Title: The genomic history and global expansion of domestic donkeys

Authors: Evelyn T. Todd[1], Laure Tonasso-Calvière[1], Lorelei Chauvey[1], Stéphanie Schiavinato[1], Antoine Fages[1], Andaine Seguin-Orlando[1], Pierre Clavel[1], Naveed Khan[1,2], Lucía Perez Pardal[3,4], Laura Patterson Rosa[5], Pablo Librado[1], Harald Ringbauer[6], Marta Verdugo[7], John Southon[8], Jean-Marc Aury[9], Aude Perdereau[9], Emmanuelle Vila[10], Matilde Marzullo[11], Ornella Prato[11], Umberto Tecchiati[11], Giovanna Bagnasco Gianni[11], Antonio Tagliacozzo[12], Vincenzo Tinè[13], Francesca Alhaique[12], João Luís Cardoso[14,15], Maria João Valente[16], Miguel Telles Antunes[17], Laurent Frantz[18,19], Beth Shapiro[20,21], Daniel G. Bradley[7], Nicolas Boulbes[22], Armelle Gardeisen[23], Liora Kolska Horwitz[24], Aliye Öztan[25], Benjamin S. Arbuckle[26], Vedat Onar[27], Benoît Clavel[28], Sébastien Lepetz[28], Ali Akbar Vahdati[29], Hossein Davoudi[30], Azadeh Mohaseb[28,30], Marjan Mashkour[28,30,31], Olivier Bouchez[32], Cécile Donnadieu[32], Patrick Wincker[9], Samantha A. Brooks[33], Albano Beja-Pereira[3,4,34,35], Dong-Dong Wu[36,37], Ludovic Orlando[1]*.

Affiliations:

[1]Centre d'Anthropobiologie et de Génomique de Toulouse, Université Paul Sabatier; Toulouse 31000, France.

[2]Department of Biotechnology, Abdul Wali Khan University; Mardan 23200, Pakistan.

[3]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto; Vairão 4485-661, Portugal.

[4]BIOPOLIS Program in Genomics, Biodiversity and Land Planning, Campus de Vairão, Universidade do Porto; Vairão 4485-661, Portugal.

[5] Department of Animal Science, Sul Ross State University, US-90, Alpine, TX 79830, United States.

[6]Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology; Leipzig 04103, Germany.

[7]Smurfit Institute of Genetics, Trinity College Dublin; Dublin D02 PN40, Ireland.

[8]Earth System Science Department, University of California; Irvine, CA 92697, United States.

[9] Genoscope, Institut de biologie François Jacob, CEA, Université d'Evry, Université Paris-Saclay ; Evry 91042, France.

[10]Laboratoire Archéorient, Université Lyon 2; Lyon 69007, France.

[11]Dipartimento di Beni Culturali e Ambientali, Università degli Studi di Milano; Milan 20122, Italy.

[12]Bioarchaeology Service, Museo delle Civiltà; Rome 00144, Italy.

[13]Soprintendenza archeologia belle arti e paesaggio per le province di Verona, Rovigo e Vicenza; Verona 37121, Italy.

[14]ICArEHB, Campus de Gambelas, University of Algarve; Faro 8005-139, Portugal.

[15]Universidade Aberta ; Lisbon 1269-001, Portugal.

[16]Faculdade de Ciências Humanas e Sociais, Centro de Estudos de Arqueologia, Artes e Ciências do Património, Universidade do Algarve ; Faro 8000-117, Portugal.

[17]Centre for Research on Science and Geological Engineering, Universidade NOVA de Lisboa; Lisbon 1099-085, Portugal.

[18]Palaeogenomics Group, Department of Veterinary Sciences, Ludwig Maximilian University; Munich, 80539, Germany.

[19]School of Biological and Behavioural Sciences, Queen Mary University of London; London E1 4DQ, United Kingdom.

[20]Department of Ecology and Evolutionary Biology, University of California; Santa Cruz, CA 95064, United States.

[21]Howard Hughes Medical Institute, University of California; Santa Cruz CA 95064, United States.

[22]Institut de Paléontologie Humaine, Fondation Albert Ier, Paris / UMR 7194 HNHP, MNHN-CNRS-UPVD / EPCC Centre Européen de Recherche Préhistorique, Tautavel 66720, France.

[23]Archéologie des Sociétés Méditéranéennes, Université Paul Valéry - Site Saint-Charles 2 ; Montpellier 34090, France.

[24]National Natural History Collections, Edmond J. Safra Campus, Givat Ram, The Hebrew University; Jerusalem 9190401, Israel.

[25]Archaeology Department, Ankara University; Ankara, 06100, Turkey.

[26]Department of Anthropology, University of North Carolina at Chapel Hill; Chapel Hill, NC 27599, USA.

[27]Osteoarchaeology Practice and Research Center and Department of Anatomy, Faculty of Veterinary Medicine, Istanbul University-Cerrahpaşa; Istanbul 34320, Turkey.

[28]Archéozoologie, Archéobotanique, Sociétés, Pratiques et Environnements, Muséum National d'Histoire Naturelle; Paris 75005, France.

[29] Provincial Office of the Iranian Center for Cultural Heritage, Handicrafts and Tourism Organisation, North Khorassan, Bojnord, Iran.

[30]Archaezoology section, Bioarchaeology Laboratory of the Central Laboratory, University of Tehran; Tehran CP1417634934, Iran.

[31]Department of Osteology, National Museum of Iran; Tehran 1136918111, Iran.

[32]GeT-PlaGe - Génome et Transcriptome - Plateforme Génomique, GET - Plateforme Génome & Transcriptome, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement; Castaneet-Tolosan Cedex 31326, France.

[33]Department of Animal Science, UF Genetics Institute, University of Florida; Gainesville, FL 32610, United States.

[34]DGAOT, Faculty of Sciences, Universidade do Porto; Porto 4169-007, Portugal.

[35]Sustainable Agrifood Production Research Centre (GreenUPorto), Universidade do Porto; Vairão 4485-646, Portugal.

[36]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences; Kunming, 650201, China.

[37]Kunming Natural History Museum of Zoology, Kunming Institute of Zoology, Chinese Academy of Sciences; Kunming, Yunnan, 650223, China.

* Ludovic Orlando, ludovic.orlando@univ-tlse3.fr

Abstract: Donkeys transformed human history as essential beasts of burden for long-distance movement, especially across semi-arid and upland environments. They remain understudied despite globally expanding and providing key support to low-to-middle income communities. To elucidate their domestication history, we constructed a comprehensive genome panel of 207 modern and 31 ancient donkeys, including 15 wild equids. We found strong phylogeographic structure in modern donkeys supporting a single domestication in Africa ~5,000 BCE, followed by further expansions in this continent and Eurasia, ultimately returning back into Africa. We uncover a new genetic lineage in the Levant ~200 BCE, which contributed increasing ancestry towards Asia. Donkey management involved inbreeding and the production of giant bloodlines at a time when mules were essential to the Roman economy and military.

One-Sentence Summary: Ancient and modern genomes elucidate the origins, spread and management practices underlying donkey domestication.

Main Text:

Domestic donkeys (*Equus asinus*) have facilitated the movement of goods and people for millennia, enabling trade and transport across a broad spectrum of landscapes (*1*). Despite their importance to ancient pastoral societies, little is known about the deep history of donkeys and the impact of human management on their genomes. This is most likely due to their undervalued status and loss of utility in modern industrialized societies. Donkeys are, however, extraordinary working animals and remain essential for developing communities, especially in semi-arid environments (*2*). Understanding their genetic makeup is not only key to assess their contribution to human history but also to improving their local management in the future.

The current archaeological record of early donkeys is limited (*1, 3*), which makes their domestic origins and spread through the world contentious. The reduced body size of zooarchaeological ass remains in Egypt at El Omari (4,800–4,500 BCE) and Maadi (4,000–3,500 BCE) have been interpreted as early evidence of domestication (*4-7*). Carvings on the Libyan palette, found in Abydos, Egypt (3,200-3,000BCE), depict lines of walking asses, cattle, and sheep, also suggesting a domestication context (*8, 9*). Together with contemporary remains from the same region that show morphological evidence for load carrying (*10*), these findings suggest that donkeys could have been first domesticated within a range extending from the northeastern Sahara, the Nile Valley, the Atbara River, the Red Sea Hills, to Eritrea. In this model, donkeys were domesticated by pastoralists to assist with mobility around 5,500-4,500 BCE due to the large-scale aridification of the Sahara (*1*). Independent evidence based on patterns of mitochondrial (*11, 12*) and nuclear sequence variation (*13*) also point to African origins of the donkey, due to their closer proximity to African wild asses (*Equus africanus* spp.), than to Asian wild asses (*Equus hemionus* spp.).

However, candidate regions outside of Africa are also proposed as alternative domestication centers. In Ash Shuman (Yemen), for example, ass remains of disputed domestic status predate those from Egypt by 2,000 years (~6,500 BCE) (*14*). Likewise, textual, iconographic and zooarchaeological material indicate a possible additional center in Mesopotamia during the 4th and 3rd millennia BCE (*15-19*), a context in which first-generation hybrids of donkeys and Syrian onagers have been identified genetically (*20*). Segregation of mitochondrial variation in two main clades may also support a dual domestication process, for which the Nubian wild ass (*Equus africanus africanus*) is securely identified as the progenitor of Clade I (*11, 12*). As for the ancestor of Clade II, it could either be the extinct Atlas wild ass (*Equus africanus atlanticus*), endemic to northern Africa, or another undescribed subspecies that potentially ranged outside of Africa. Whether a single, maternally inherited marker captures the whole complexity of underlying ancestries can, however, be questioned, following recent results from other animals (e.g., horses (*21*)). Furthermore, previous analyses of nuclear genetic variation in African and non-African donkeys have failed to disentangle their origins (*13, 22*). Overall, this lack of consensus between genetic and archaeological data means that the geographic and temporal origin of donkeys and whether they were domesticated more than once remains uncertain. The global spread of donkeys is also unclear as their worldwide patterns of genomic diversity lacks extensive characterization.

Modern donkeys originated in Africa and spread into Eurasia

To address these issues, we sequenced 49 modern donkey genomes from underrepresented regions, and combined these with 158 publicly available to capture worldwide diversity (*13, 23-25*) (Fig. 1A, Table S1). We constructed a fine-scale recombination map from genomes encompassing all phylogenetic groups, which we used to phase 13,013,551 variants (Table S3,

S4). Principal Component Analysis (PCA; (*26*)) revealed strong geographical sub-structuring, with donkeys from Africa, Europe, and Asia forming distinct genetic clusters (Fig. 1B, S1, S2).

A Treemix phylogenetic reconstruction, grouping modern donkeys according to sampling locations (*27*), confirms the earliest split between African (Clade A) and mostly non-African donkeys (Clade B) (Fig. 1C). Further structure within Clade A separates donkeys from the Horn of Africa (Ethiopia and Somalia) plus Kenya, from those from western Africa (Ghana, Mauritania, Nigeria, and Senegal). Within Clade B, we find another major divergence between European and Asian donkeys, with east-to-west affinities in Europe from the Balkans (Croatia and Macedonia) to Iberia, Denmark, and Ireland. Conversely, Asian subpopulations show west-to-east sub-structuring from Iran and Central Asia to China and Mongolia. Combined, these findings suggest expansions from a central source into both continents.

In Clade B, some of the most basal donkeys are from the southern Arabian Peninsula (Oman and Yemen), whereas the single donkey from Saudi Arabia analyzed here shows European affinities indicative of a secondary translocation. Similarly, the Pega donkey from Brazil is nested within Iberia, mirroring the colonization history of the Americas. Clade B also includes donkeys from Nubia (Egypt and Sudan) showing affinities to the Levant (Syria) and Anatolia (Turkey), as well as donkeys from Maghreb (Tunisia), with closer genetic proximity to European subpopulations. This suggests gene flow into Africa from donkeys native to Anatolia and the Levant, but not to the Arabian Peninsula. Overall, this phylogenetic reconstruction is compatible with both models of donkey domestication: a unique origin in Africa followed by dispersals out and back, or dual origins in Africa and the southern Arabian Peninsula.

The unique origin model posits a demographic expansion in Africa first, and subsequent waves into Europe and Asia. In contrast, dual origins would result in an earlier split of demographic trajectories between African and Eurasian subpopulations, given their deep phylogenetic divergence. To test these, we first performed demographic modelling using SMC++ (*28*), which revealed the first expansion around 5,200 BCE (7,186±742 years ago), in line with archaeological evidence of domestication occurring at this time (Fig. 1D, S5). Additionally, when modelled from a possible African source, SMC++ trajectories indicated more recent and nearly coincidental expansions into Asia around 2,600 BCE (4,573±577 years ago) and Europe around 2,800 BCE (4,806±671 years ago) (Fig. 1D). This is in line with the unique origin model and the earliest archaeological evidence of donkeys in Asia (Iranian Plateau and the Indus Valley), and Europe (Portugal, Greece and Cyprus) in the mid-to-late 3rd millennia BCE (*29-34*).

Furthermore, Yemen and Oman subpopulations do not branch basal to Clade B according to fineSTRUCTURE (*35*), in contrast to the expectations of the dual origins model, but within Asian subpopulations (Fig. 2A, 2B). Lastly, pairwise genetic distances between Ethiopia and non-African subpopulations were greater than those from Yemen (Fig. S5). They both increased linearly with geographic distances and supported identical dispersal rates (Fig. S5; *p*-value=0.775), in line with a single wave of expansion at a constant pace. Therefore, our analyses support an early domestication in Africa, spreading at an even rate into the Arabian Peninsula and Eurasia, and flow back into Nubia and Maghreb. Modern subpopulations from the Horn of Africa and Kenya so far best represent the descendants of earliest donkeys.

