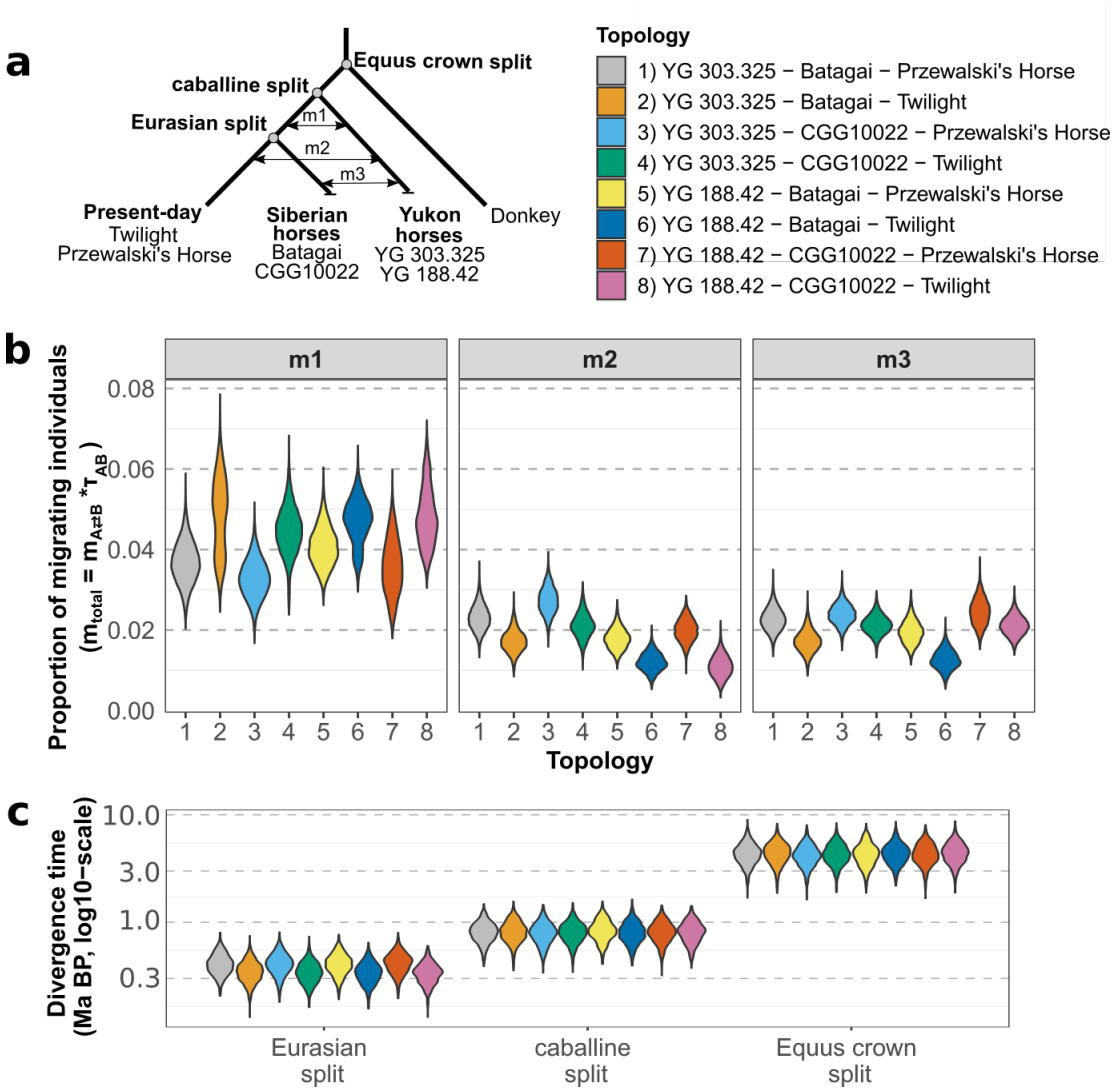

Figure 3. Demographic history of caballine horses reconstructed with G-PhoCS. a) A schematic of the tested demographic scenario. **b)** Estimates of the migration rates. Since forward and backward migration rates were nearly identical, we collapsed their values when plotting the graph. **c)** Divergence times. In the dataset of 4,215 putatively neutral autosomal loci, there were 62,647 polymorphisms in the analysis with the donkey outgroup, including 26,627 sites segregating only in caballine horses. In each G-PhoCS run we used one genome per population (a), producing eight possible configurations of the tree topology as noted on the figure legend.

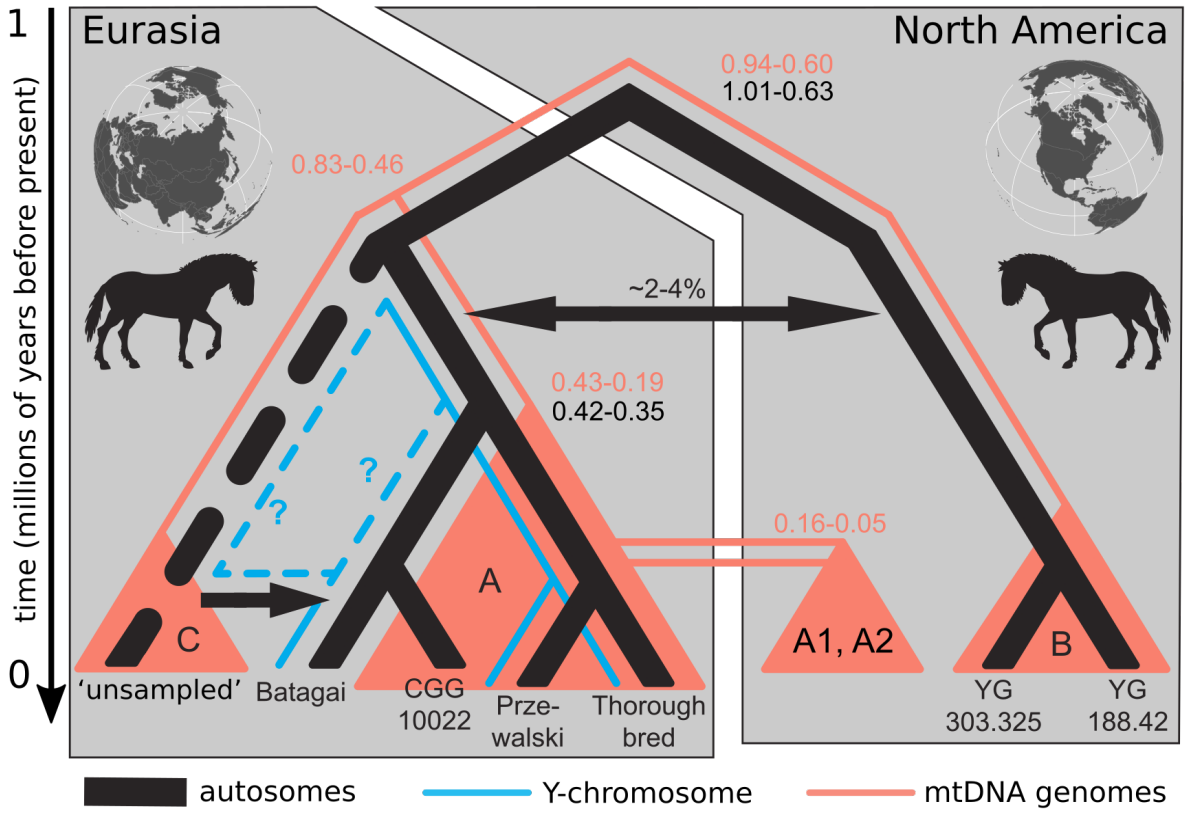

Figure 4. A tentative demographic scenario of caballine horse evolution. The Eurasian caballine horses (mitochondrial clades A and C) diverged from North American caballines (clade B) ~1.0-0.8 Ma via dispersal across the Bering Land Bridge (BLB). The evolution of horses continued on both continents in the presence of gene flow across the BLB. Cold conditions during the Middle Pleistocene and MIS 6 openings of the BLB facilitated local dispersals, leading to cross-continental migration of Eurasian horses back to North America (clades A1, A2). Many extinct lineages of horses are still undersampled (Librado & Orlando, 2020; Orlando, 2020), including an as yet unsampled population that diverged from other Eurasian horses prior to substantial gene flow between Eurasian and North American horses. We assume that this population with unsampled nuclear genome ancestry belongs to the divergent mitochondrial Clade C. We further hypothesize that this unsampled population contributed to the ancestry of the Batagai horse based on our D-statistic and D_{FOIL} results. This hypothesis may be further supported by the phylogenetic signal of Batagai's uniparental loci (Fages et al., 2019). All North American horses (clades A1, A2, B) and clade C went extinct by the early Holocene ~11 ka. Divergence times for mtDNA genomes are from the BEAST analyses, whereas those for the autosomes are from the G-PhoCS analyses. Arrows represent gene flow, as inferred by admixture analysis, D-statistics, D_{FOIL}, and G-PhoCS. Horse silhouettes are from phylopic.org: image by Mercedes Yrayzoz (vectorized by T. Michael Keesey and reused under a CC-BY-3.0 license. No changes were made other than vertically flipping the image on the left). Globes are plotted in R with rworldmap (South, 2011; Wickham, 2016).