Ancient donkey genomes reveal early and rapid dispersal into Asia and secondary contacts between Europe and western Africa

Patterns of genetic variation within modern subpopulations may reflect recent breeding history rather than early domestication (*36*). Additionally, they could under-represent the contribution of lineages that were once important but have since declined (*37*). Dating population splits also assumes constant, yet unknown generation intervals. To address these caveats and validate the domestication history reconstructed above, we generated a genomic time series spanning the last ~4,000 years, that included 31 ancient donkeys from 11 different sites, ranging from the Atlantic shores (Portugal) to Central Asia (eastern Iran/Turkmenistan) (Fig. 3A, Table S2).

Ancient genomes sequenced to 0.77-5.05 fold coverage (Table S6) were analyzed using two complementary methods: pseudo-haploidization following (*21*) resulting in 4,833,570 nucleotide transversions, and genotype imputation following (*38*), at 7,161,029 polymorphic sites present at more than 5% frequency in modern donkeys (*38, 39*). Imputation accuracy was confirmed by high consistency rates between imputed and observed genotypes following down-sampling of high-coverage modern genomes, and downstream analyses largely consistent with those based on pseudo-haploidized data (Fig. S6, S7, S9, S10, S11, S12, S14, Table S6).

The three oldest samples from our dataset consist of donkeys from Anatolia (Acemhöyük, Turkey), radiocarbon dated to 2,564-2,039 BCE. Their age and phylogenetic placement within Clade B confirm an early expansion out of Africa by ~2,500 BCE, in agreement with SMC++ time estimates (Fig. 2A, S10). These samples, and a donkey from eastern Iran/Turkmenistan affiliated to the Bactria–Margiana Archaeological Complex (BMAC, ~2,050 BCE; Chalow3), branch prior to the formation of modern subpopulations from central Asia (Kazakhstan, Kyrgyzstan, Turkmenistan) and eastern Asia (China, Mongolia, Tibet) (Fig. 2C). These subpopulations thus diverged after ~2,050 BCE, but potentially before the radiocarbon age of the

donkey from Doshan Tepe (1,049-928 BCE), which appears closer to modern subpopulations from central Asia in one Treemix analysis (Fig. 2D, S10).

Ancient samples from Iran (Shahr-i-Qumis, 800 BCE-800 CE), including one Sassanid (AM805) are not more closely related to central than to eastern Asian modern subpopulations, although their exact phylogenetic placement remains unclear (Fig. 2A, B, S10). Their fineSTRUCTURE affinities to modern Iran, Anatolia (Turkey), the Levant (Syria), and Maghreb (Tunisia) support different genetic ancestry profiles from those inferred at the nearby site of Doshan Tepe. This indicates a population turnover in Iran after ~1,000 BCE but before ~500 CE, corresponding to the radiocarbon time interval of Doshan Tepe and a single specimen from Shahr-i-Qumis.

Strikingly, all our ancient specimens from Europe cluster within modern European domesticates, supporting differentiation within this continent prior to the oldest European samples analyzed (Tarquinia, 803–412 BCE, ~2,500 years ago; Fig. 3C). However, a donkey from a Roman context in Marseille, a major seaport trading center in southern France (Centre Bourse Marseille, 0–500 CE), displayed strong genetic affinities with modern individuals from western Africa (Fig. 3B, 3D). Additionally, SNP and haplotype sharedness with modern western Africa were also found in European donkeys from Islamic era in Portugal (Albufeira, 1,228–1,280 CE) and Roman times in Northern France (Boinville-en-Woëvre, 200-500 CE) (Fig 3B, 3E). This reveals multiple contacts between Europe and western Africa from the Antiquity to Middle Ages. Despite ancient European donkeys showing western African ancestry, these contacts have impacted western Africa more than Europe, in line with Treemix inferring gene flow predominantly in this direction than the reverse (Fig. 1C). Interestingly, all modern Irish donkeys and the two Etruscan samples from Tarquinia are devoid of western African ancestry. This

suggests the preservation of old European genetic lineages, at least in some modern subpopulations of this continent.

Donkey management involved inbreeding and introgression from divergent lineages

Inbreeding is a common reproductive strategy for breeding animals with desirable traits (*40*). To assess whether ancient donkey breeders made use of inbreeding, we measured the proportion of autosomal runs of homozygosity (ROH) using three independent techniques, all of which provided consistent results (Fig. S13, S14). We detected inbreeding, but no significant changes in levels between modern and ancient donkeys (Wilcoxon rank sum test, p-value = 0.3951) (Fig. 4A, 4B, 4C). Conversely, modern horses show higher inbreeding levels than their ancient counterparts (Wilcoxon rank sum test, p-value<0.001), mirroring previous reports of reduced heterozygosity and increased deleterious mutation load in recent times (*21, 41*) (Fig. 4D, E, F). Longer ROH tracts are more common in modern horses and donkeys than in the past, consistent with inbreeding from closer generations in their genealogies (Fig. 4C, 4F). Overall, our analyses support recent major changes in reproductive management inflating inbreeding in horses, but not in donkeys.

Admixture modelling suggests ongoing introgression from African wild asses into modern donkeys from Africa and the southern Arabian Peninsula (with between 0.24–6.99% of admixture, Fig. 1A, S3, Table S5). This is in line with free-ranging local management practices allowing for continued interbreeding with wild and feral subpopulations (*4, 42*). The limited but significant amount of wild genetic material from kiangs in one modern donkey from China also supports admixture between taxa generally regarded as separate species. This confirms previous reports of mitochondrial introgression (*43*) and genomic admixture despite different karyotypes

(*24*). Interestingly, all but one ancient donkey (Tur168) carried remnants of outgroup material (0.21-4.15%; Fig. 3B), potentially resulting from recent range contractions of wild subpopulations and ancient management practices providing more opportunities for wild introgression.

The genome of MV242, a donkey from Israel dating to the Hellenistic period (350-58 BCE), displayed the largest fraction of divergent genetic material (Fig. 3B, 4.15%±0.019). In Treemix, this sample showed a deeper placement than all donkeys present in our panel, except the Somali wild ass (*E.a.som)* (Fig. 2E). Significantly positive f4(*E.a.som*, MV242; Horn+Ken, x) statistics revealed MV242-related genetic ancestry in some modern subpopulations (x), especially towards central and eastern Asia (Fig. 5E). This ancestry was already present in the BMAC sample from Iran (Chalow3, Fig. 5F), indicating contact ~2050 BCE at the latest. It was, however, absent in Acemhöyük at that time, suggesting that the MV242 divergent lineage ranged into eastern Iran/Turkmenistan, but not Turkey. This lineage also left genetic material in modern Anatolia, the Levant, Nubia, and Maghreb, but not in western Africa, consistent with donkeys carrying MV242-related ancestry flowing back into some African regions. Finally, this ancestry was also present in southwestern European subpopulations (CYK, ESP, PTG), but neither in the modern Balkans and Ireland, nor in any ancient European sample analyzed here (Fig. 5E-F). Combined, our results suggest a range for the MV242-related lineage from the Levant into Asia, rather than Europe and Africa.

Despite its divergent genetic makeup, MV242 carries a mitochondrial haplotype characteristic of Clade II (Fig. 5A). Our tip-calibrated coalescent analyses revealed that the time to the most common recent ancestor of that Clade was 32,226 BCE and not 332,580-142,980 BCE (Fig. 5B), as previously reported (*12, 44*). Since the same holds true for Clade I, both clades could have

coexisted in sympatry 25,000 years later as donkeys were first domesticated (Fig. 5B). Additionally, no phylogeographic structure is apparent in patterns of mitochondrial variation, both in modern and ancient subpopulations, as ancient specimens from Asia and Europe, sometimes from the same archaeological sites, were placed across both clades (Fig 5A). Y-chromosomal variation was also associated with little, if any, population structure (Fig. 5C, D). Combined, our results dismiss mitochondrial DNA and the Y-chromosome as reliable markers of domestication history in donkeys.

Romans bred improved donkeys for producing mules essential for their military power and economy

Beyond documenting domestication history at the global scale, our genomic dataset also included 3 jennies (females) and 6 jacks (males) from the same archaeological site (Boinville-en-Woëvre) (Fig. 3A). These were found in a dedicated farming area of a Roman villa, providing insights into local management practices in Roman Northern France (200-500 CE). One jack (GVA349) appeared particularly inbred with long ROH indicative of recent consanguinity (Fig. 4A) and was genetically related to four jacks and one jenny (family group GVA1, including GVA125, GVA347, GVA348, GVA349, GVA353, and GVA354; Table S10). Additionally, two jennies showed genetic relatedness coefficients equivalent to full siblings (family group GVA2 GVA355 and GVA358; Table S10). This indicates breeding management within close kin, potentially aimed at selecting for desirable traits. Genotype imputation at *TBX3* (*13*) revealed the presence of dun and derived colored coats, but no evidence for the dominant alleles associated with white spots or long hair was found in the sequence alignments at *KIT* (*45*) and *FGF5* (*46*) (Table S7-9). The latter two phenotypes are, however, relatively common in modern breeds from France, suggesting post-Roman selection for these traits.

297 The abundance of donkeys at Boinville-en-Woëvre stands as an exception in Roman France, as
298 mules dominated all other assemblages from this time (*47*). Contemporaneous Roman sites
299 report mules of a large and uniform size, indicating selective breeding in the parental species for
300 expensive animals of exceptional stature (Varo (2, 6)) (*48*). Interestingly, morphometric
301 measurements previously revealed five donkeys from family group GVA1 as giant (148-156cm
302 at the withers) (*47*). We found that GVA359 had a similarly large size (144cm) and genetic
303 affinities to western Africa. This may indicate restocking to enhance body size from distant
304 bloodlines carrying divergent ancestry, or from wild populations.

305 Interestingly, outgroup admixture was significantly higher at Boinville-en-Woëvre than in other
306 ancient donkeys except the divergent MV242 specimen (p-value = 0.045). Significantly negative
307 f4(kiang, MV242; Fiumarella1, Boinville-en-Woëvre) statistics support restocking into family
308 group GVA1 only, from a lineage more divergent than MV242 (Fig. 5G). Additionally, f4(kiang,
309 *E.a.som*; Fiumarella1, GVA1) statistics reject unbalanced allele sharedness between *E.a.som* and
310 GVA1, ruling out restocking from *E.a.som* or more divergent populations (Fig. 5H). Combined,
311 these findings uncover a lineage, phylogenetically intermediate between MV242 and *E.a.som*,
312 contributing to the genetic makeup of some Roman donkeys at Boinville-en-Woëvre. Together
313 with the evidence of genetic relatedness and inbreeding, this suggests Boinville-en-Woëvre as a
314 likely mule production center that maintained bloodlines of giant donkeys selected through
315 familial breeding and restocking. This center may illustrate how Romans sustained the enormous
316 demand for mules, which is documented in the nearby Rhine frontier (*49*), and has fueled
317 transportation networks throughout the Empire (*47*).

318 **Discussion**

Our study solves long-standing debates about donkey domestication. We support domestication starting from a unique African source ~5,000 BCE. Donkeys subsequently spread into Eurasia from ~2,500 BCE, and central and eastern Asian subpopulations differentiated ~2,000-1,000 BCE. Genetic affinities characteristic of modern western Europe were already formed by 500 BCE. Following early domestication, African donkeys further differentiated in the West and the Horn of Africa plus Kenya, but also received streams of genetic ancestry from western Europe as well as a region encompassing the Levant, Anatolia and Mesopotamia. Donkey domestication involved limited, but significant wild introgression. It did not entail inflated inbreeding in recent times, in contrast to horses. In fact, the processes of donkey and horse domestication dramatically differed, as horses were domesticated twice (*50*) and rapidly spread across Eurasia from the lower Don-Volga region ~2,000 BCE (*21*). Their regional differentiation remained relatively limited due to strong connectivity at continental distances early on and until oriental bloodlines were propagated throughout the world during the last 1,000 years (*41, 51*). The extent to which the different domestication trajectories of donkeys and horses were only driven by their respective roles in human societies or also reflected management practices adapted to their respective mating and social behavior (*52*), remains to be explored.

This work clarifies global patterns of donkey domestication and movements, but also highlights many directions for future research. For example, it remains unknown whether domestic donkeys only dispersed out of Africa by land through the Sinai Peninsula, or across the Red Sea from Ethiopia to Yemen. Additionally, modern subpopulations from the Horn of Africa plus Kenya were found to be the first expanding. This may suggest early domestication there, or donkeys domesticated elsewhere in Africa entering the region more recently. Further research is needed to clarify the timing of pastoral spread into the Red Sea Sudanese region and the Horn of Africa.

342 Current dates range from ~2500 BCE in Ethiopia and Eritrea (*53*) to ~3000 BCE in northern
343 Kenya (*54*). Donkeys are not present in the archaeological record of western Africa before the
344 beginning of the common era either (*55*), which postdates by 3,000 years the time when donkey
345 populations from Horn of Africa plus Kenya and western African are inferred to have split
346 genetically. This may indicate an early, yet undocumented arrival in the region, or a slow
347 migration westward, only reaching the modern range later. Improving the current African
348 archaeological record thus appears paramount to refining the exact context underlying early
349 donkey domestication and subsequent population movements.