The legends of all the supplementary figures and tables

Table S1. Metadata for the new caballine horse samples.

Table S2. Metadata for the previously published mitochondrial genomes used in the current study.

Table S3. Log output with ESS estimates for two BEAST analyses. (a) Trace for the BEAST that

included outgroup sequences, (b) trace for the BEAST CTMC run that did not include outgroups. For each (a) and (b) we ran two MCMC chains for 100M steps and skipped 25M initial steps as burn-in. The results are reported for the log-combined trace files excluding the burn-in.

Table S4. Time to the most recent common ancestor (TMRCA) for caballine horse mitochondrial clades estimated by time-calibrated BEAST analyses. Clade notations follow the main text figures 1b, 4. BEAST with outgroup included three non-caballine Equids: E. ovodovi, E. zebra and E. asinus. The continuous-time Markov chains (CTMC) (Minin & Suchard, 2008a, 2008b) inferred the timing and number of migrations across the Bering Land Bridge (Figure 1c) and included only caballine horses (Figure 1b, c).

Table S5. DNA degradation and sequencing error rates in ancient genomes. We estimated sequencing errors by comparing our data to the Twilight horse. We used the donkey genome to polarize the alleles, and assumed derived alleles shared between the "ideal" genome, Twilight, and our genomes to be sequencing errors. (see Appendix S2 for details). We used MapDamage v. 2.0.8 (Jónsson et al., 2013) to estimate frequency of G>A and C>T substitutions.

Table S6. Direction of gene flow between North American and Eurasian caballine horses. We used D FOIL to test a symmetric five-population model in a form of (((P1-YG188.42, P2-YG303.325), (P3,P4)), Outgroup - donkey). The schematic of the tree topology model and percent estimates for window counts are presented in the Figure S10.

Appendix S1. Supplemental material: radiocarbon dating protocol.

Appendix S2. Supplemental material: estimating error rates in whole-genome data.

Figure S1. Best-scoring maximum likelihood tree estimated in RAxML for 78 new and 112 previously-published horse mitochondrial genomes. Branch supports are tested using 500 bootstrap iterations (marked above the branches). Outgroups have been removed to improve readability.

Figure S2. Time-calibrated maximum clade credibility tree reconstructed with BEAST. Removal

of age-uninformative samples resulted in 176 sequences for analysis, including three outgroup samples: E. ovodovi, E. zebra and E. asinus. We calibrated radiocarbon ages of the Late Pleistocene samples reported in Table S1 using the IntCal13 (Reimer et al., 2013) and OxCal v4.2 (Ramsey, 2009), and assigned the median calibrated age of each sample as tip date prior. For samples published in Fages et al., 2019 that have stratigraphic information and not a C 14 isotope date, we assigned the mean of the tentative age interval as a tip prior (Table S2). We further calibrated BEAST using caballine/non-caballine Equus divergence estimate of 4.5–4.0 Ma.

Figure S3. Expanded version of the time-calibrated mitochondrial genealogy reconstructed in unrooted BEAST CTMC analysis and presented in the main text Figure 1b. Sample labels are listed next to the tree tips. Posterior supports are marked next to the tree nodes. To account for root polytomy between the Middle Pleistocene YG148.20/TC2, samples from New Siberian Islands, Russian Far East, and the rest of the tree, we constrained the monophyly on two nodes (marked with stars). Each sample was assigned to either Eurasian or North American continent based on their sampling location relative to the present-day Bering Strait. Gray bars represent 95% highest probability density (HPD) interval of node heights. We calibrated radiocarbon ages of the Late Pleistocene samples reported in Table S1 using the IntCal13 curve (Reimer et al., 2013) and OxCal v4.2 (Ramsey, 2009), and assigned the median calibrated age of each sample as prior information. If a sample was dated using stratigraphic information (Fages et al., 2019), we assigned the mean of the tentative age interval as a date prior (Table S2). Samples without age information were excluded from the analysis.

Figure S4. Ancient nuclear DNA damage patterns for YG188.42. The lack of a deamination signal is due to the USER-enzyme pre-treatment of YG188.42 DNA extracts.

Figure S5. Ancient nuclear DNA damage patterns for YG303.325.

Figure S6. Depth of coverage for nuclear genomes of North American caballine horses YG188.42 and YG303.325. We trimmed adapters from paired-end reads and filtered the remaining reads for quality, merged overlapping reads and aligned both merged and unmerged reads to the EquCab2.0 E. caballus reference genome with BWA aln and seed disabled. We estimated coverage of each genome with bedtools v2.25.0.

Figure S7. Genome-wide type-specific error rates relative to the Twilight horse. North American caballine horse genomes sequenced in this study are marked with a star. Overall per genome error rates are reported in the Table S4. Method behind the error rate estimation is described in the Appendix S2. We omitted the TC21 from the plot due to >50% error rate in this 1-fold coverage sample.

Figure S8. Admixture analysis of genotype likelihoods with PCAngsd and K=3 to K=5. Estimates are based on a total of 6,690,973 autosomal SNP sites across eight horse samples. We limited the analysis only to nucleotide transversions to exclude possible presence of false positive SNP calls due DNA post-mortem deamination.

Figure S9. D-statistic test for admixture in sex chromsomes of sampled horse linages. The test in configuration (((P1,P2),P3), outgroup donkey) estimates whether the sample P3 shares a greater proportion of alleles with P1 or P2. Negative D-statistics suggest gene flow between P1 and P3, while positive statistics suggest gene flow between P2 and P3. We tested all sites on the X-chromosome (19,213 ABBA+BABA sites on average across all combinations of P1, P2, and P3). We repeated the analysis removing potential transitions to account for sites resulting from DNA damage (5,918 ABBA+BABA sites on average across all combinations of P1, P2, and P3).

Figure S10. Gene flow among caballine horses detected with DFOIL. The inset shows the five-population model tested for treeness. We used the autosomal dataset that excluded transitions and split each genome into 11,225 200kb non-overlapping windows. We run the test four times with P3 and P4 sample combinations as shown on the figure legend. The ticks and arrows on the x-axis indicate the direction of the gene flow detected by D FOIL for each run. Raw window counts are presented in the Table S4.

Figure S11. Selection of putatively-neutral nuclear loci for G-PhoCS analysis. The "filter" coordinates correspond to exons, genome assembly gaps, tandem and simple repeats, paralogs, CpG islands, and clustered SNPs (genomic intervals with two or more clustered SNPs within 5, 10, and 50 bp windows).

Figure S12. Maximum likelihood phylogenetic analysis of seven high coverage genomes built on 4,215 putatively-neutral nuclear 1kb loci. To verify that the ascertained neutral loci data set follows the expected phylogenetic relationships between the samples, we concatenated all loci into an alignment of seven genomes, which included the donkey as outgroup, and reconstructed phylogeny in Geneious v. 2020.0.5 (https://www.geneious.com) using RAxML v.8.2.11 plugin (Stamatakis, 2014). There were 62,647 variable sites in the alignment, including 26,627 if the outgroup donkey genome was excluded. We used RAxML with the GTRCAT model of nucleotide substitution and assessed branch support confidence with 500 rapid bootstrap-replicates.

Figure S13. Changes in effective population size of caballine horses (a) reconstructed with G-PhoCS using the topology model (b). We used G-PhoCS analysis on the dataset of 4,215 putatively neutral autosomal loci containing 62,647 polymorphisms with the donkey outgroup, including 26,627 polymorphisms segregating only in caballine horses. In each G-PhoCS run we used one genome per population as noted in the figure legend.

Figure S14. Genome-wide mutation rate (A) and generation time (B) used for re-scaling of G-PhoCS coalescent estimates into calendar years. To account for uncertainty in estimates of the mutation rates and variable generation time we created gamma-distribution for each of these parameters with R function 'dgamma'. To calibrate G-PhoCS runs we randomly sampled values from these distributions.

Figure S15. PSMC demographic inference for ancient caballine horses. We ran PSMC on quality-filtered genome alignments to the EquCab2.0 reference. We re-scaled Θ for each ancient sample so that their demographic history reconstruction starts at the time of their calibrated radiocarbon age (see Table 1).