350 Further genomic studies in other regions would also largely benefit the understanding of donkey
351 diversity and history. Resolving the genetic structure of equine remains from the 3rd millennia
352 BCE of southwest Asia will be challenging due to postmortem DNA decay, but essential to map
353 the geographic range of the divergent lineage identified here (MV242), as well as to understand
354 dispersal mechanisms in greater detail. The same holds true for Chalcolithic and Bronze Age
355 Europe, which remain genetically undocumented in our dataset, and onwards. Developing
356 genetic knowledge of ancient European donkeys will further clarify patterns of exchange across
357 the Mediterranean region, including during and after Roman times, as revealed in this study. It
358 will also provide insights into the dispersal mechanisms underpinning the genetically supported
359 presence of donkey remains in Portugal ~2,200 BCE (*33*). Genetic characterization of local
360 archaeological sites at the population scale may uncover additional mule breeding centers, other
361 than the one reported here. This will shed light on the diversity of breeding management
362 strategies developed by Romans to supply their continental-wide economy and military with
363 adequate animal resources (*49*). For now, both the absence of mules and rarity of horse mares at
364 Boinville-en-Woëvre (*47*) suggest that mares were brought in for mating before returning

365 pregnant to their owners. Alternatively, donkey breeders may have visited other farms with their
366 jacks to cover mares.

367 Efforts should continue to characterize the modern donkey diversity around the world, especially
368 in Saudi Arabia, which is currently characterized by a single individual, as well as in Africa, for
369 which no populations located south of the Equator have been sampled. Such efforts may not only
370 refine the historical legacy of past populations into the modern world, but also uncover the
371 genetic basis of desert adaptations, which could prove invaluable for future donkey breeding in
372 the face of global warming.

373

References and notes

1. P. Mitchell, *The donkey in human history: an archaeological perspective* (Oxford University Press, Oxford, UK, 2018).
2. S. L. Norris, H. A. Little, J. Ryding, Z. Raw, Global donkey and mule populations: figures and trends. *PLOS ONE* **16**, e0247830 (2021).
3. B. Kimura, F. Marshall, A. Beja-Pereira, C. Mulligan, Donkey domestication. *Afr Archaeol Rev* **30**, 83-95 (2013).
4. F. Marshall, *Rethinking agriculture: archeological and ethnoarcheological perspectives* (Left Coast Press, Walnut Creek, CA, USA, 2007).
5. S. Bokonyi, "The animal remains of Maadi, Egypt: A preliminary report" in *Studi di paletnologia in onore di Salvatore M. Puglisi* M. Liverani, A. Palmieri, R. Peroni, Eds. (Università di Roma "La Sapienza,, Rome, 1985), pp. 495-499.
6. J. Boesneck, A. Von den Driesch, "Tierreste aus der vorgeschichtlichen siedlung von El-Omari bei" in *El Omari : a neolithic settlement and other sites in the vicinity of Wadi Hof, Helwan,* F. Debono, B. Mortensen, Eds. (1990), pp. 99-107.
7. J. Boesneck, A. von den Driesch, R. Ziegler, "Die Tierreste von Maadi und dem Friedhof am Wadi Digla" in *Maadi III,* I. Rizkana, J. Seeber, Eds. (1989), pp. 87-128.
8. K. M. Cialowicz, *Les palettes Egyptiennes aux motifs zoomorphes et sans decoration. Etudes de l'art predynastique* (Jagiellonian University, Krakow, Poland, 1991).
9. D. J. Brewer, D. B. Redford, S. Redford, *Domestic plants and animals: the Egyptian origins* (Warmister: Aries and Phillips, 1994).

10. S. Rossel, F. Marshall, J. Peters, T. Pilgram, M. D. Adams, D. Connor, Domestication of the donkey: timing, processes, and indicators. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3715 (2008).

11. A. Beja-Pereira, P. R. England, N. Ferrand, S. Jordan, A. O. Bakhiet, M. A. Abdalla *et al.*, African origins of the domestic donkey. *Science* **304**, 1781 (2004).

12. B. Kimura, F. B. Marshall, S. Chen, S. Rosenbom, P. D. Moehlman, N. Tuross *et al.*, Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication. *Proc. R. Soc. B* **278**, 50-57 (2011).

13. C. Wang, H. Li, Y. Guo, J. Huang, Y. Sun, J. Min *et al.*, Donkey genomes provide new insights into domestication and selection for coat color. *Nat. Commun.* **11**, 6014 (2020).

14. M. Cattani, S. Bokonyi, "Ash-Shumah. An early Holocene settlement of desert hunters and mangrove foragers in the Yemeni Tihama" in *Essays on the Late Prehistory of the Arabian Peninsula. Serie Orientale Romana XCIII,* S. Cleuziou, T. Maurizio, Z. Juris, Eds. (Istituto Italiano per l'Africa e l'Oriente, Roma, 2002).

15. R. H. Meadow, H. P. Uerpmann, *Equids in the ancient world, vol. 2* (Wiesbaden, 1991).

16. J. Clutton-Brock, The process of domestication. *Mamm Rev.* **22**, 79-85 (1992).

17. J. Zarins, R. Hauser, The domestication of equidae in third-millennium BCE Mesopotamia. *Cornell University Studies in Assyriology and Sumerology* **24**, XI + 432 (2014).

18. E. Vila, "Data on equids from late fourth and third millennium sites in Northern Syria" in *Equids in Time and Space: Papers in Honour of Véra Eisenmann.,* M. Mashkour, Ed. (Oxford, Oxbow, 2006), pp. 101-123.

19. J. Boessneck, A. von den Driesch, U. Steger, Tierknochenfudne der ausgrabungen des Deutschen Archäologischen Instituts Baghdad in Uruk-Warka, Iraq. *Baghdader Mitteilingen* **15**, 149-189 (1984).

20. A. E. Bennett, J. Weber, W. Bendhafer, S. Champlot, J. Joris Peters, G. M. Schwartz *et al.*, The genetic identity of the earliest human-made hybrid animals, the kungas of Syro-Mesopotamia. *Sci. Adv.* **8**, eabm0218 (2022).

21. P. Librado, N. Khan, A. Fages, M. A. Kusliy, T. Suchan, L. Tonasso-Calvière *et al.*, The origins and spread of domestic horses from the Western Eurasian steppes. *Nature* **598**, 634–640 (2021).

22. S. Rosenbom, V. Costa, S. Chen, L. Khalatbari, G. H. Yusefi, A. Abdukadir *et al.*, Reassessing the evolutionary history of ass-like equids: Insights from patterns of genetic variation in contemporary extant populations. *Mol. Phylogenet. Evol.* **85**, 88-96 (2015).

23. G. Renaud, B. Petersen, A. Seguin-Orlando, M. F. Bertelsen, A. Waller, R. Newton *et al.*, Improved de novo genomic assembly for the domestic donkey. *Sci. Adv.* **4**, eaaq0392 (2018).

24. H. Jónsson, M. Schubert, A. Seguin-Orlando, A. Ginolhac, L. Petersen, M. Fumagalli *et al.*, Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 18655 (2014).

25. L. Zeng, H. Q. Liu, X. L. Tu, C. M. Ji, X. Gou, A. Esmailizadeh *et al.*, Genomes reveal selective sweeps in kiang and donkey for high-altitude adaptation. *Zool. Res.* **42**, 450-460 (2021).

26. A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, D. Reich, Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904-909 (2006).

27. J. K. Pickrell, J. K. Pritchard, Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).

28. J. Terhorst, J. A. Kamm, Y. S. Song, Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303-309 (2017).

29. D. Reese, Faunal remains from early Helladix II Lerna (Argolid-Greece). *Mediterr. Archaeol. Archaeom.* **13**, 289-320 (2013).

30. B. A. Knapp, Bronze Age Mediterranean Island Cultures and the Ancient Near East. *The Biblical Archaeologist* **55**, 52-72 (1992).

31. S. Ratnagar, *Trading Encounters: from the Euphrates to the Indus in the Bronze Age* (Oxford: Oxford University Press, 2004).

32. M. A. Zeder, "The equid remains from Tal-e Malyan, southern Iran" in *Equids in the Ancient World,* R. H. Meadow, H. P. Uerpmann, Eds. (Wiesbaden: Reichert, 1986), pp. 164-193.

33. J. L. Cardoso, J. T. Vilstrup, V. Eisenmann, L. Orlando, First evidence of Equus asinus L. in the Chalcolithic disputes the Phoenicians as the first to introduce donkeys into the Iberian Peninsula. *J. Archaeol. Sci.* **40**, 4483-4490 (2013).

34. S. Amiri, M. Mashkour, F. Mohaseb, Naseri R, "The Subsistence Economy of a Highland Settlement in the Zagros during the Bronze and Iron Ages. The Case of Gūnespān (Hamadan, Iran)" in *Archaeozoology of Southwest Asia and Adjacent Areas XIII: Proceedings of the Thirteenth International Symposium, University of Cyprus, Nicosia,*

Cyprus, June 7-10, 2017. , J. Daujat, A. Hadjikoumis, R. Berthon, J. Chahoud, V. Kassianidou, J. D. Vigne, Eds. (2021), pp. 199-219.

35. D. J. Lawson, G. Hellenthal, S. Myers, D. Falush, Inference of population structure using dense haplotype data. *PLoS Genet.* **8**, e1002453 (2012).

36. L. Girdland Flink, R. Allen, R. Barnett, H. Malmström, J. Peters, J. Eriksson *et al.*, Establishing the validity of domestication genes using DNA from ancient chickens. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6184-6189 (2014).

37. L. A. F. Frantz, D. G. Bradley, G. Larson, L. Orlando, Animal domestication in the era of ancient genomics. *Nat. Rev. Genet.* **21**, 449-460 (2020).

38. R. Hui, E. D'Atanasio, L. M. Cassidy, C. L. Scheib, T. Kivisild, Evaluating genotype imputation pipeline for ultra-low coverage ancient genomes. *Sci. Rep.* **10**, 18542 (2020).

39. B. l. Browning, S. R. Browning, Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* **98**, 116-126 (2016).

40. T. N. Kristensen, A. C. Sørensen, Inbreeding – lessons from animal breeding, evolutionary biology and conservation genetics. *Animal Science* **80**, 121-133 (2007).

41. A. Fages, K. Hanghøj, N. Khan, C. Gaunitz, A. Seguin-Orlando, M. Leonardi *et al.*, Tracking five millenia of horse management with extensive ancient genome time series. *Cell* **177**, 1419-1435.e1431 (2019).

42. F. Marshall, L. Weissbrod, "The consequences of women's use of donkeys for pastoral flexibility: Maasai ethnoarchaeology" in *Tracking Down the Past. Ethnohistory Meets Archaeozoology,* G. Grupe, J. Peters, G. McGlynn, Eds. (Rahden/Westf. : M. Leidor, 2009), pp. 59-79.

43. E. A. Bennett, S. Champlot, J. Peters, B. S. Arbuckle, S. Guimaraes, M. Pruvost *et al.*, Taming the late Quaternary phylogeography of the Eurasiatic wild ass through ancient and modern DNA. *PLOS ONE* **12**, e0174216 (2017).

44. L. Wang, G. Sheng, M. Preick, S. Hu, T. Deng, U. H. Taron *et al.*, Ancient mitogenomes provide new insights into the origin and early introduction of Chinese domestic donkeys. *Front. Genet.* **12**, (2021).

45. B. Haase, S. Rieder, T. Leeb, Two variants in the KIT gene as candidate causative mutations for a dominant white and a white spotting phenotype in the donkey. *Anim. Genet.* **46**, 321-324 (2015).

46. R. Legrand, L. Tiret, M. Abitbol, Two recessive mutations in FGF5 are associated with the long-hair phenotype in donkeys. *Genet. Sel. Evol* **46**, 65 (2014).

47. S. Lepetz, B. Clavel, D. Alioğlu, L. Chauvey, S. Schiavinato, L. Tonasso-Calvière *et al.*, Historical management of equine resources in France from the Iron Age to the Modern Period. *J. Archaeol. Sci. Rep.* **40**, 103250 (2021).

48. G. K. Kunst, Archaeozoological evidence for equid use, sex structure and mortality in a Roman auxiliary fort (Carnuntum-Petronell, lower Austria). *Anthropozoologica* **31**, 109-118 (2000).

49. C. Johnstone, "Commodities or logistics? : The role of equids in Roman supply networks" in *Feeding the Roman Army: The Archaeology of Production and Supply in NW Europe,* S. Stalibrass, R. Thomas, Eds. (Oxbow Books, Oxford, 2008).

50. C. Gaunitz, A. Fages, K. Hanghøj, A. Albrechtsen, N. Khan, M. Schubert *et al.*, Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science* **360**, 111-114 (2018).

51. S. Felkel, C. Vogl, D. Rigler, V. Dobretsberger, B. P. Chowdhary, O. Distl *et al.*, The horse Y chromosome as an informative marker for tracing sire lines. *Sci. Rep.* **9**, 6095 (2019).

52. F. B. Marshall, K. Dobney, T. Denham, J. M. Capriles, Evaluating the roles of directed breeding and gene flow in animal domestication. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6153-6158 (2014).

53. J. Lesur, E. A. Hildebrand, G. Abawa, X. Gutherz, The advent of herding in the Horn of Africa: new data from Ethiopia, Djibouti and Somaliland. *Quat* **343**, 148-158 (2014).