Supplemental Information for:

Ancient horse genomes reveal the timing and extent of dispersals across the Bering Land Bridge

Alisa O. Vershinina, Peter D. Heintzman, Duane G. Froese, Grant Zazula, Molly Cassatt-Johnstone, Love Dalén, Clio Der Sarkissian, Shelby G. Dunn, Luca Ermini, Cristina Gamba, Pamela Groves, Joshua D. Kapp, Daniel H. Mann, Andaine Seguin-Orlando, John Southon, Mathias Stiller, Matthew J. Wooller, Gennady Baryshnikov, Dmitry Gimranov, Eric Scott, Elizabeth Hall, Susan Hewitson, Irina Kirillova, Pavel Kosintsev, Fedor Shidlovsky, Hao-Wen Tong, Mikhail P. Tiunov, Sergey Vartanyan, Ludovic Orlando, Russell Corbett-Detig, Ross D. MacPhee, Beth Shapiro

Table of Contents:

Figure S4. Ancient nuclear DNA damage patterns for YG188.42.	
Figure S5. Ancient nuclear DNA damage patterns for YG303.325.	
Figure S6. Depth of coverage for nuclear genomes of North American caballine horses YG188.42 and YG303.325.	
Figure S7. Genome-wide type-specific error rates relative to the Twilight horse.	
Figure S8. Admixture analysis of genotype likelihoods with PCAngsd and K=3 to K=5.	
Figure S9. D-statistic test for admixture in sex chromsomes of sampled horse linages.	
Figure S10. Gene flow among caballine horses detected with D_{FOIL}.	
Figure S11. Selection of putatively-neutral nuclear loci for G-PhoCS analysis.	
Figure S12. Maximum likelihood phylogenetic analysis of seven high coverage genomes built on 4,215 putatively-neutral nuclear 1kb loci.	
Figure S13. Changes in effective population size of caballine horses (a) reconstructed with G-PhoCS using the topology model (b).	
Figure S14. Genome-wide mutation rate (A) and generation time (B) used for re-scaling of G-PhoCS coalescent estimates into calendar years.	
Figure S15. PSMC demographic inference for ancient caballine horses.	

Table S1. Metadata for the new caballine horse samples.

https://docs.google.com/spreadsheets/d/1AAqYqKrcUb8aaIu-pbATzqfwWUw5C8sG-6SIMRr5BhA/edit?usp=sharing

Table S2. Metadata for the previously published mitochondrial genomes used in the current study.

https://docs.google.com/spreadsheets/d/1XAE1ONTIAsrsKJ_TbIi9_TgeVquzfJs3B9Ov5haSlqU/edit#gid=655037552

Table S3. Log output with ESS estimates for two BEAST analyses. (a) Trace for the BEAST that included outgroup sequences, **(b)** trace for the BEAST CTMC run that did not include outgroups. For each (a) and (b) we ran two MCMC chains for 100M steps and skipped 25M initial steps as burn-in. The results are reported for the log-combined trace files excluding the burn-in.

https://drive.google.com/file/d/1w_1MNZ1tibbUOXEg6mWiAFwGiRWmY4Zp/view?usp=sharing

Table S4. Time to the most recent common ancestor (TMRCA) for caballine horse mitochondrial clades estimated by time-calibrated BEAST analyses. Clade notations follow the main text figures 1b, 4. BEAST with outgroup included three non-caballine Equids: *E. ovodovi, E. zebra* and *E. asinus*. The continuous-time Markov chains (CTMC) (Minin & Suchard, 2008a, 2008b) inferred the timing and number of migrations across the Bering Land Bridge (Figure 1c) and included only caballine horses (Figure 1b, c).

Clade	**BEAST with outgroup**		**BEAST CTMC**	
	TMRCA median	**95% HPD**	**TMRCA median**	**95% HPD**
((A,C),B), outgroup	4.17E6	[3.88E6,4.47E6]	-	-
(A,C),B	7.22E5	[5.89E5,8.59E5]	8.45E5	[7.59E5,9.4E5]
A,C	6.14E5	[4.56E5,7.75E5]	7.11E5	[5.98E5,8.32E5]
A	2.36E5	[1.86E5,3.03E5]	3.75E5	[3.05E5,4.47E5]
A1	6.06E4	[5.16E4,7.31E4]	1.36E5	[1.17E5,1.6E5]
A2	3.89E4	[4.29E4,5.02E4]	1.17E5	[9.32E4,1.44E5]
B	6.35E5	[5.19E5,7.54E5]	7.6E5	[7.23E5,8.05E5]
C	2.87E5	[1.91E5,3.89E5]	3.98E5	[3.29E5,4.71E5]

Table S5. DNA degradation and sequencing error rates in ancient genomes.

We estimated sequencing errors by comparing our data to the Twilight horse. We used the donkey genome to polarize the alleles, and assumed derived alleles shared between the "ideal" genome, Twilight, and our genomes to be sequencing errors. (see Appendix S2 for details). We used MapDamage v. 2.0.8 (Jónsson et al., 2013) to estimate frequency of G>A and C>T substitutions.

Specimen ID	Overall Error Rate (%)	Frequency of G>A at 3p' ends (1st position)	Frequency of C>T at 5p' ends (1st position)	UDG treatment	Reference
Batagai	0.02	0.01	0.02	no	(Librado et al., 2015)
CGG10022	0.1	0.02	0.03	no	(Schubert et al., 2014)
CGG10023	0.14	0.03	0.03	no	(Schubert et al., 2014)
YG303.325	0.08	0.13	0.14	no	This study
YG188.42/ YT03-40	0.09	0.02	0.01	yes	This study
YG148.20/ TC21	52.04	0.28	0.33	no	(Orlando et al., 2013)

Table S6. Direction of gene flow between North American and Eurasian caballine horses.

We used D_{FOIL} to test a symmetric five-population model in a form of (((P1-YG188.42, P2-YG303.325), (P3,P4)), Outgroup - donkey). The schematic of the tree topology model and percent estimates for window counts are presented in the Figure S10.

		Count of non-overlapping 200kb windows			
	P3:	**Batagai**	**Batagai**	**CGG10022**	**Przewalski**
Gene flow direction	**P4:**	**Przewalski**	**Twilight**	**CGG10023**	**Twilight**
P12<->P3		229	182	860	23
P12<->P4		25	24	1	78
P1->P3		51	41	220	12
P1->P4		8	11	0	32
P2->P3		2	7	8	4
P2->P4		0	1	0	4
P3->P1		600	535	64	193
P3->P2		6	2	0	0
P4->P1		98	132	3	302
P4->P2		1	2	0	4

Appendix S1. Supplemental material: radiocarbon dating protocol.

Aliquots of ~200 mg of cortical bone were cleaned mechanically and sampled with a high speed rotary Dremel tool, crushed to mm-sized chips, and placed in 13 mm culture tubes with vented caps. To remove contaminating conservation materials, we sonicated samples in acetone, methanol and ultrapure Milli-Q (MQ) water (1 hr each) in a water bath cooled to ~ 45 °C, well below the melting point of collagen. The bone was decalcified overnight with 1N HCl at room temperature, using an amount of acid sufficient enough to dissolve all of the bone mineral if no collagen was present. The samples were then sonicated in acidified methanol (5% 0.2N HCl) and in MQ water (1 hr each) to remove any contaminating polymerized shellac that might be present. The water was then pipetted off and replaced with 3 mL of 0.01N HCl and the samples were gelatinized at 60 °C overnight. The resulting gelatin was pipetted into precleaned Vivaspin 15 Turbo ultrafiltration devices, avoiding transferring any solids, and centrifuged to select the >30 kDa molecular weight fraction. The >30 kDa retentate was then pipetted into preweighed thick walled 13 mm culture tubes, frozen, and dried overnight in a vacuum centrifuge.

The tubes containing the lyophilized collagen were reweighed to determine collagen yields, and 2 mg aliquots for ^{14}C dating were transferred to preweighed 6 mm quartz combustion tubes. CuO oxidizer and silver wire getter were added and the tubes were sealed under vacuum and combusted at 900 °C. CO_2 graphitization and ^{14}C measurement by AMS were carried out at the Keck AMS laboratory at University of California Irvine. Aliquots containing 0.7 mg of collagen were wrapped in 5 x 9 mm tin foil cups and placed in the carousel of a Fisons NA1500NC elemental analyzer interfaced to a Finnigan Delta Plus isotope ratio mass spectrometer for elemental analyses and $\delta^{13}C$ and $\delta^{15}N$ measurements.

Appendix S2. Supplemental material: estimating error rates in whole-genome data.

We used the method described in Orlando et al. (2013) to estimate overall and type-specific error rates in our autosomal dataset. Briefly, the method compares an ideal genome, the tested genome mapped to a reference, and an outgroup. The "ideal" genome is usually a high quality dataset assumed to be error free while the outgroup is used to polarize the alleles. We expect each tested sample and the "ideal" genome to have the same number of derived alleles. Excess derived alleles are attributed to errors.

We estimated overall error rates in ANGSD v0.918 (Korneliussen et al., 2014) using the '-doAncError 2' flag. We used Twilight as an "ideal" horse genome and the donkey genome as an "ancestral" individual for an outgroup. Both genomes were converted from BAM alignments into fasta files using '-doFasta', as implemented in ANGSD. For all analysed genomes, we sampled a random base at each position, including bases with coverage in all three individuals.

The estimated error rates (Table S4, Figure S7) demonstrate an excess of cytosine deamination in the ancient samples as expected due to post-mortem DNA damage. Interestingly, two ancient North American genomes have similar overall error rates despite the fact that one genome, YG188.42, was treated with Uracil-DNA Glycosylase (UDG) to remove the effect of deamination. The UDG treatment, however, resulted in a lower type-specific error rate in YG188.42 compared to YG303.325 (Figure S7). The highest error rates are observed in previously published TC21 and CGG10023 individuals which are the lowest coverage genomes in our panel (Table S4).

Batagai has a lower error rate (0.2%) compared to CGG10022 and CGG10023 (~0.1%) (Table S4). The Batagai genome BAM alignment was published with a strict quality filter cut-off (mapQ≥30, baseQ≥20) which is potentially responsible for this difference in error rates (see section 2.7 in Librado et al., 2015). Since D_{FOIL} and D-statistic tests are affected by individual sequencing error rates (Rasmussen et al., 2011), we used mapQ≥30 and baseQ≥30 for all genomes in these analyses to adjust for this difference. Furthermore, assuming a maximum proportion of 0.001 sites affected by errors, in our D-statistic analysis of ~0.8M ABBA+BABA allele patterns per comparison

one would expect less than 1 shared allele to result from deamination or sequencing error.

Figure S1. Best-scoring maximum likelihood tree estimated in RAxML for 78 new and 112 previously-published horse mitochondrial genomes. Branch supports are tested using 500 bootstrap iterations (marked above the branches). Outgroups have been removed to improve readability.

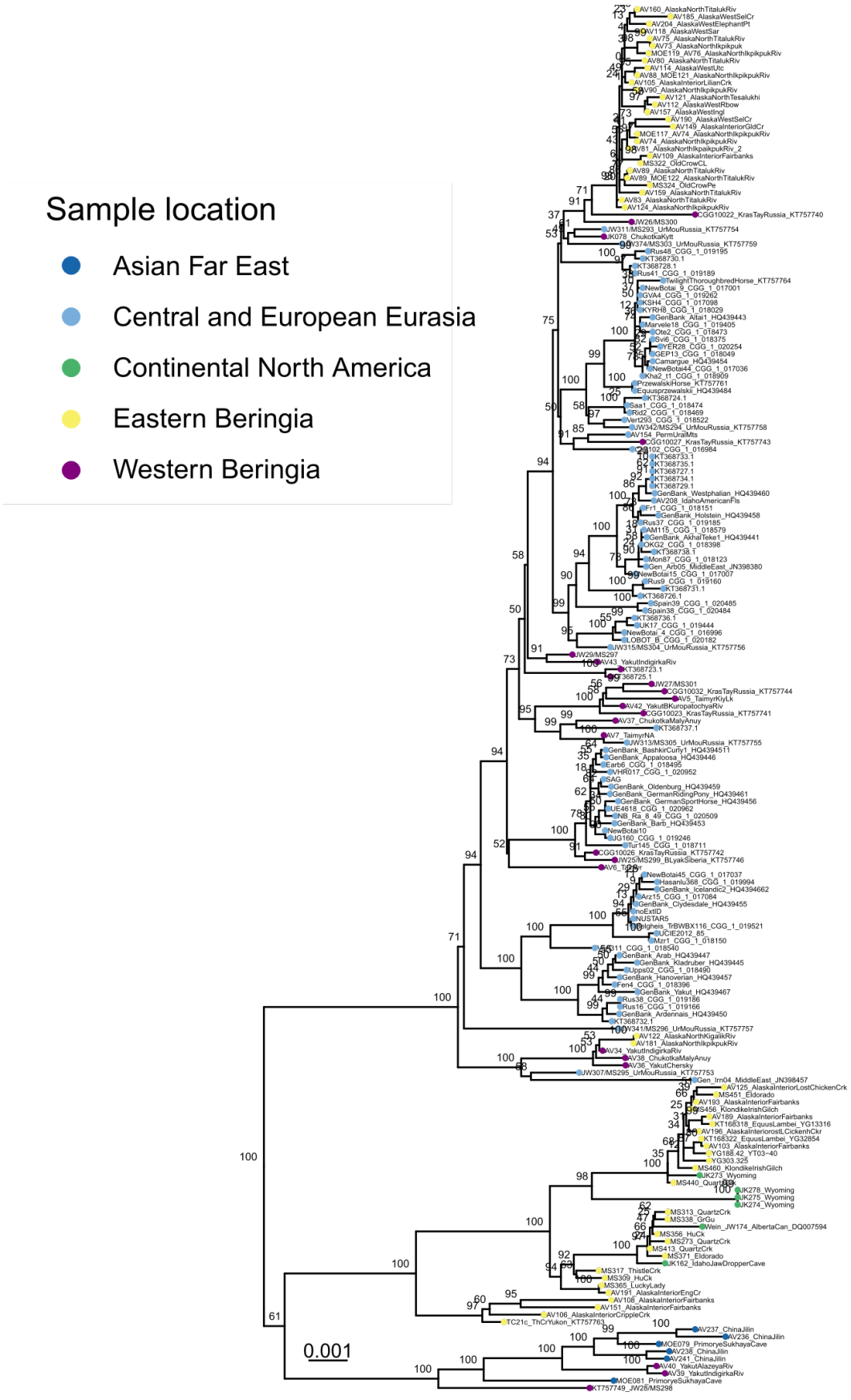
Sample location
Asian Far East
Central and European Eurasia
Continental North America
Eastern Beringia
Western Beringia
0.001

Figure S2. Time-calibrated maximum clade credibility tree reconstructed with BEAST. Removal of age-uninformative samples resulted in 176 sequences for analysis, including three outgroup samples: *E. ovodovi, E. zebra* and *E. asinus*. We calibrated radiocarbon ages of the Late Pleistocene samples reported in Table S1 using the IntCal13 curve (Reimer et al., 2013) and OxCal v4.2 (Ramsey, 2009), and assigned the median calibrated age of each sample as tip date prior. For samples published in (Fages et al., 2019) that have stratigraphic information and not a C^{14} isotope date, we assigned the mean of the tentative age interval as a tip prior (Table S2). We further

calibrated BEAST using caballine/non-caballine Equus divergence estimate of 4.5–4.0 Ma.

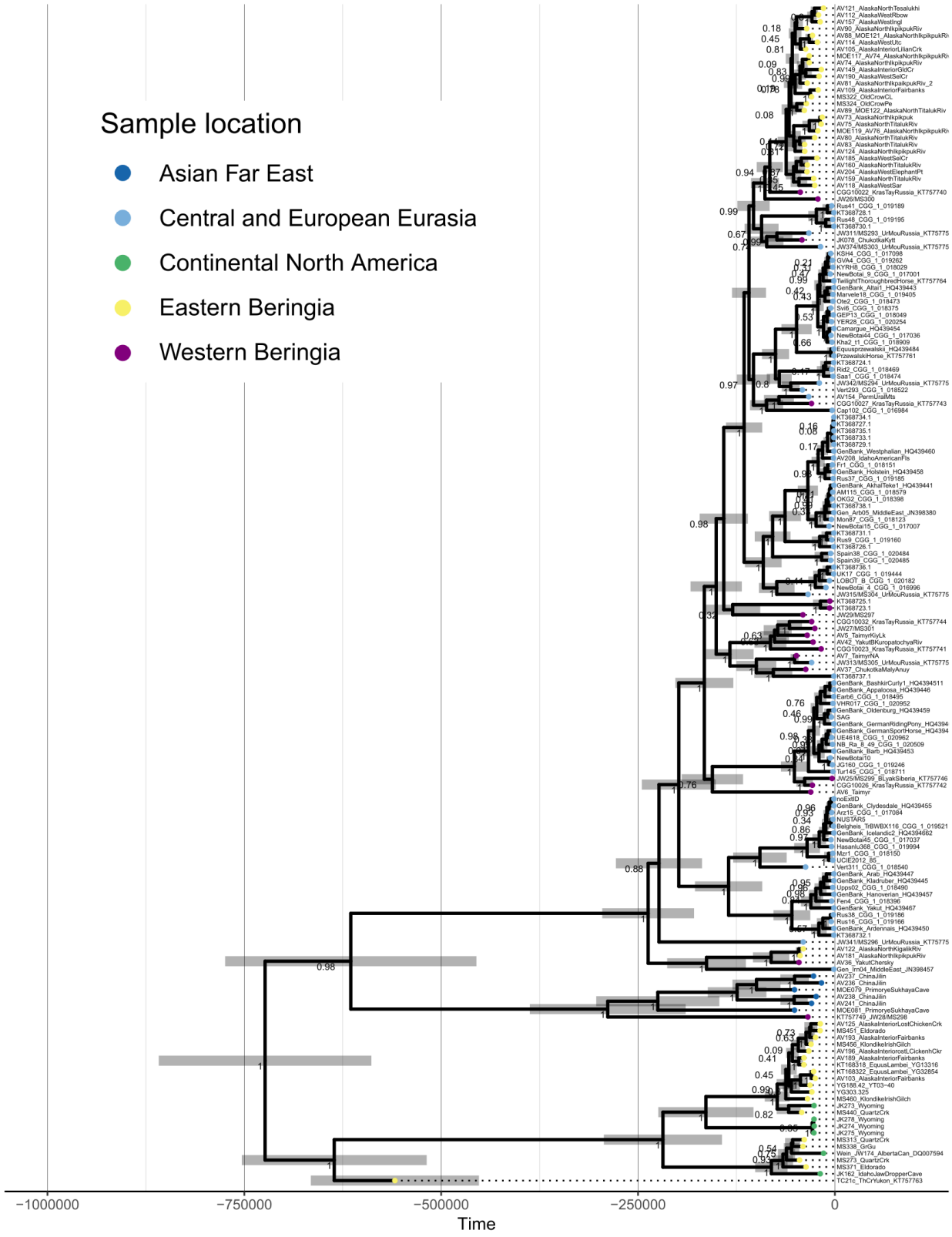
Sample location
Asian Far East
Central and European Eurasia
Continental North America
Eastern Beringia
Western Beringia
−1000000
−750000
−500000
−250000
0
Time