54. E. A. Hildebrand, K. M. Grillo, E. A. Sawchuk, S. K. Pfeiffer, L. B. Conyers, S. T. Goldstein *et al.*, A monumental cemetery built by eastern Africa's first herders near Lake Turkana, Kenya. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 8942-8947 (2018).

55. K. MacDonald, R. Hutton MacDonald, "The origins and development of domesticated animals in arid West Africa" in *The origins and development of African livestock: archaeology, genetics, linguistics and ethnography.,* R. M. Blench, K. MacDonald, Eds. (University College London Press, London, UK, 2000).

56. D. H. Alexander, K. Lange, Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform.* **12**, 246 (2011).

57. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).

58. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan *et al.*, Ancient admixture in human history. *Genetics* **192**, 1065 (2012).

59. F. G. Vieira, A. Albrechtsen, R. Nielsen, Estimating IBD tracts from low coverage NGS data. *Bioinformatics* **32**, 2096-2102 (2016).

60. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler *et al.*, IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530-1534 (2020).

Fig. 1: Modern donkey dataset and population evolutionary history. A) Number and geographical distribution of modern donkey samples (*n*=207). Pie charts show the ADMIXTURE proportion of domestic ancestry (grey), African wild ass ancestry (white) and kiang ancestry (black) averaged across all individuals from each country (*56*). For visualization, the total surface of each pie chart is scaled to 2%. B) Smartpca (*57*) of modern donkeys, with the imputed ancient samples in black. C) Treemix phylogeny of modern domesticates (excluding individuals with high wild introgression, *n*=201) (*27*). Node supports are estimated from 100 bootstrap pseudo-replicates (confidence < 90% in red). Percentage values indicate admixture proportions inferred from Treemix (*27*). D) SMC++ demographic trajectories (colored) and split time estimates (black) for pairs of main geographic regions (*28*), repeating the analysis on two datasets of three individuals per population (the second dataset is shown in semi-transparency). Modern donkeys are colored and shaped according to geographical location and continents in all panels.

Fig. 2: Haplotype sharedness and phylogenetic placement of ancient European donkeys. A) Haplotype sharedness clustering of modern (*n*=168) and ancient donkeys (*n*=31) reconstructed using fineSTRUCTURE (*35*). Modern domesticates are colored following Fig. 1 and ancient individuals are numbered according to Fig. 3A. Cluster supports are shown in percentage on each node if >0.8. MV242 placement is incongruent with Treemix (Fig. 2E), due to the limited representation of divergent ancestries in the modern reference panel used for imputation. B) Co-ancestry matrix based on haplotype sharedness. Co-ancestry values averaged for co-clustered individuals. C-E) Treemix phylogenies of three ancient specimens shown in black (C: Chalow3, D: Doshan Tepe, E: MV242) placed within the subpopulations defined in Fig. 1C (*27*). Branches that are not scaled are shown as dashed lines.

Fig. 3: Ancient donkey dataset, genetic affinities to outgroups and modern donkeys. A) Geographical distribution, estimated age and sample names of ancient donkeys (*n*=31). Pie charts represent the proportion of individuals with dun coat color (white), heterozygotes (grey), and derived coat color (red) at each site. Genotype probabilities >=0.99 denoted with ** and >=0.9 with *. B) Heatmap displaying outgroup f3-statistics in the form of (modern, ancient; kiang) (*58*). Bar charts represent the proportion of wild ancestry (kiang, onager, zebra, *E.a.som*) in each ancient individual with standard errors estimated from ADMIXTURE with 100 bootstrap pseudo-replicates (*56*). C-E) Treemix phylogenies of ancient specimens from three archaeological sites shown in black (C: Tarquinia (Tarquinia214, Tarquinia501), D: Bourse (BourseB, BourseC), E: Albufeira) placed within the subpopulations defined in Fig. 1C (*27*). Branches that are not scaled are shown as dashed lines.

Fig 4: Inbreeding in domestic donkeys and horses. A) Distribution of total runs of homozygosity (ROH) length in modern versus ancient donkeys. B) The total length of ROH in donkey genomes through time. C-D) Same as A-B, but for 79 modern and 75 ancient horses. ROH tracts were identified using ngsF-HMM (*59*).

Fig. 5: Uniparental marker phylogenies and introgression of divergent lineages. A) Mitochondrial phylogeny constructed using IQ-TREE (*60*) with 100 bootstrap pseudo-replicates marked with a black triangle if >90%. B) Posterior distributions of the time to the most recent common ancestors of all mitochondrial haplotypes, Clade I and Clade II labelled with their modes. C-D) Same as A-B for the Y-chromosome. E-H) f4-statistics (*58*) exploring the genetic

578 contribution of divergent lineages into modern and ancient donkeys. Z scores were corrected for
579 multiple testing, and red bars with asterisks show p-value<0.05.

Acknowledgments:

Funding: LKH would like to thank the director of the Tel es-Safi/Gath excavation, Prof. Aren Maeir, for facilitating sampling. The DonkeyBank collection of modern donkey DNA samples is supported by the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement Number 857251). Lucia Perez-Pardal is funded by national funds from FCT – Fundação para a Ciência e a Tecnologia, I.P. Genoscope and GeT-PlaGe sequencing platforms are partly funded by France Génomique National infrastructure, funded as part of Investissement d 'avenir' program managed by Agence Nationale pour la Recherche (contract ANR-10-INBS-09). This project has received funding from the CNRS, University Paul Sabatier (AnimalFarm IRP), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreements 885729-AncestralWeave, 295729-CodeX, 853272-PALAEOFARM, and; 681605-PEGASUS).

Author contributions:

Conceptualization: LO

Materials and Reagents: JMA, AP, PW, ABP, DDW, LO

Archaeological samples and contextual information: MV, EV, MM, OP, UT, GBG, AT, VT, FA, JLC, MJV, MTA, NB, AG, LKH, AÖ, BA, OV, BC, SL, AAV, HD, AM, MM.

Investigation: SBB, DLA, MPW, WCB.

Sampled modern donkeys: ABP, DDW, LPR, SAB.

Radiocarbon dating: JS.

DNA sequencing: MTC, LC, SS, ASO, AP, OB, CD, PW.

Data analysis: ETT, PL, LO.

Writing – Supplemental Information: ETT, with input from LO

Writing – Main Article: ETT, PL, LO, with input from all co-authors.

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: The sequence data generated in this study is available for download at the European Nucleotide Archive (Accession number = PRJEB52849). The accession numbers for each individual sample and all other data used in this study are included in Table S1, S2 and S11 of the Supplementary Information.

Supplementary Materials:

Materials and Methods

Figs. S1 to S15

Tables S1 to S11

References 61-137

Science
AAAS

Supplementary Materials for

The genomic history and global expansion of domestic donkeys

Evelyn T. Todd, Laure Tonasso-Calvière, Lorelei Chauvey, Stéphanie Schiavinato, Antoine Fages, Andaine Seguin-Orlando, Pierre Clavel, Naveed Khan, Lucía Perez Pardal, Laura Patterson Rosa, Pablo Librado, Harald Ringbauer, Marta Verdugo, John Southon, Jean-Marc Aury, Aude Perdereau, Emmanuelle Vila, Matilde Marzullo, Ornella Prato, Umberto Tecchiati, Giovanna Bagnasco Gianni, Antonio Tagliacozzo, Vincenzo Tinè, Francesca Alhaique, João Luís Cardoso, Maria João Valente, Miguel Telles Antunes, Laurent Frantz, Beth Shapiro, Daniel G. Bradley, Nicolas Boulbes, Armelle Gardeisen, Liora Kolska Horwitz, Aliye Öztan, Benjamin S. Arbuckle, Vedat Onar, Benoît Clavel, Sébastien Lepetz, Ali Akbar Vahdati, Hossein Davoudi, Azadeh Mohaseb, Marjan Mashkour, Olivier Bouchez, Cécile Donnadieu, Patrick Wincker, Samantha A. Brooks, Albano Beja-Pereira, Dong-Dong Wu, Ludovic Orlando.

Correspondence to: ludovic.orlando@univ-tlse3.fr

This PDF file includes:

Summary

This document describes the methods that have been involved in this study. The first part of these analyses focusses on a panel of 207 modern donkey and 15 wild equid genomes, 49 of which are newly described in this study. These genomes were used to: 1) call variants (GraphTyper (version 2.5.1) (*61*)); 2) create a recombination map (LDHat (version 2.2) (*62*)); 3) call phased haplotypes (BEAGLE (version 5.1) (*39*)) ; and 4) infer the population history and structure (PLINK (version 1.9) (*63*), ADMIXTURE (version 1.3.0) (*56*), qpAdm (version 810) (*64*), Treemix (version 1.13) (*27*), SMC++ (version 1.15.4) (*28*) and ADMIXTOOLS2 (*58, 65*)).

Additionally, the second part of the analysis leverages the modern genome panel, supplemented with 31 ancient donkey genomes spread across central Asia to western Europe and spanning the last 4,500 years. We created two datasets to fully exploit the genetic information of these samples, both pseudo-haploidising genomes at transversion sites (n= 4,833,570), and imputing genomes for the set of variants identified in the modern panel (n=7,161,029) (BEAGLE versions 4.0 and 5.1). Those datasets were used to infer the past population dynamics and assess breeding management through ADMIXTURE, PLINK, Treemix, fineSTRUCTURE (version 4.1.1) (*35*), KING (version 2.2.7) (*66*), NgsRelate (version 2) (*67*), NgsF-HMM (version 1) (*59*) and qpDstat (version: 751) (*58, 65*).

Finally, modern and ancient sequences aligned against the mitochondrial genome and Y-chromosome were used to infer phylogenetic relationships within both maternal and paternal lineages and reconstruct their past demographic trajectory (IQ-TREE (version 1.6.12) (*60*) and BEAST (version 2.6.5)) (*68-70*).

Materials and Methods

Sample collection, DNA extraction and genome sequencing of modern samples

We extracted and sequenced DNA from 48 tissue samples of domestic donkeys kindly provided from the existing collection of Dr. Albano Beja-Pereira, which were collected between 2000 and 2002 (DonkeyBank, CIBIO-InBIO, University of Porto). The sampling was revised and approved by CIBIO bioethic board. Samples from this collection have been used across the years in several published studies (*11, 12, 71*). The curation of this sample bank is oriented by the principles of the 3Rs, avoiding unnecessary sampling of animals whenever the collection has samples representing a region or the desired donkey phenotype. Only from 2015 onward, did export and ethical and animal welfare permits start to be required from the samples stored in this collection. Up to this date, it was not a general practice to require such permits from domestic animals, and unfortunately, even less in the case of the donkey. When these samples were collected, the owners first approached the animal to calm them down. The marginal region of the ear was cleaned with 70% ethanol and a single-use sterile punch biopsy was used to take a tiny piece of skin about 0.2 cm^3 from each individual. Particular attention was devoted to collecting the tissue along the margin and not across the ear, as this area is poorly irrigated and not sensitive. The punch biopsy device automatically cauterizes the possible small capillary vessels from the place where a sample was taken. Usually, this takes a split second and does not require holding the animals for blood sampling and animals do not generally react. After sampling, blue spray disinfectant was applied to the region. Nervous or frightened animals were avoided and the animal was observed for some minutes after having been sampled.

Around 20 plucked hairs (with roots) were instead collected from animals for which the owner expressed a preference for plucking hairs instead of tissue. Normally, dorsum or neck hairs were individually plucked from the animal without the need of restraining the animal. The collected hairs or tissues were stored in the plastic tube and completely submerged in preservative (96% alcohol) with at least three parts of ethanol for each part of the tissue. DNA from DonkeyBank tissues were extracted from the tissues using the JetQuick™ Tissue DNA Spin Kit (Genomed, GmbH) and the concentration of DNA extracts was measured using a Qubit Fluorimeter (Thermo Fisher Scientific).

A single specimen from a Pega donkey was provided by the Brooks Equine Genetics Lab (University of Florida, Gainesville, FL, USA). The sampling was revised and approved by the UF - IACUC Protocol #201408411. The hair sample, including hair roots, was pulled from the tail of the individual, and stored in a clean paper envelope. DNA from this sample was extracted using a modified lysis protocol described by Cook and colleagues (*72*).

Publicly available fastq files for 158 domestic donkeys, 2 Asiatic wild asses, 1 *E. africanus somaliensis* (*E.a.som*), 1 *E. zebra hartmannae*, 1 *E. zebra grevyi*, 1 *E. zebra burchelli*, 2 *E. hemionous* and 7 *E. kiang* were downloaded from the National Library of Medicine and Genome Sequence Archive database (*13*, *24*, *73*).

Details and accession numbers for all samples sequenced and downloaded from public databases can be found in Table S1.

Archaeological samples and context (Provenance)

The following section describes the archaeological contexts associated with all ancient donkeys sequenced in this study. The full name of each site is composed of the modern country where the excavation site lays followed by the age in Before Common Era (BCE) or Common Era (CE) as estimated from radiocarbon dating or inferred from archaeological context. The accession number and associated metadata for each ancient donkey genome included in this paper can be found in Table S2.

- **TUK_2564-2039BCE:** Acemhöyük, Turkey (samples: AC14380, AC14415, MV051).

Acemhöyük is a large mound site located in the Aksaray province of central Turkey representing an important urban center in the Early and Middle Bronze Age (EBA and MBA, ~2,800-1,700 BCE). The site is located at an elevation of approximately 950 m above sea level on the alluvial fan of the Melendiz river near the central Anatolian Great Salt Lake (Tüz Gölü). Acemhöyük consists of twelve major occupational levels with deposits representing EBA, MBA, Early Iron Age, Hellenistic, and modern occupations. The site is best known for its well-preserved Sarıkaya and Hatıplar 'palace' structures, which were built in the early 18th century BCE and destroyed by a violent fire in the mid-18th century BCE (*74*). These remains were excavated and studied by Dr. Nimet Özgüç, who documented extensive connections between the Sarıkaya palace at Acemhöyük and Kültepe-Kanesh, the kingdom of Karkemis on the Syrian-Turkish border, as well as the Assyrian kingdom of Šamši Adad (*75*). More recent excavations by Dr. Aliye Öztan have explored administrative buildings within the city center associated with the MBA occupation (including the 'Hizmet binası'), which were also destroyed by the fire that likely ended the settlement's role as a political center towards the end of the MBA (*74*). Moreover, Öztan's excavations have uncovered extensive exposures of the EBA occupation including 75 meters length of the EBA city wall on the south-eastern margin

of the mound as well as associated buildings dating to mid to late 3rd millennium BCE (EBIII) (*76-78*). Deposits associated with EBA levels XI and X include evidence for the destruction of city wall including as many as 1500 biconical clay balls interpreted as sling stones, human remains subject to violent death, as well as extensive pits filled with burnt and ashy deposits (*79, 80*). Based on a direct radiocarbon date on human bone from area AB/52 associated with these deposits (Sk4: BETA464596, 3920±30 bp, 95% 2480 - 2299 cal BC), this destruction is dated to the second half of the 3rd millennium BCE (*80*).

All three donkey specimens from Acemhöyük utilized in this study are petrosal portions of the temporal bone derived from grid square EB/50 and assigned to stratigraphic level XI or X dating to the EBA (EBIII). Specimen AC14380 (derived from mekan C) was recovered on the 23rd of July, 2012. The specimen AC14415 was recovered on the 24th of July 2015; while specimen MV051 (recorded as specimen AC13084 in the Acemhöyük zooarchaeological database) was recovered on August 17th, 2012. All of these specimens derive from deposits representing multiple complete or partial donkey burials located in close proximity to the level XI city wall (MNI of 8 donkeys recovered from this area). They were recovered from shallow deposits directly under the remains of structures associated with the modern village, which currently surrounds the mound and were initially thought to be modern pits related to the disposal of donkey remains. However, it became clear that these donkey burials, as well as others in adjacent areas DB/50 (MNI=3) and DB/48 (MNI=3) are associated with the EBA occupation of the city. Specimen MV051 has been directly dated by radiocarbon assay placing it in the last quarter of the 3rd millennium BCE, which corresponds with the phasing of the stratigraphic context to levels XI and X (UBA-30288, 3784±41 BP; 2285-2141 cal. BCE). Samples AC14380 and AC14415 were also radiocarbon dated and returned the same measurement (UCIAMS-199621 and UCIAMS-199619, 3945±20 BP; 2564-2346 cal. BCE) (Table S2).

- **IRA_ 2400BCE-2039BCE:** Chalow, Iran (sample: Chalow3)

Chalow cemetery is located in the North Khorasan Province in the North East of Iran. It was first located by Dr. Ali Akbar Vahdati in 2006 and the discovery of material culture placed this site in the Middle-Late Bronze / Bactria-Margiana Archaeological Complex (BMAC) (2200 to 1900 BCE) (*81*). In Trench 41E, Grave 6 East, excavated by Dr. Vahdati and Dr. Raffaele Biscione in 2015, an equid was discovered buried beside a human skeleton. This equid was later identified as a donkey and included in the current study (Table S2).

- **IRA_1049BCE-928BCE**: Doshan Tepe, Iran (sample: DoshanTepe).

Doshan Tepe is one of the five archaeological sites of the Ozbaki archaeological zone located in the Savojbolagh plain, 75 kilometers towards the north-west of Tehran, with excavations starting in 1998. The site of Doshan Tepe is located 250 meters to the west of the Main Tepe (Ozbaki Median Fortress). The plain was occupied form the 6th millennium BCE with excavations leading to the identification of 3 periods of the Iron Age. The latest is contemporaneous to the Median period and the two earliest periods are dated from the second half of the second millennium to the advent of the Median dynasty. The presence of grey pottery suggests non-local traditions. Doshan Tepe had also an important role in the region since cuneiform tablets were found in the Ozbaki archaeological zone. Studies of the faunal remains identified numerous equids at this sites, including 29 donkeys, 11 hemiones, 8 horses and 4 probable hybrids and 93 unidentified equids (*82*). The donkey sample in this study

belongs to the Iron Age II chronology in Iran. It was directly radiocarbon dated to 1049-928 cal. BCE (UCIAMS-223195, 2840±15 BP) (Table S2).

- **ITA_803-412BCE:** Tarquinia, Italy (samples :Tarquinia214, Tarquinia501).

The 'monumental complex' of Tarquinia offers the extraordinary opportunity to monitor the cultural development of an Etruscan area sacred to the major female goddess of the Etruscans. Archaeological evidence sheds light on the continuity and memory of the sacred area over the centuries up to the encounter with Rome. From the end of the 10th century BCE, offerings located by a natural cavity show the cult of a divinity of Nature, who catalyzed the very first community. Ritual sealing of a number of votive pits of different size contain a considerable number of animal bones (*83*). Samples Tarquinia214 and Tarquinia501 were found in the texture of pavements of structures belonging to the Archaic phase of the site. Both samples were directly radiocarbon dated. The date obtained for specimen Tarquinia214, 750-412 BCE (UCIAMS-224884, 2445±20 BP), overlaps the archaeological context (~550BCE). Two dates were obtained for specimen Tarquinia501, which returned a range slightly older than those estimated based on the archaeological context (803-547 BCE vs 520-500BCE) (UCIAMS-224885 and UCIAMS-224886, 2515±20 BP and 2656±20 BP) (Table S2).

- **ISR_350_58BCE:** Nizzana, Israel (sample: MV242).

Nizzana (sometimes also written as Nessana) is located 52km to the South-West of the city Beersheba. The site was first occupied in the Hellenistic period, and settlement continued throughout the Roman and Byzantine periods until its abandonment in the Early Islamic period. Architectural remains include residential buildings, a Late Roman military fort, three Byzantine churches and a monastery and notably was a 6–7th century CE papyrus archive (*84-86*). The sample MV242 dates back to the Hellenistic period and was radiocarbon dated to 350-58 cal. BCE (UCIAMS-199283, 2150±20 BP) (Table S2).

- **IRA_800BCE-800CE:** Shahr-i-Qumis, Iran (samples: AM39, AM44, AM66, AM71, AM805, AM89).

Shahr-i-Qumis is a site in Northeast Iran, consisting of several isolated mounds spread across an area of 28 kilometers. This site dates back to the Parthian and Sassanian periods, although some recent radiocarbon dating of faunal remains show a longer period of occupation, from the 8th century BCE to the 8th century CE (*87*, *88*). The site has been identified as Hekatompylos (*41, 88*), the capital of the Parthian Empire and major hub of the Silk Road and Great Khorasan Road. Excavations at Shahr-i-Qumis revealed a very large quantity of equine skeletons. Sample AM805 was radiocarbon dated to 415-542 CE (UCIAMS-223584 and UCIAMS0223188, 1615±20 and 1585±15; Table S2). This places it either during the kingdom of Yazdegerd II (438–457 CE) or his brother Peroz I (457–484 CE). In the beginning of the 5th century CE, nomadic groups (in particular the Hephthalites or White Huns) attacked Persia several times, invading parts of eastern Persia for several years. These events may have also impacted the equine population. A large set of animal bones including an important assemblage of equine bones has been studied by Dr. Marjan Mashkour and Dr. Azadeh Mohaseb from 2002 and later other collaborators (Hossein Davoudi, Homa Fathi, Sansaz Beizaee Doost and Roya Khazaeli) at the British Institute of Persian studies in Tehran (*89*). The assemblage was then transferred to the National Museum of Iran where Azadeh Mohaseb is currently performing a morphometric geometric study of the equid bones.

- **FRA_200-500CE:** Boinville-en-Woëvre, France (samples : GVA125, GVA347, GVA348, GVA349, GVA353, GVA354, GVA355, GVA358, GVA359).

The Gallo-Roman villa of Boinville-en-Woëvre, Déviation Est d'Etain, is located in the department of Meuse, in Northern France. The excavation was carried out in 2005 under the direction of S. Viller (Inrap). Within the Pars Rustica, approximately fifteen pits were discovered, containing 22 complete or sub-complete skeletons of horses and donkeys (9 of which were included in this study, Table S2). Individuals are dated from the Late Antiquity (200-500 CE) (*47*).

- **FRA_0-500CE:** Centre Bourse Marseille, France (samples: BourseB, BourseC)

The equid bones come from ancient excavations carried out in 1968-1969 at the horn of the ancient port of Marseille. They are dated to late Roman times and were studied by Lucien Jourdan who delivered one of the first archaeozoological theses for the Roman period in 1976. Although the chronological resolution is limited, the assemblages can be associated to a complex of carcass deposits accumulated by marine movements between 0-500 CE (*90*). A horse and two donkeys were identified from this site (*47*).

- **TUK_552-987CE:** Yenikapi, Turkey (samples: Tur168, Tur177, Tur179, Tur277)

Yenikapi excavations area is located at the Yenikapi section of Istanbul which lies at the west of Namik Kemal Avenue leading from Aksaray down to the Marmara Sea. The site occupies approximately 58,000 m^2 and covers 1.5 km inlands from the Marmara Sea. During the construction work of the Marmaray and Metro railway project at Yenikapı a large number of antique shipwrecks and animal skeletons were discovered. In the light of these important findings, organized excavations began as early as 2004. The results of the analyses indicate various dates ranging from the early through to the late Byzantine period. About 57 animal species have been identified from the faunal assemblage of the site, and the majority of them are comprised from horse, donkey and mule remains (*91-93*).

- **PTG_1228-1280CE:** Albufeira , Portugal (sample: Albufeira1x1)

This site is in the historic center of Albufeira, on an old peninsula which is surrounded by an inlet to the east and the north. Two silos were found located to the east of the small church of Misericórdia. One silo, that was uncovered during construction work was filled with archaeological material. The finding of coins indicated that this material is no older than the 13^{th} century, during the last phase of Islamic Rule (Almohad Period). The Almohad dominion of Albufeira lasts until 1249 and was the last Alcazaba (city) to be conquered by the Christians. The ceramic materials found at this site are typical from the Almohad period and one of the coins is from the reign of King Afonso III (1248-1279).

Abundant remains of mammological and malacological fauna were identified, including deciduous teeth of a horse and donkey on the top layer of the silo (*94*). The radiocarbon date for the donkey sample (1228-1280 cal. CE; UCIAMS-208877, 765±15 BP, Table S2) suggests its death in the last decades of the Almohad period or shortly after the conquest. However, Islamic people remained in Algarve under the rule of the Christians, so the sample has been considered as Late Islamic.

- **ITA_1683-1936CE:** Fiumarella, Italy (sample: Fiumarella1)

The site of Riparo della Fiumarella di Tortora is located in the valley of the Fiumarella di Tortora stream, close to the modern town of Tortora (Cosenza, Calabria, Southern Italy) and

not far from the Tyrrhenian coast. This strategic position, along one of the routes between the coast and more inland territories, may hint to the importance of the site in the region.

The site was excavated in 2000 by the Soprintendenza Speciale al Museo Preistorico Etnografico "Luigi Pigorini", now part of the Museo delle Civiltà (Rome). The stratigraphic sequence and the archaeological materials evidenced that the site was in use from at least the late Chalcolithic to the MBA (*95*). The chronological and cultural attributions are based on ceramic typology and few bronze artifacts.

The site is now a rock-shelter, but in the past, possibly until the beginning of the MBA, it was a larger cave that collapsed just before the last phases of prehistoric occupation. The relatively small faunal assemblage (n=299) from all the archaeological layers includes mainly domestic mammals, although some remains of red deer and wild boar as well as tortoise were also recovered. Most of the remains represent food refuses although animals probably used for other purposes (e.g., dog, equids) are also present.

Caprine herding represents the main economic activity especially in the MBA when there is a corresponding decrease in the number of pig remains, while cattle rearing was not relevant throughout the archeological sequence. Dogs were extremely rare. Hunting was moderately important during the EBA occupation (20% of the identified specimens). Of particular interest to this study is the presence of two remains of small equids: a femur head from the EBA 2 levels (ca. 1950-1650 BCE), and a third lower molar belonging to a young individual from the MBA 3 - Apennine Culture levels (ca. 1450-1350 BCE). Based on genetic analyses, both specimens were identified as donkey and the latter one was included in the present study due to its high content in endogenous DNA.
The presence of donkeys at such an early date was unexpected because according to current archaeozoological data the earliest occurrence of domestic donkey in Italy is documented only at sites referable to more recent phases of the Bronze age (e.g., Spina, Monte Titano, Coppa Nevigata, Madonna del Petto; (*96-99*)). Therefore, to assess the actual antiquity of the tooth, the specimen was directly dated (UCIAMS-229410, 165±25 BP; Table S2). Unfortunately, the results indicated that the specimen represents modern intrusive material within the Bronze Age levels, however its genetic data have been integrated in this research.