Figure S3. Expanded version of the time-calibrated mitochondrial genealogy reconstructed in unrooted BEAST CTMC analysis and presented in the main text Figure 1b. Sample labels are listed next to the tree tips. Posterior supports are marked next to the tree nodes. To account for root polytomy between the Middle Pleistocene YG148.20/TC2, samples from New Siberian Islands, Russian Far East, and the rest of the tree, we constrained the monophyly on two nodes (marked with stars). Each sample was assigned to either Eurasian or North American continent based on their sampling location relative to the present-day Bering Strait. Gray bars represent 95% highest probability density (HPD) interval of node heights. We calibrated radiocarbon ages of the Late Pleistocene samples reported in Table S1 using the IntCal13 curve (Reimer et al., 2013) and OxCal v4.2 (Ramsey, 2009), and assigned the median calibrated age of each sample as prior information. If a sample was dated using stratigraphic information (Fages et al., 2019), we assigned the mean of the tentative age interval as a date prior (Table S2). Samples without age information were excluded from the analysis.

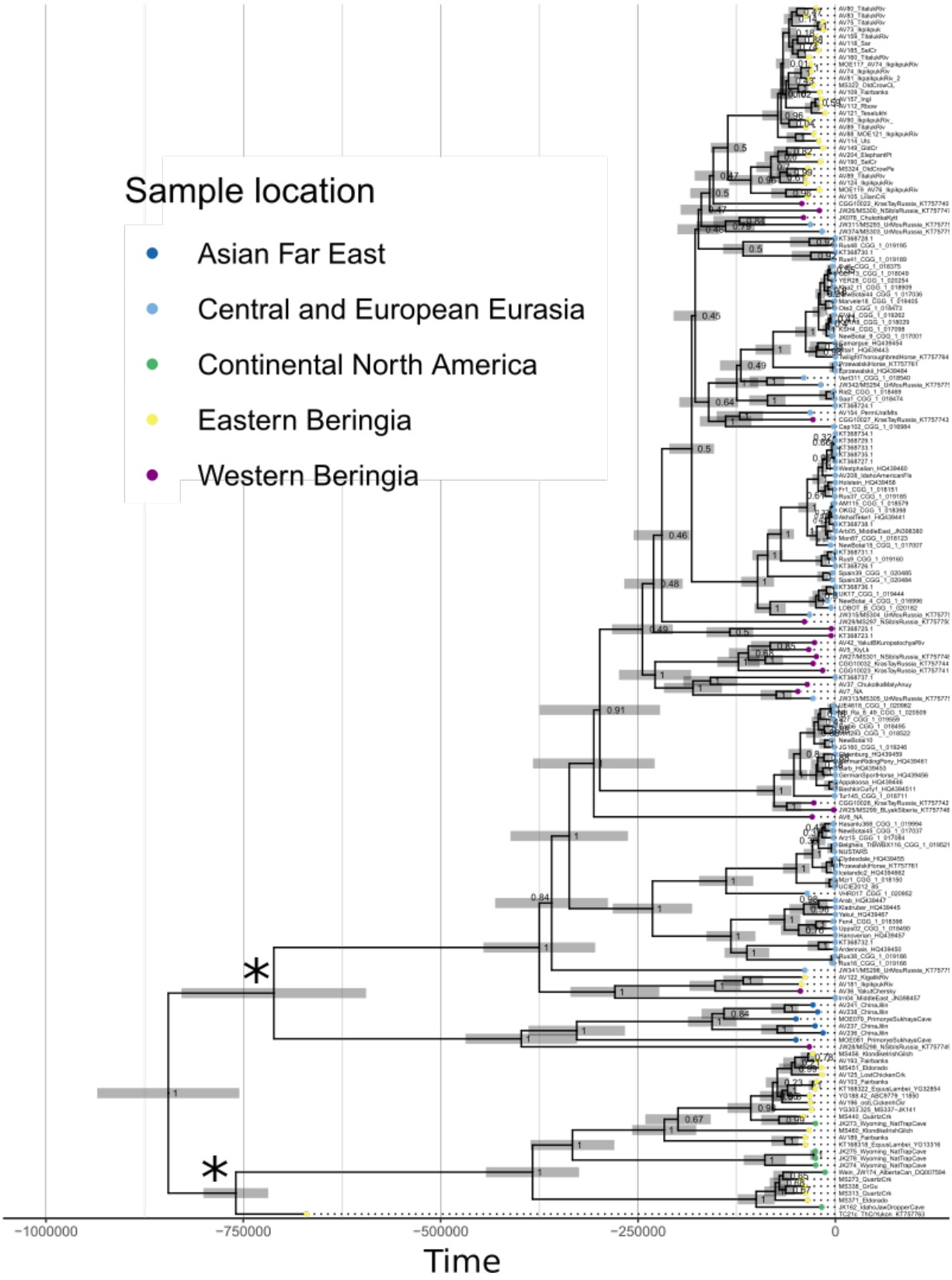
Sample location
Asian Far East
Central and European Eurasia
Continental North America
Eastern Beringia
Western Beringia
−1000000
−750000
−500000
−250000
0
Time

Figure S4. Ancient nuclear DNA damage patterns for YG188.42.

The lack of a deamination signal is due to the USER-enzyme pre-treatment of YG188.42 DNA extracts.

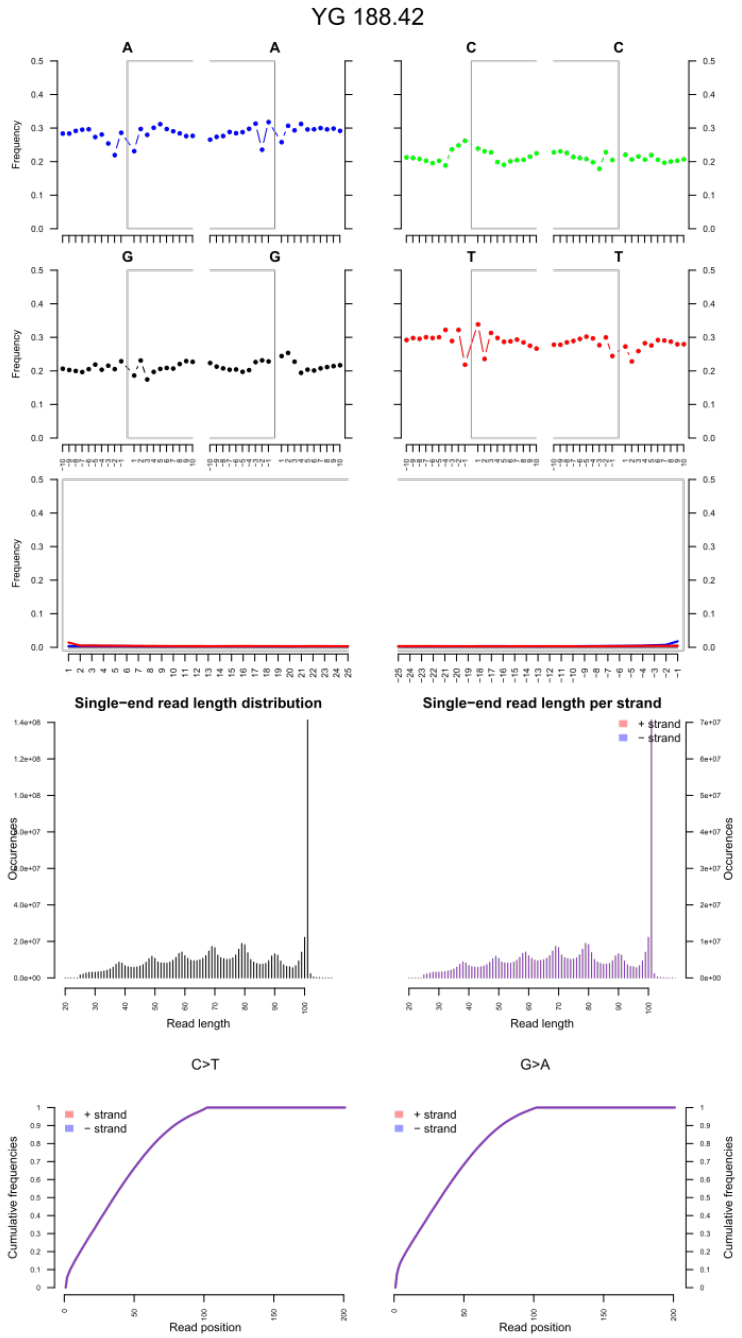

Figure S5. Ancient nuclear DNA damage patterns for YG303.325.

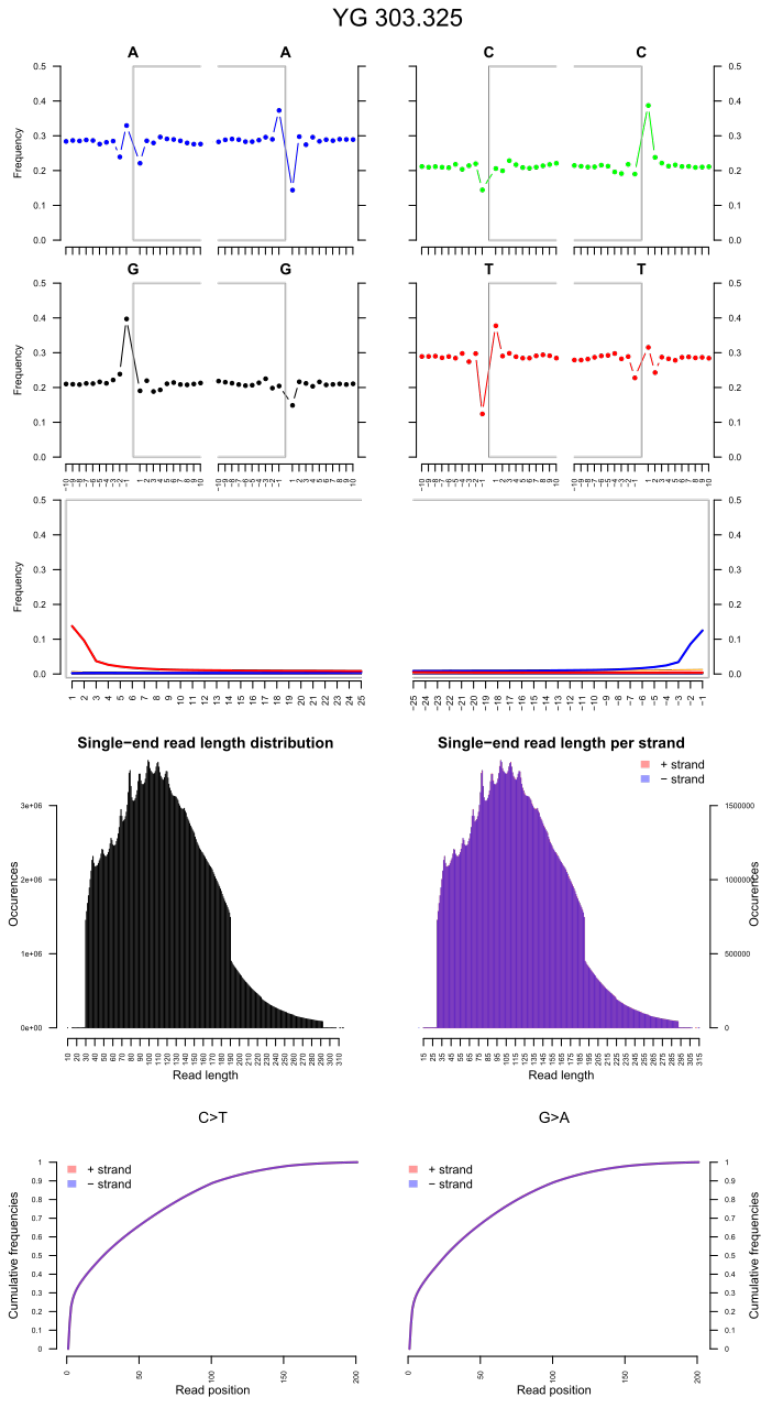

Figure S6. Depth of coverage for nuclear genomes of North American caballine horses YG188.42 and YG303.325.

We trimmed adapters from paired-end reads and filtered the remaining reads for quality, merged overlapping reads and aligned both merged and unmerged reads to the EquCab2.0 *E. caballus* reference genome with BWA aln and seed disabled. We estimated coverage of each genome with bedtools v2.25.0.

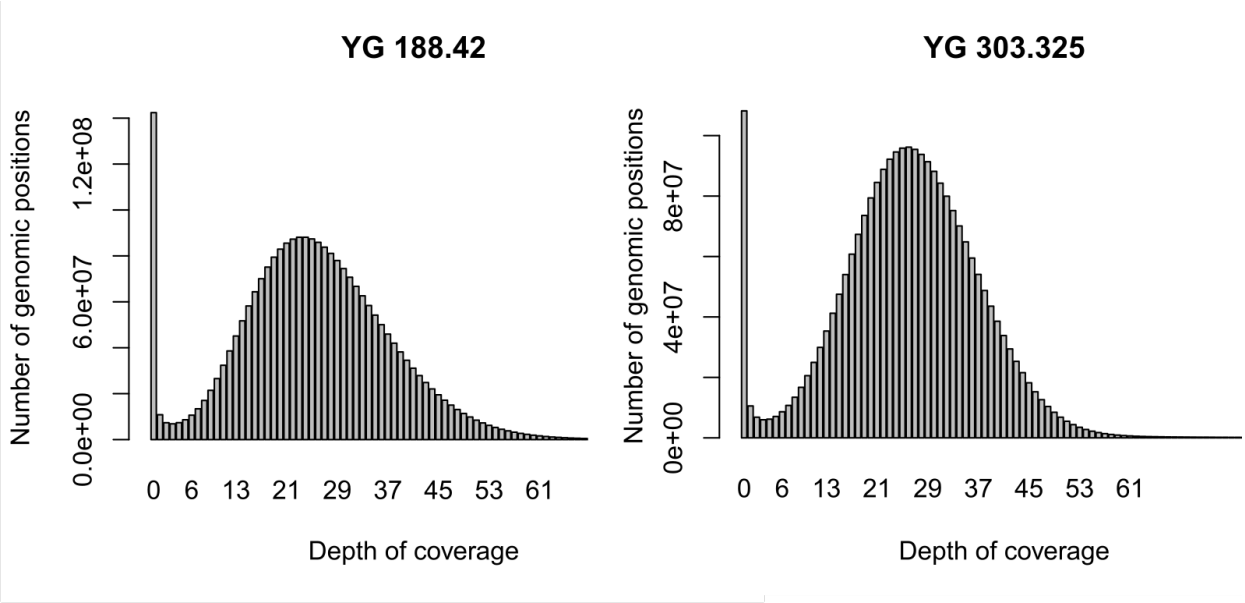

Figure S7. Genome-wide type-specific error rates relative to the Twilight horse. North American caballine horse genomes sequenced in this study are marked with a star. Overall per genome error rates are reported in the Table S4. Method behind the error rate estimation is described in the Appendix S2. We omitted the TC21 from the plot due to >50% error rate in this 1-fold coverage sample.

Figure S8. Admixture analysis of genotype likelihoods with PCAngsd and K=3 to K=5. Estimates are based on a total of 6,690,973 autosomal SNP sites across eight horse samples. We limited the analysis only to nucleotide transversions to exclude possible presence of false positive SNP calls due DNA post-mortem deamination.