DNA extraction and genome sequencing of ancient samples

The procedures of DNA extraction, library construction and shallow sequencing followed the procedures outlined by Seguin-Orlando and colleagues (*100*) and Librado and colleagues (*21*). The drilling and DNA extractions from osseous material of ancient equids were carried out in the ancient DNA facilities of the Centre for Anthropobiology and Genomics of Toulouse (CAGT), France. Briefly, the methods involved: 1) powdering a total of 100-590mg of osseous material using the Mixel Mill MM200 (Retsch) Micro-dismembrator; 2) extracting the DNA following the procedure outlined by (*101*), which was tailored to facilitate the recovery of even the shortest DNA fragments; 3) treating DNA extracts with the USER™ (NEB) enzymatic cocktail to eliminate a fraction of post mortem DNA damage (*102*); 4) constructing from double-stranded DNA templates DNA libraries in which two internal indexes are added during adapter ligation and one external index is added during PCR amplification; and 5) amplification, purification and quantification of DNA libraries before pooling 20–50 DNA

libraries for low-depth sequencing. After screening for library content using a Miniseq instrument (high-output 80PE mode) at the CAGT (France), sequencing was performed on various Illumina platforms, including HiSeq2500 instruments, at the Centre for GeoGenetics (University of Copenhagen, Denmark) and HiSeq4000 instruments at the Genoscope (Evry, France). Sequence trimming, mapping, filtering and base calibration at damaged sites were carried out following the methodology from Librado and colleagues (*21*).

Radiocarbon dating

Radiocarbon dates were estimated for 14 of the 31 (45%) ancient donkey samples in this study. Dating was carried out at the Keck Carbon Cycle AMS Laboratory, UC Irvine following collagen extraction and ultra-filtration from approximately 1 g of osseous material. IntCal20 calibration (*103*) was performed using OxCalOnline (*104*). Calibrated dates are provided in Table S2. The ages of ancient samples that were not radiocarbon dated were inferred from their established archaeological contexts.

Read alignment, rescaling and trimming

For each raw fastQ file, sequencing reads were demultiplexed, collapsed and trimmed using AdapterRemoval2 (version 2.3.0) (*105*) following the methodology from Gaunitz and colleagues (*50*) for single indexed DNA libraries, and the methodology from Librado and colleagues (*21*) for triple indexed libraries. AdapterRemoval2 also ensured that paired-end reads showing sufficient sequence overlap were collapsed and trimmed (truncated) if ends showed insufficient qualities. Collapsed, truncated and those paired end reads not collapsed (paired) were then parsed through PALEOMIX version 1.2.13.2 (*106*) for Bowtie2 mapping against the donkey mitochondrial (CM027722.1), and nuclear reference sequence (GCA_016077375.1,https://ftp.ncbi.nlm.nih.gov/genomes/genbank/vertebrate_mammalian/Equus_asinus/all_assembly_versions/GCA_016077325.1_EquAsi1.0). Finally, the optimized parameters recommended by Poullet and Orlando (*107*) were considered for mapping, and alignments were locally realigned around indels using the IndelRealigner procedure from GATK (*13*). Sequence alignments shorter than 25 nucleotides, and/or representing PCR duplicates were removed, as well as reads with mapping quality scores inferior to 25.

Subject to trimming, the software mapDamage2 (*24*) was used to check for the presence of nucleotide mis-incorporation profiles characteristic of ancient DNA data at the library level, randomly selecting 100,000 reads. We observed the expected increase of C to T (G to A) mis-incorporation rates at read starts (read ends) for both USER™-treated and non-USER™-treated data, although of lower magnitude for the former, as expected. Furthermore, genomic positions preceding read starts were higher in purines in non-USER™ read alignments, consistently with post-mortem DNA fragmentation being depurination-driven. In USER™-treated read alignments, these positions were enriched in cytosine residues, in line with the excision of deaminated cytosines by the sequential activities of Uracil DNA glycosylase and Endonuclease VIII enzymes present in the USER™ mix. In order to limit the impact of remnant mis-incorporations in downstream analyses, we applied the computational procedure combining end trimming and base quality rescaling based on the post-mortem DNA damage profiles, as described in Seguin-Orlando and colleagues (*100*) and Librado and colleagues (*21*). Briefly, this procedure relies on PMDtools (*108*) to sort read alignments into those likely affected by and those devoid of post-mortem DNA damage. The former alignments were then subjected to base rescaling at those positions likely incorporating nucleotide mis-

incorporations reflecting post-mortem cytosine deamination using mapDamage2 (*109*), before trimming their ends for 10 nucleotides, while the latter were directly subjected to end trimming for 5 nucleotides.

Variant calling pipeline

- **Alignment to the reference genome and rescaling of modern individuals**

We determined the sex of each individual by comparing the relative depth of reads between the autosomes and X chromosomes in the bam files using the "depth" function in SAMtools (version 1.7-12-g17a2483)(*110*). Individuals with a relative depth of 1 between the autosomes and X chromosome were considered to be female and an autosomal depth twice that of the X chromosome were considered to be male (Table S1, S2).

- **Variant calling and quality control filtering of modern individuals**

We called variants (single nucleotide polymorphisms (SNPs) and insertions or deletions of bases (INDELs)) from the mapped and rescaled bam files of modern equids using Graphtyper, running each chromosome in parallel (version 2.5.1) (*61*) (n=45,031,411 variants, Table S3). We then applied the recommended variant filters using the "vcffilter" function from Vcflib (version 1.0) (*111*): ABHet < 0.0, ABHet > 0.33, BHom < 0.0, ABHom > 0.97, MaxAASR > 0.4, MQ > 30. We used GATK (version 4.0.8.1) (*112*) and BCFtools (version 1.8) (*110*) to apply the following genotype filters: Phred score > 20, minor allele frequency (MAF) >= 0.01, Hardy-Weinberg equilibrium p-value >=0.001 and genotype missingness=< 0.2, and conditioning on biallelic variants only. After filtering, we removed the 18 scaffolds with no variants remaining and the sex chromosomes, leaving the variants on the 30 autosomes for further analysis (*n*=13,013,551 variants, Table S4).

- **Generation of the recombination map and phasing of modern individuals**

We selected 25 donkeys to generate a recombination map for all autosomal variants that passed QC filters (*n*=13,013,551). In order to select individuals that provided a representative subset of all subpopulations, we constructed a Principal Component Analysis (PCA) using PLINK (version 1.9) (*63*) with all domestic donkey samples (*n*=206). We finally selected 25 domestic donkeys representing the different geographical locations sampled, so no two individuals were chosen from the same country. In order to prevent selecting individuals with high levels of inbreeding, we estimated levels of inbreeding as runs of homozygosity (ROH) across all autosomes using PLINK (version 1.9) (*63*). Considering that the data used to generate the recombination map were unimputed, we also selected individuals with the lowest proportion of missing SNPs (Table S1).

To calculate the effective population size of the 25 donkeys, we used the formula $N_e = \theta/4\mu$, where μ is the per generation mutation rate, N_e is the effective population size, and θ is the nucleotide diversity. We used a per generation per site μ value of 7.242e-09 as estimated for horses (*113*), assuming a generation interval of 8 years. We calculated theta (θ) for the 25 selected individuals by calling variants using ANGSD (version 0.930) (*114*), conditioning only on variants that passed the previous quality control filters with the parameters: " -GL 1 -C 50 -minQ 25 -minmapq 30 -doMaf 1 -baq 1". We estimated θ as 0.000875 for autosomal variants, and N_e for domestic donkeys as 30,222.

To generate the recombination map, we first calculated the population scaled recombination rate (ρ) between each variant using of LDHat (version 2.2) (*62*). To achieve this, we split each chromosome into overlapping windows of 2,000 variants with an overlap of 200 variants between each window. We generated a log likelihood lookup table for 50 chromosomes for the 25 diploid individuals using the θ estimated using ANGSD with the "complete" function of LDHat. We then estimated ρ for each region using the "intervals" function of LDHat with the parameters: "-its 10000000 -samp 2000 -bpen 5". We discarded the first 20 million burnins and averaged the remaining iterations using the "stat" function of LDHat with the parameter: "-burn 50", before combining the ρ values for each window back into complete chromosomes and converting the ρ values to centimorgans (cM) using the estimated *Ne* value (Table S4, Fig. S1). We found that the average rate of recombination 0.599 cM/Mb per chromosome, which is lower than a previous estimate for horses (1.16 cM/Mb)(*115*), and in the lower range for mammalian species. Next, we used the recombination map to phase missing variants for each individual using BEAGLE (version 5.1) (*39*).

Population genetic analysis of modern donkeys

We used the phased variants to construct PCA analyses using PLINK for three subsets of the population: all individuals (n=222 individuals, Fig. S2), domestic donkeys and *E.a.som* (n=208 individuals, Fig. S3), domestic donkeys only (n=206 individuals, Fig. 1B).

A PCA of all samples (n=222, Fig. S2) showed that domestic donkeys clustered closely together compared to the wild equids, which is consistent with all individuals originating from a single domestication process. The closest wild equid to the cluster of domestic donkeys was *E.a.som*, in agreement with previous findings that donkeys were most likely domesticated from wild African ass species (*11-13*). Early evidence of hunted *Equus a. africanus* at Gebel Gharbi (modern day Libya, radiocarbon dated to 16,750 years ago) suggests a long history of human contact with wild asses in Africa (*116*). However, the absence of the other two African wild ass subspecies in the dataset (*E. a. africanus* or *E. a. atlanticus*) makes it impossible to determine which of these subspecies is genetically closest to the donkey. Interestingly, the 7 kiangs in the dataset separated into two clusters which diverged on the PC2 axis only, which may represent two different subspecies of kiang that have previously been found to be genetically distinct (*22*). Of the two publicly available samples labelled as "Asiatic Wild Ass" (Accession numbers: AW_1 (SRS3167373) and AW_2 (SRS3167374)), one clustered with a group of kiangs and the other was most genetically similar to *E.hemionus*.

The PCA including only domestic donkeys and their closest relative showed *E.a.som* as divergent from the domestic donkeys but closest to East African donkeys (Fig. S3). One donkey from Ethiopia clustered between *E.a.som* and the other domesticates, which is indicative of wild genetic material being present in the genome of this individual (Fig. S3). Additionally, another donkey from Ethiopia and one from Algeria also shifted closer to *E.a.som* compared to PCA plots with domesticates only, also indicating the presence of wild genetic material in these individuals.

Within the domestic donkey population only, we observed strong sub-structuring of donkeys from different geographical locations (Fig. 1B). African donkeys were diverged from the rest of the donkeys on all PCA plots. European donkeys were genetically differentiated on the PC1 axis, with Irish donkeys highly drifted from individuals sampled from mainland Europe. There

was a further spread of donkeys along the bottom half of the PC2 axis moving through Asia with all Chinese, Mongolian and Tibetan donkeys clustering together at the bottom of the PC2 axis.

We also found genetic differentiation between donkeys sampled from the same country. Ethiopian donkeys cluster closely together with other African individuals, except for one donkey clustering close to individuals from the Balkans (Macedonia and Croatia). One individual from Turkey clustered distinctly as well, between Egyptian and European donkeys, so was most likely the product of interbreeding between donkeys from different regions. Additionally, Somalian donkeys form two distinct clusters. Two donkeys cluster with individuals from the neighbouring countries of Ethiopia and Algeria. However, three donkeys are more genetically similar to individuals sampled from Tunisia, Turkey, Syria and Iran, seemingly the result of secondary translocations of donkeys from the Middle East back into this region of the world.

We conducted an admixture analysis for all modern equids using ADMIXTURE (version 1.3.0) (*56*) (Fig. S4). We thinned the variants using the "--indep-pairwise 50 10 0.2 --maf 0.05" parameters in PLINK, leaving 531,322 unlinked variants. We used these variants for ADMIXTURE analysis, with K values between 2-5. The ADMIXTURE analysis showed a distinctive (red) ancestral component that differentiates wild equids from domestic donkeys for all K values. The optimal K value of 4 showed a green ancestral component, which almost completely makes up the genetic material of Irish donkeys with the navy component predominating the genetic makeup of Asian donkeys, with the Kenyan samples showing a yellow ancestral component. The additional (blue) ancestral component at K=5 was predominate in donkeys from the Canary Islands, Spain and Portugal. These findings agree with the substructures seen on the PCA and indicate that genetic drift has occurred in some subpopulations of donkeys, mostly those from more geographically isolated locations such as Ireland, Iberia and the Horn of Africa plus Kenya.

We found that the genomes of all kiangs, onagers and zebras consisted entirely of the red ancestral component (named "wild ancestry"). However, only half the genome of the single *E.a.som* individual only consisted of wild ancestry, which may be due to high levels of inbreeding and genetic drift due to low population size in this species or because it is the closest genetic ancestor to domesticates (*23*, *24*). We found that the red ancestral component was also present in the genomes of some domesticated individuals (named "wild ancestry"). To determine the proportion of wild ancestry in the genome of each domestic donkey, we reran the ADMIXTURE analysis with 100 bootstrap pseudo-replicates. We estimated the average proportion of wild ancestry and the standard deviation for each domestic donkey across the bootstraps Individuals with a standard deviation larger than the average wild ancestry proportion (with ancestry proportion estimates intercepting zero) were assigned a wild ancestry proportion of 0 (Fig. 1A). Donkeys with a proportion of wild ancestry larger than their standard deviation were considered to carry significant admixture proportions and were named "admixed donkeys" (n=20 individuals).