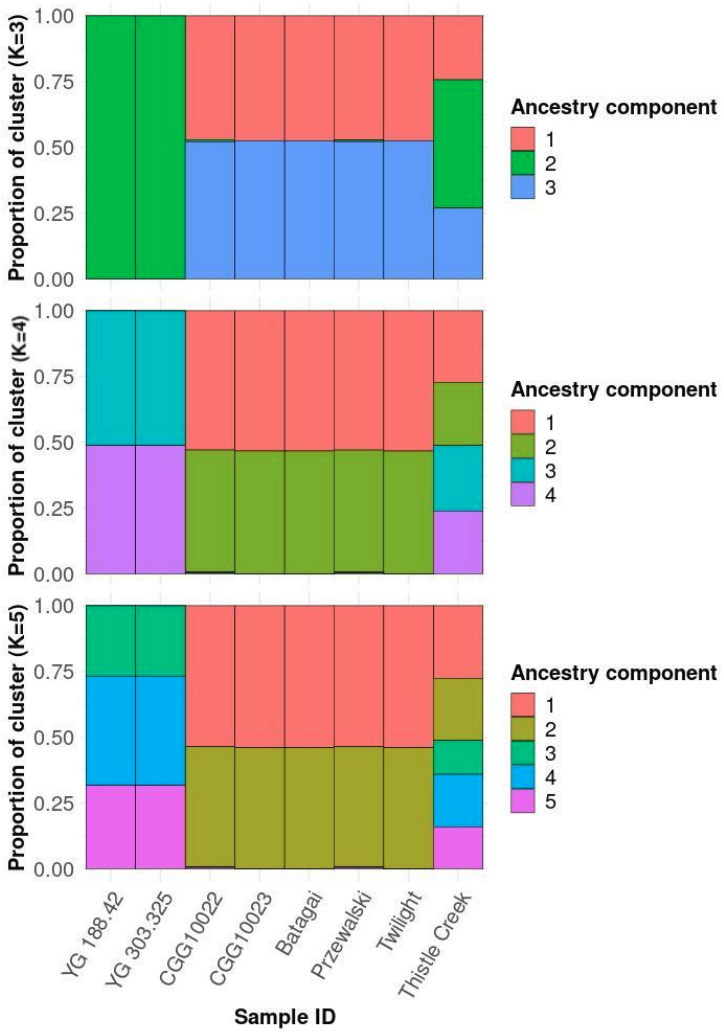

Figure S9. D-statistic test for admixture in sex chromsomes of sampled horse linages. The test in configuration (((P1,P2),P3), outgroup donkey) estimates whether the sample P3 shares a greater proportion of alleles with P1 or P2. Negative D-statistics suggest gene flow between P1 and P3, while positive statistics suggest gene flow between P2 and P3. We tested all sites on the X-chromosome (19,213 ABBA+BABA sites on average across all combinations of P1, P2, and P3). We repeated the analysis removing potential transitions to account for sites resulting from DNA damage (5,918 ABBA+BABA sites on average across all combinations of P1, P2, and P3).

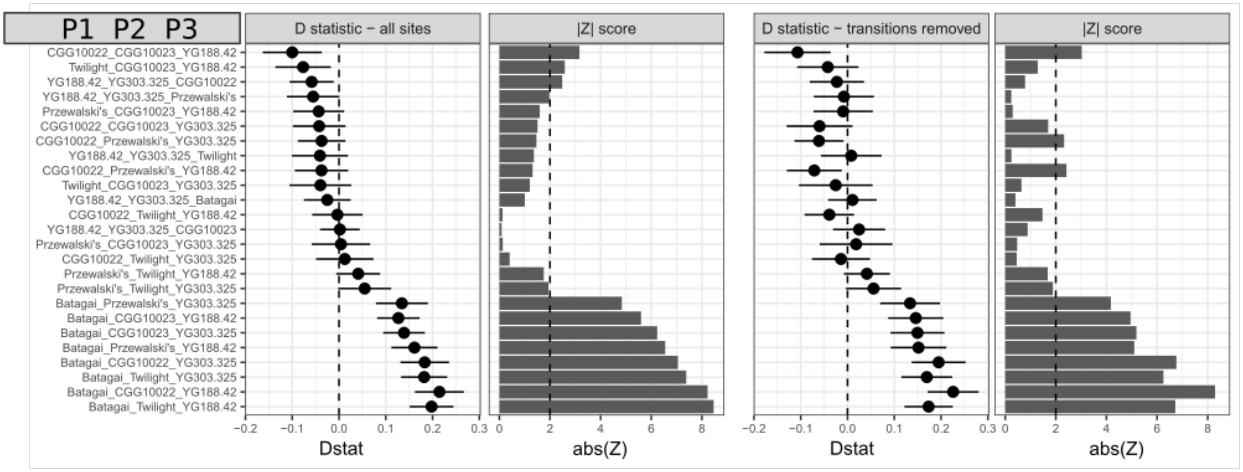

Figure S10. Gene flow among caballine horses detected with D_{FOIL}. The inset shows the five-population model tested for treeness. We used the autosomal dataset that excluded transitions and split each genome into 11,225 200kb non-overlapping windows. We run the test four times with P3 and P4 sample combinations as shown on the figure legend. The ticks and arrows on the x-axis indicate the direction of the gene flow detected by D_{FOIL} for each run. Raw window counts are presented in the Table S4.

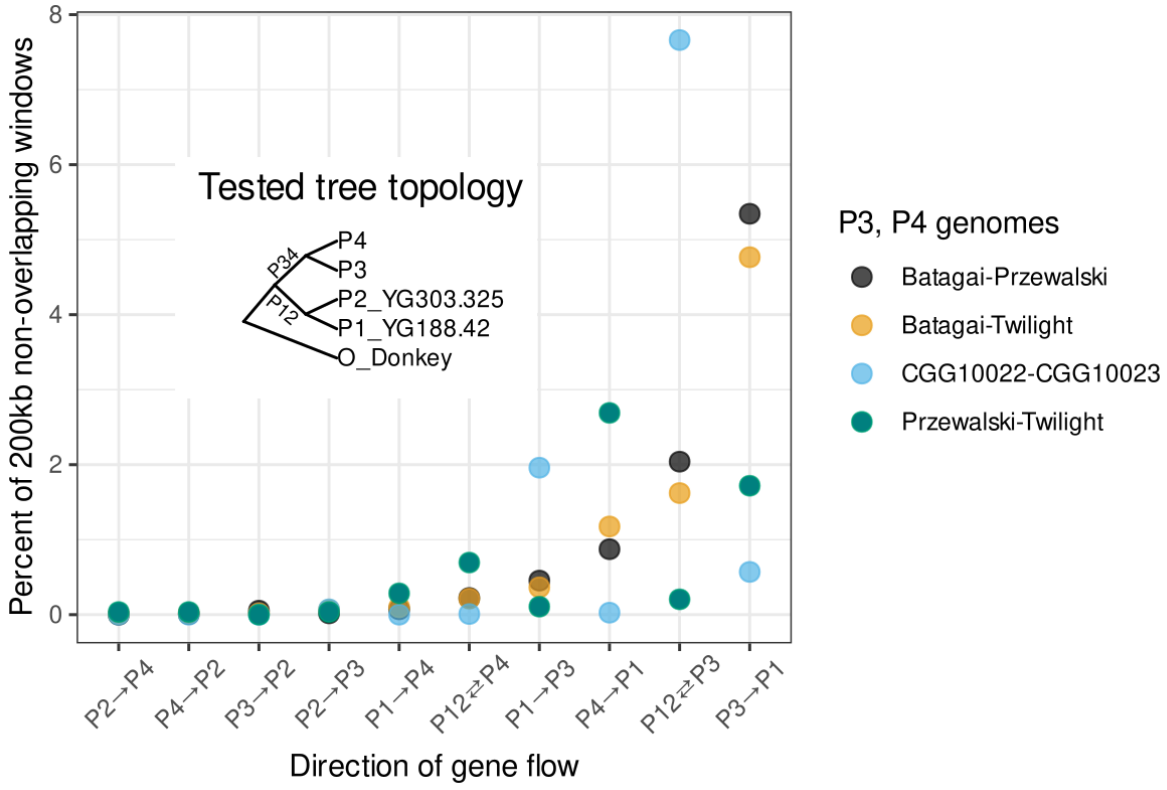

Figure S11. Selection of putatively-neutral nuclear loci for G-PhoCS analysis.

The “filter” coordinates correspond to exons, genome assembly gaps, tandem and simple repeats, paralogs, CpG islands, and clustered SNPs (genomic intervals with two or more clustered SNPs within 5, 10, and 50 bp windows).

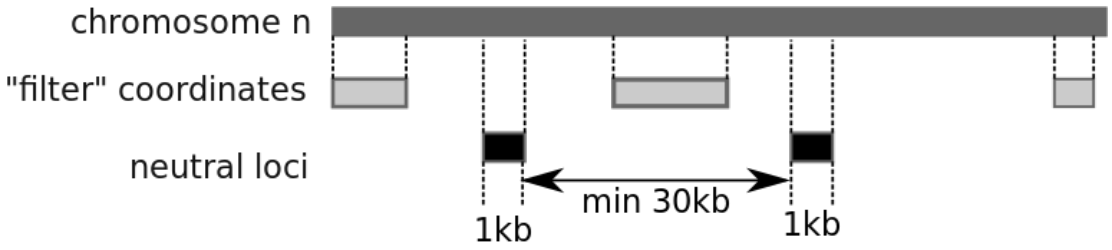

Figure S12. Maximum likelihood phylogenetic analysis of seven high coverage genomes built on 4,215 putatively-neutral nuclear 1kb loci.

To verify that the ascertained neutral loci data set follows the expected phylogenetic relationships between the samples, we concatenated all loci into an alignment of seven genomes, which included the donkey as outgroup, and reconstructed phylogeny in Geneious v. 2020.0.5 (https://www.geneious.com) using RAxML v.8.2.11 plugin (Stamatakis, 2014). There were 62,647 variable sites in the alignment, including 26,627 if the outgroup donkey genome was excluded. We used RAxML with the GTRCAT model of nucleotide substitution and assessed branch support confidence with 500 rapid bootstrap-replicates.

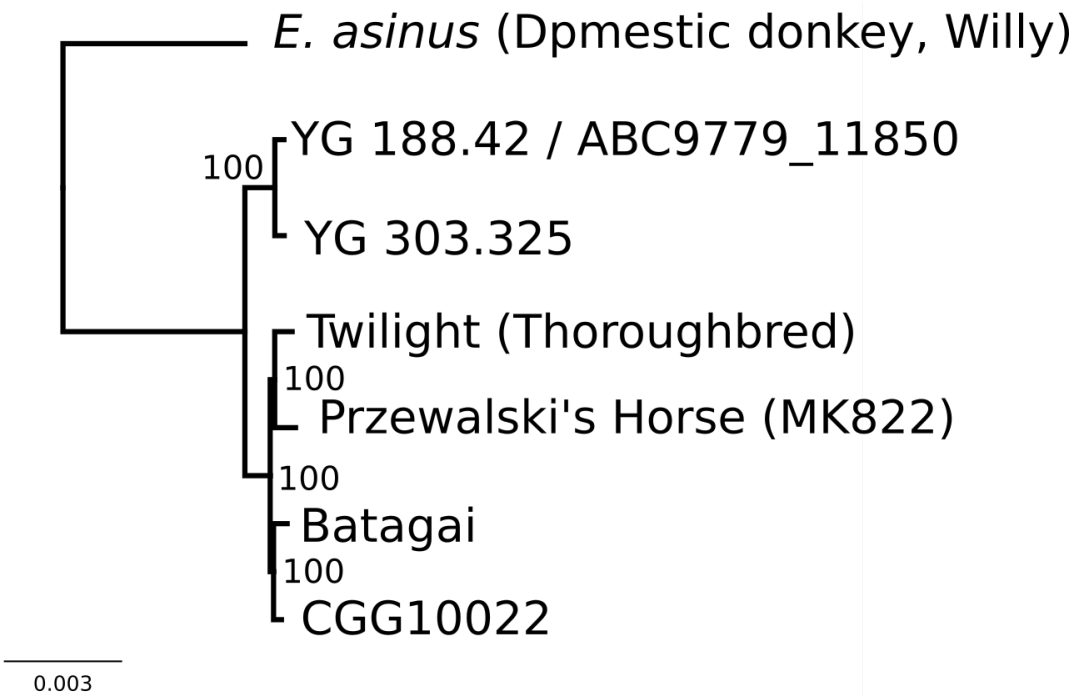

Figure S13. Changes in effective population size of caballine horses (a) reconstructed with G-PhoCS using the topology model (b).

We used G-PhoCS analysis on the dataset of 4,215 putatively neutral autosomal loci containing 62,647 polymorphisms with the donkey outgroup, including 26,627 polymorphisms segregating only in caballine horses. In each G-PhoCS run we used one genome per population as noted in the figure legend.

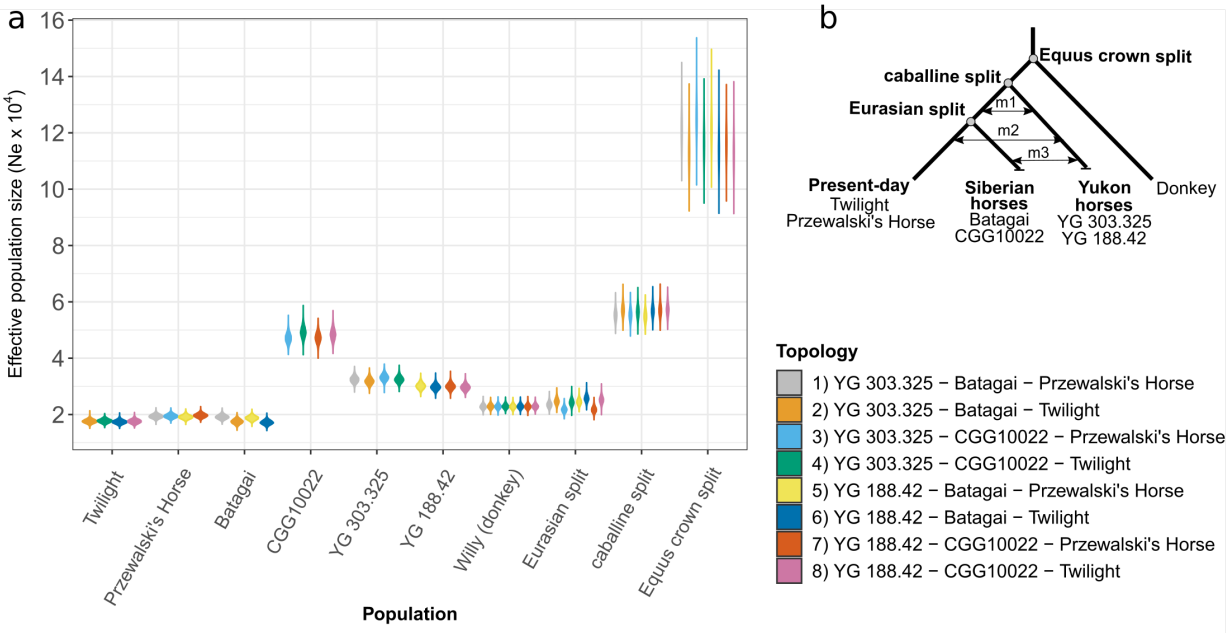

Figure S14. Genome-wide mutation rate (A) and generation time (B) used for re-scaling of G-PhoCS coalescent estimates into calendar years.

To account for uncertainty in estimates of the mutation rates and variable generation time we created gamma-distribution for each of these parameters with R function 'dgamma'. To calibrate G-PhoCS runs we randomly sampled values from these distributions.

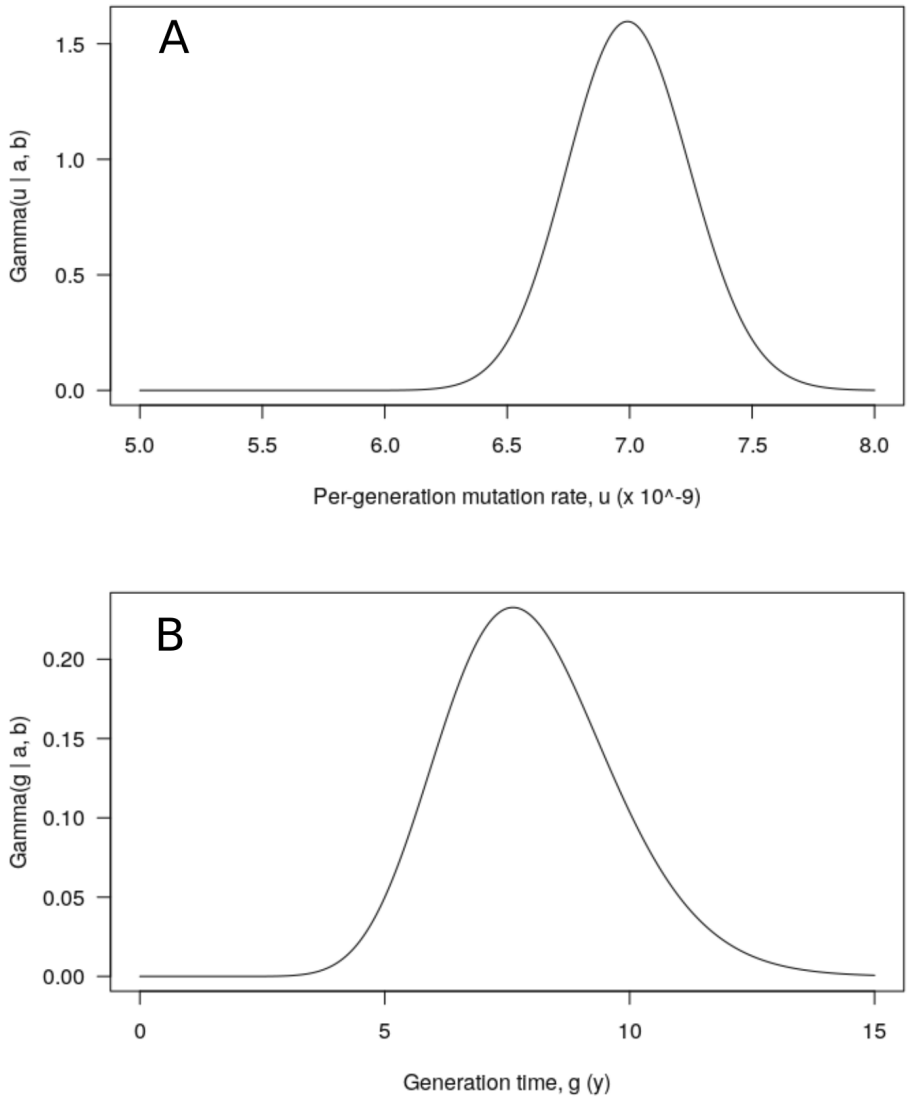

Figure S15. PSMC demographic inference for ancient caballine horses.

We ran PSMC on quality-filtered genome alignments to the EquCab2.0 reference. We re-scaled Θ for each ancient sample so that their demographic history reconstruction starts at the time of their calibrated radiocarbon age (see Table 1).