Within the domestic samples, one individual sampled from Ethiopia had a high proportion of wild ancestry (6.99%), and was also identified on the PCA as showing a closer genetic relationship with *E.a.som* compared to the other domesticates. We found measurable levels of

wild introgression in 18 other individuals from Africa and the southern Arabian Peninsula (Yemen and Oman), and one individual from China (Fig. 1A).

To determine which wild equid population contributed wild ancestry to the hybrid donkeys, we constructed qpAdm models (version 810) (*64*). The right (reference) populations consisted of two outgroup domestic donkey populations (determined as donkey populations on different clades to the individual of interest with no admixture from Treemix models and with differential genetic components from the ADMIXTURE analysis) and two wild populations (Table S5). To investigate possible sources of admixture, we selected domestic donkeys that showed a similar genetic makeup to the target individuals based on the ADMIXTURE analysis and a wild equid population as another potential ancestral group.

Population modelling with qpAdm identified the source of wild admixture in all individuals from the Horn of Africa + Kenya and the southern Arabian Peninsula was from a closely related source to *E.a.som*. However, without whole genome sequence data for the other African wild ass species, it was not possible to determine whether this wild admixture occurred from *E.a.som* directly or another sister subspecies. One individual from China showed admixture from kiangs which are a native wild equid species found in the area and may have been the result of human experimentation.

Interestingly, donkeys from Yemen and Oman also showed introgression from African wild asses despite being outside the species historical and current habitat range. This is possibly due to sustained trade of donkeys across the Red Sea with Africa. Additionally, introgression of wild African asses was also found in donkeys sampled from western Africa despite this region also being outside the species historical and current habitat range, which may be due to the wider distribution of African wild asses in the past (*12*). Wild introgression into domestic donkeys is consistent with the extensive reporting of interbreeding between donkeys and wild asses throughout history (*52, 117, 118*), as well as observations in other domesticated species including sheep (*119*) and cattle (*120, 121*). Such practices may have aimed to a further fitness advantage by providing a new phenotype or increasing heterozygosity levels. Further sampling of domestic donkeys in the future would confirm if wild introgression is continuing to occur or if management practices have changed in recent times.

We constructed phylogenetic models using Treemix (version 1.13) (*27*) with 0-5 migration edges for domestic donkeys + *E.a.som* (n=200). We excluded donkeys with the highest levels of wild genetic material (n=6 with over 0.5% wild genetic material, as determined by the ADMIXTURE analysis, and n=2 that were hybrids between multiple subpopulations), as they introduced unnecessary complexity to the graph. Inclusion of these individuals resulted in strong migration edges to the outgroup and each other, making it impossible to see admixture between other groups of donkeys. We grouped the remaining donkeys into subpopulations based on their geographical location, and then thinned the variants using the "--indep-pairwise 50 10 0.2 --maf 0.05" parameters in PLINK (632,429 variants remaining after pruning). We estimated the optimal number of migration edges using a mixed linear model implemented in the optM R package (https://cran.r-project.org/web/packages/OptM/index.html). Using the tree with the optimal number of migration edges (m=3), we estimated bootstrap confidence intervals for each node using modified scripts from the BITE package with 100 pseudo-replicates (*122*) (Fig. 1C).

The Treemix analyses showed distinctive population sub-structuring within domestic donkeys from different geographical locations, with two main branches forming between African (Clade A) and African and non-African donkeys (Clade B), with further differentiation of Asian and European donkeys into separate clusters. The Pega donkey from Brazil was highly divergent but most genetically similar to individuals from the Canary Islands and Iberia. Therefore, the genetic makeup of this rare breed of donkey suggests that is most likely the result of importation of stocks from Iberia during Portuguese colonisation.

With the optimal number of migration edges (m=3) and exclusion of hybrid individuals, there was evidence of shared genetic material between donkeys from the Clade A (Horn of Africa + Kenya and western Africa) with individuals from Sudan (34.5%), which cluster on Clade B. Bootstrapping the tree revealed low confidence at this node (Fig. 1C), which is likely due to the high level of admixture with donkeys from Clade A. Most likely donkeys in this region are bred from stocks sourced from Egypt in the north and other donkey populations in Africa. A migration edge with a lower weight (21.7%) is also observed between the cluster of donkeys from Spain, Portugal, the Canary Islands, Saudi Arabia and Brazil with individuals from western Africa, which likely reflects trade over the Mediterranean, resulting in the importation of donkeys between these regions. Finally, a migration edge between the single donkeys sampled from Saudi Arabia and Brazil (39.8%) was also observed. The genetic similarity between the donkey sampled from Saudi Arabia with the European donkeys compared to others from the Arabian Peninsula (Yemen and Oman) is likely due to translocations of stocks back into this region.

To further elucidate whether modern individuals are derived from one or two domestication processes, we plotted the correlation between the genetic versus the geographic distance of each subpopulation compared to donkeys from Ethiopia (Clade A) and Yemen (Clade B) (Fig. S5). First, we determined regions of the genome contributed by wild ancestors by modelling the admixed individuals in PCAdmix (*123*) with the ancestral populations as determined by ADMIXTURE and qpAdm using the default parameters. We then created a masked VCF file of all domestic donkeys by removing all variants from regions attributed to wild ancestry (n=11,576,248 variants remaining after filtering). We then estimated the genetic distance (f2) between populations using ADMIXTOOLS2 (*124*, *125*) and the geographic distance between populations as the haversine distance using the geosphere package in R (https://cran.r-project.org/web/packages/geosphere/index.html).

To avoid closely related subpopulations confounding regression trends, we excluded those from the same geographic regions which clustered on Treemix with Ethiopia (Kenya and Somalia) and Oman (Yemen). We calculated a separate regression line for individuals from western Africa (Ghana, Mauritania, Nigeria and Senegal), as our demographic trajectories indicated that they split from the subpopulations in the Horn of Africa+ Kenya early on before the expansion out of Africa (Fig. 1D). We also excluded individuals that were translocations back into geographic regions (ALG, BRA, SAU).

We found a strong linear trend of increasing genetic distance verses geographic distance from Ethiopia ($r=0.767$, $r^2=0.460$) and Oman ($r=0.662$, $r^2=0.438$). The strong linear correlations fits with modern donkeys being derived from a single source population similar to Ethiopia, as a break in the trend would indicate that individuals out-of-Africa contained genetic material from another source. The Z-statistic between the coefficients of the two models found no significant

difference (p-value=0.775). The same rate of regression from Oman and Ethiopia further suggests that donkeys expanded out from a single source in Africa into the Arabian Peninsula and then into Eurasia.

To determine the demographic history and split timing of donkey subpopulations, we selected 4 main subpopulations based on Treemix modelling, ADMIXTURE and PCA analysis comprising of individuals from the Horn of Africa + Kenya (Horn+Ken), western Africa (WAfrica), Asia and Europe. We selected three individuals from each subpopulation and converted the variants in the VCF file to SMC++ format, masking regions with wild introgression and tandem repeats using the "vcf2smc" function in the SMC++ package (version 1.15.4) (*28*). We then constructed pseudo-bootstrap replicates of each file by randomly resampling 90% of each chromosome in chunks with 10 replicates based on a modified script from MSMC2 package (*126, 127*), which was developed and implemented by Zheng and colleagues (*127*). We then modelled the population split timing between subpopulations using the split function in SMC++. Next, we obtained the split times from each model using the standard plot function from SMC++ with a generational interval of 8 years (Fig. 1D, S6). We estimated the mean and standard deviation for the split times of each model across the 10 bootstrap pseudo-replicates (Table S14). Additionally, we repeated the same analysis using a different subset of three individuals from each subpopulation to confirm the robustness of the model outputs.

Our demographic modelling using SMC++ showed a decrease followed by a rapid expansion in effective population size for all donkey subgroups around 5,000 BCE, in line with theories that donkeys are derived from a domestication process in Africa around the time of the aridification of the Sahara desert (*1*) (Fig. S6). Further, the models estimated that the first population split occurred between donkeys now found in the Horn of Africa plus Kenya and western Africa, indicative of early genetic isolation occurring within the African continent (Fig. 1D).

Concurrent population split times of European and Asian subpopulation with donkeys from the Horn of Africa plus Kenya indicates a rapid population expansion out of Africa, which suggests that donkeys spread almost simultaneously and extremely rapidly throughout the Old World by the third millennium BCE. This, and the strong phylogeographic structure detected amongst modern populations, indicate that early herders maintained high local reproductive stocks within the areas where donkeys were imported to sustain their further geographic spread. In contrast, effective population size of the donkeys now found in western African only achieved stabilisation around 1,000 years ago.

Imputation of ancient genomes

We imputed the ancient genomes based on the pipeline developed and tested by Hui and colleagues (*38*). In line with this method, we created a reference panel consisting of all modern domestic donkeys (n=206) and variants with a MAF >=0.05. We selected only ancient donkeys with a genome coverage of over 0.75X as candidates for imputation (n=31 individuals, Fig. 3A, Table S2). Before imputation we pseudo-haploidized the ancient individuals using the "dohaplo" flag in ANGSD, conditioning only on positions found in the modern reference panel. We then projected the ancient individuals onto the PCA of modern domesticates using the "lsqproject" function in the smartpca program from the EIGENSOFT package (version 6.1.4) (*26, 57*) (Fig. S7). We found that all ancient individuals clustered closely with the modern

domesticates, indicating that they have a similar genetic makeup and that the reference panel of modern variants can be used for the imputation of the ancient samples.

After confirming that the ancient samples clustered with the modern individuals, we genotyped all variants found in the modern reference panel using ANGSD with the following parameters: "-doMajorMinor 3 -GL 1 -doMaf 1 -snp_pval 1e-6 -doGeno 4 -doPost 1 -postCutoff 0.99 -remove_bads 1 -C 50 -minMapQ 25 -minQ 30 -uniqueOnly 1 -baq 1". After variant calling the genotypes in the ancient samples from the reference panel of variants, we compared the proportion of missing variants to the level of coverage in each sample (Table S7). We found that the level of coverage was approximately inversely proportional to missingness in our ancient samples. The lowest rate of missing variants was 0.558 (55.8%) in a sample with 4.92X coverage and the highest proportion of missing variants was 0.973 for the samples with the lowest level of coverage (0.77X and 0.93X).

We applied a pre-imputation filter of "GP >=0.99" using BEAGLE (version 4.0) to our ancient variant panel We then imputed the genotypes of our ancient individuals with BEAGLE (version 5.1), using only the filtered variants, the reference panel of modern donkeys and the recombination map previously generated. We reapplied the filter "GP >=0.99" post-imputation (*n*=7,161,029 variants (TI/TV=2.17), and *n*=2,245,992 variants (TI/TV=2.21) that were present in all ancient individuals after post-imputation filtering). We then merged the variants from ancient and modern individuals into a single file using the "merge" function in BCFtools.

To examine the accuracy of this method on the imputation of donkey genomes, we randomly knocked out an increasing proportion of variants (0.2, 0.5, 0.5, 0.9, 0.92, 0.94, 0.96 and 0.99) from ten modern individuals with the lowest rates of missing SNPs (pre-phasing and excluding the donkey that was used for the reference genome). We then re-imputed the variants for these individuals using the same imputation pipeline as outlined above. and after filtering, compared them with the original variants for the same sample to measure the accuracy of imputing samples with different rates of missingness (Fig. S8). Based on this imputation accuracy test, we predicated that all samples have an overall imputation accuracy between 98.1% and 98.6% (Fig. S8, Table S6).

After imputation, we projected the ancient, imputed samples onto the PCA with the non-imputed, pseudo-haploidized data for the same ancient donkeys and the modern donkeys used in the reference dataset (Fig. S7). We found that after imputation each ancient individual clustered very similarly to the non-imputed data, albeit moving away from the 0,0-axis due to more data being available (including heterozygous variants). This further provided an indication that the imputation did not change the genetic makeup of the ancient samples relative to the modern individuals, but helped gain resolution.

To test for the effects of post-mortem damage on the accuracy of imputation in ancient samples, we genotyped alleles for the ancient donkey with the highest coverage (GVA348, 5.05X), using ANGSD and conditioning on sites with a coverage of at least 8X ("setMinDepth 8").We then compared these genotyped alleles to the imputed variants and found that we recovered the same alleles for 99.99% of sites (541,969 out of 541,981 sites), further providing evidence that our method is highly accurate for imputing variants in samples with post-mortem damage.

Population genetic analysis using imputed variants

We performed an ADMIXTURE analysis conditioning on all modern equids and ancient donkeys using imputed variants (Fig. S4). We first thinned all imputed autosomal variants in PLINK using the parameters: "--indep-pairwise 50 10 0.2 --maf 0.05", then calculated admixture proportions for models with K values between 2 and 5 using ADMIXTURE (*n*=253

individuals and *n*=494,050 variants after filtering). An optimal K value of 4 was estimated by comparing the cross-validation values of the different models.

PCA analysis showed that ancient donkeys clustered most closely with modern donkeys, and also showed a similar genetic makeup on the ADMIXTURE analysis. However, an ancient donkey from Israel (MV242; Nizzana, 350-58BCE) showed high amounts of ancestry from a divergent wild outgroup. Bootstrapped ADMIXTURE (100 bootstrap pseudo-replicates) found that MV242 contained 4.15±0.19 % wild genetic material (Fig. 2C).

We conducted a haplotype-based clustering analysis of all modern and ancient domestic donkeys using fineSTRUCTURE (version 4.1.1) (*35*). We converted the variants in the VCF file and the recombination map present in all individuals (*n*=2,245,992) to the required input file formats using custom R scripts and the provided perl scripts from the fineSTRUCTURE package. We excluded 58 Chinese and Tibetan donkeys so as to avoid overrepresenting this region. Additionally, we removed modern individuals that were identified in the previous ADMIXTURE analysis as having a high proportion of wild admixture (*n*=6) and admixture between different populations (*n*=2), which were found to confound the output, resulting in a final dataset of 172 individuals. FineSTRUCTURE was run with default parameters to paint the chromosomes and model haplotype sharing between individuals. The maximum likelihood tree and co-ancestry matrix was plotted from the output files using modified versions of the R scripts provided with the fineSTRUCTURE package (Fig. 2A, 2B, S9).

To estimate the genetic sharedness between each ancient individual with the modern subpopulations, we calculated outgroup f3-statistics in the form of (modern, ancient; kiang) using ADMIXTOOLS2 (*58, 65*), using only variants present in all individuals (*n*=2,245,992).We used the mean and standard error from the outgroup f3-statistics to plot a heatmap comparing relatedness between the ancient individuals to the modern populations (Fig. 3B).

To further confirm the genetic makeup of our ancient individuals, as inferred by fineSTRUCTURE analysis and outgroup f3-statistics, we constructed Treemix models using the imputed matrix with variants present in all individuals, first pruning the matrix in PLINK using the parameter "--indep-pairwise 50 10 0.2" (*n*=175,093 variants after filtering). In accordance with earlier Treemix models (Fig. 1B), we removed modern donkeys with high proportions of wild admixture or that were hybrids between different regions, and included *E.a.som,* with the kiangs as an outgroup (*n*=207 modern individuals). We then grouped modern donkeys according to the branches on Fig. 1B into HORN+KEN (ETH,KEN,SOM), WAFR (GHA, MAU,NIG, SEN), SAPEN (OMA,YEM), CASIA (TKM,KYR,KAZ), EASIA (CHI,TIB,MON), IRA (IRA), TTS (TUK, TUN, SYR), NUBIA (EGY,SUD), EEUR (YUM,YUC), IRE (IRE,Eas), and WEUR (ESP,PTG,CYK,BRA). Ancient donkeys were added to the Treemix model separately, grouped according to their archaeological site (Table S2, Fig. S10). However, in two sites, fineSTRUCTURE analysis showed potentially different genetic makeup in individuals from Yenikapi and Shahr-i-Qumis, so were modelled separately. Each Treemix model was run for 0-10 migration events with 5 replicates and a k value of 1000. The optimal migration edges were inferred using optM, and the 100 bootstraps were preformed using the BITE package as above (Fig. 2C, 2D, 2E 3C, 3D, 3E, S10).

A deletion in *TBX3* has been found to be responsible to the phenotypic change from a grey dun coat to a coloured coat in donkeys (*13*). A single nucleotide deletion in the *TBX3* gene (CT>C-

) results in derived coat colours in homozygous individuals, which has previously been annotated to JADWZW010000009.1:42742556 on this version of the assembly (*13*). We genotyped all ancient and modern individuals in our dataset. As a confirmation of the validity of this genotyping, we found that all wild individuals were genotyped for the dun coat colour, but the reference individual (a black donkey) was genotyped for a derived coat colour.

With a post-imputation filter of GP>=0.99, 19 out of 31 ancient donkeys were genotyped for the TBX3 locus. However, with a GP>=0.9, the *TBX3* genotype of 25 ancient individuals could be inferred. Coat color phenotypes in ancient donkeys showed that derived coat colors were present across multiple locations, ranging from western Asia (Iran, Shahr-i-Qumis) to Iberia (Portugal, Albufeira) (Fig. 3A). Colored coats appeared almost simultaneously in our dataset in samples from Shahr-i-Qumis and Boinville-en-Woëvre. However, one of our oldest samples (Chalow3) was heterozygous, indicating that this variant was segregating in donkeys by at least this time (~2050BCE). The presence of black donkeys have been recorded in Iraq (Assur) in the 2nd millennium BCE, which further suggests that the mutation in the *TBX3* gene was present in early donkey populations (*128*). Derived coat colors appeared at high frequencies in modern domesticates out of Africa, indicating that selection in more modern times may have favored derived coat colors in donkeys in some regions of the world (Fig. S11).

However, we found that variants underlying long hair and white spots were not present in our phased variant panel for modern donkeys. Two recessive mutations in the *FGF5* gene have been associated with long hair in donkeys (*46*). The missense mutation (G>A) was mapped to JADWZW010000004.1:161390091 and a frameshift deletion (delAT) to JADWZW010000004.1:161397694 on the reference genome used in this study. Additionally, a dominant mutation associated with white spotting has been identified in splice donor site in the KIT gene (T>A, JADWZW010000004.1:139925278) (*45*). Analysis of sequence alignments of the 31 ancient donkeys did not find any individuals homozygous for either *FGF5* mutation, although one individual was heterozygous for the missense mutation (AM89) (Table S7), and another for the deletion (Tur179) (Table S8). This indicates that these mutations were segregating in ancient donkeys, but likely reached higher frequency in some modern breeds at later dates. None of the ancient donkeys carried the mutation associated with white spotting, suggesting that this phenotype was not commonly found in the past (Table S9).

To gain insights into the breeding management of ancient donkeys, we estimated the level of relatedness between ancient donkeys from the same site using KING (version 2.2.7) (*66*) on the panel of imputed variants, conditioning on transversions that were common across all individuals (*n*=31 individuals, *n*=619,981 transversions, Table S10). We found evidence of close familial relatedness between 6 donkeys from Boinville-en-Woëvre. Two other donkeys from this site had a high level of genetic relatedness, indicative of full siblings. Additionally, the two donkeys from Tarquinia showed a 4th degree of genetic relatedness. No close genetic relatedness was inferred between donkeys at any other site. However, ancient donkeys from the same site may be from different generations, which could explain the lack of genetic relatedness between them.

Errors in imputation may lead to over- or underestimates of relatedness between ancient individuals. Therefore, we also estimated the relatedness between modern and ancient donkeys using NgsRelate (version 2) (*67*). Variants were first called using ANGSD for all modern and ancient donkeys (*n*=238) separately for each chromosome with the following parameters: "-

baq 1 -doCounts 1 -C 50 -skipTriallelic 1 -doMajorMinor 1 -SNP_pval 1e-6 -doMaf 1 -rmTriallelic 1e-4 - -minQ 30 -minMapQ 25 -uniqueOnly 1 -remove_bads 1 -doPost 1 -beagleProb 1 -doGlf 2 -GL 2 -P 2 -MAF 0.05", with sites covered in at least 75% of individuals. Transitions were removed and the separate chromosome files were merged together, before running NgsRelate (*n*=473,263 variants). High correlations between the KING coefficient estimated using NgsRelate and the IBD coefficient estimated for the phased and imputed data using the KING software (r=0.871, r^2=0.759) showed that accurate relationship inferences could be inferred using imputed data (Fig. S12).

We estimated inbreeding as runs of homozygosity (ROH) for all modern and ancient donkeys using three methods. First, using PLINK using the "--homozyg" function for all imputed transversions (*n*=238 individuals, *n*= 1,949,850 transversions), with a cut off length of at least 1 MB. Estimating runs of homozygosity requires dense haplotypes, however imputation errors in the low-coverage ancient samples may lead to inaccurate calculations of inbreeding levels. To account for imputation errors which may break up ROHs, we allowed for up to 4 heterozygous variants in each 50 SNP sliding window (Fig. S13A).

We examined the effects of imputation errors on ROH estimations using imputed variants in PLINK by down-sampling and re-imputing 10 high coverage modern donkey genomes: 5 with the highest ROH and 5 with the lowest total length of ROH (as estimated by PLINK). We found little change in the total length of ROH when up to 96% of variants were knocked out and re-imputed, which was the highest rate of missingness in our ancient samples (Fig. S14A). This agrees with the estimations of high imputation accuracy in these samples and provides evidence that ROHs can still be inferred using PLINK with a low rate of errors.

To further test the robustness of the imputed data in accurately estimating ROHs in our ancient samples, we also estimated ROHs using the method implemented NgsF-HMM (version 1) (*59*) on the unimputed data from all modern and ancient donkeys (*n*=238 individuals). We estimated ROHs using NgsF-HMM, using the same files as generated for NgsRelate (*n*= 473,263 variants), and using a minimum epsilon of 1e-8. We then filtered the ROHs to only select those with a total length over 1MB, containing more than 100 SNPs and with at least one SNP per 50KB on average, in line with the parameters defining an ROH in PLINK (Fig. 4A, Fig. 4B).

We also estimated ROHs from the bam files of the modern and ancient donkeys by searching for regions with a low density of heterozygous variants. First, we down sampled the bam file for each modern and ancient donkey to the lowest coverage sample in our dataset (0.77X) using SAMtools. Next, we generated counts files using ANGSD with the parameters: "-doCounts 1 -dumpCounts 4", conditioning only on sites with a MAF >=0.05 in modern donkeys. We then filtered the sites for each individual for a depth greater than 2, then grouped the remaining sites into bins of 200 SNPs. Bins with less than 6 heterozygous variants (a frequency of 0.03) were considered to be a ROH. These parameters were optimised by comparing the size and distribution of ROHs in high coverage modern individuals to those estimated in PLINK. We then summed the length of all ROH bins together to obtain the total proportion of the genome in ROH for each individual (Fig. S13B). We then compared the total ROH in the genome of each individual to that estimated by PLINK and ngsF-HMM. The three methods showed high correlation, indicating that the estimates were robust to imputation or phasing errors (Fig. S14).

We plotted the total length of ROH in the genome of each donkey as a function of time for each of the three methods (Fig. 5B, Fig. S13), separating the modern donkeys by continent and

grouping the ancient donkeys by site and inferring their age through radiocarbon dates where available or the archaeological context of the sample. Visually, little change was seen in the overall proportion of ROH in the genomes of modern versus ancient donkeys. A Wilcoxon rank sum test using the NgsF-HMM output confirmed that there was no significant difference in the total length of ROH between the two groups (W=2904 *p*-value=0.395, *n*=238) (Fig. 5A). In line with their close familial relationships, a Wilcoxon rank sum test determined that the five donkeys from Boinville-en-Woëvre had significantly higher proportions of their genomes in ROH compared to the other ancient individuals (Wilcoxon rank sum test, W = 139, *p*-value = 0.045, *n*=31).

Next, we estimated ROH from publicly available whole genome sequences of 75 ancient and 79 modern horses, using NgsF-HMM with the same method as for donkeys (Table S11, Fig. 4C, 4D, *n*=963,418 transversions). A Wilcoxon rank sum test confirmed that modern horses were more significantly inbred than ancients (W= 4541, *p*-value>0.001, *n*=154), in contrast to donkeys (Fig. 5C). The total ROH for each horse was plotted as a function of time, as for donkeys (Fig. 5D).

Pseudo-haploidized matrix

Variation in ancient individuals that is not represented in modern populations may affect the accuracy of population models conditioning on modern variation only. To confirm the accuracy of our analyses using imputed ancient genomes that were conditioned on modern variation, we constructed a pseudo-haploidized matrix for the ancient and modern individuals included in the Treemix analysis, following the procedure from Gaunitz and colleagues (2018) and Librado and colleagues (2021) (*21, 50*). Variants were called in ANGSD with the parameters: "-minQ 20 -minMapQ 25 -remove_bads 1 =uniqueOnly 1 -baq 1 -C 50 -doHaploCall 1", conditioning only on transversions (*n*=4,833,570 transversions). We used this matrix for Treemix analyses using the same method as above, LD pruning the variants (*n*=496,697 after pruning). We added ancient donkeys from each site to the Treemix models separately, then estimating the optimal number of migration edges and performed 100 bootstrap pseudo-replicates for each model. We found that placement on ancient donkeys on the Treemix models constructed using imputed and pseudo-haploidized data was highly similar, confirming the accuracy of our imputation panel (Fig. S10). Next, we constructed a neighbour joining tree to further confirm the population structure of the modern and ancient donkeys. We first calculated pairwise genetic distances between all samples using PLINK, then retrieved the tree topology by implementing the bioNJ algorithm in FastME (version 2.1.4)(*129*), with 100 bootstrap pseudo-replicates to assess node supports (Fig. S15).

The genome of MV242 was found to contain divergent genetic material, as confirmed by ADMIXTURE analysis and Treemix phylogenies using imputed data (Fig. 2E, 3A, S4, S10, S15). However, because there may be errors in the imputed haplotypes of this individual due to the divergent genetic makeup, we used pseudo-haplodized data for further analysis. We modelled f4(*E.a.som*, MV242; HORN+KEN, x) statistics to determine whether genetic material from this lineage was present in modern donkey subpopulations (x) using qpDstat (version 751) from the Admixtools package (*58, 65*). We grouped modern donkeys into the same subpopulations used on the Treemix models (Fig. S10). *P*-values were obtained through multiple test correction of Z-scores with a significance threshold of 0.05. Positive and

significant f4-statistics provided evidence of MV242 ancestry in modern donkeys from eastern Asia, Nubia, central Asia, Turkey, Syria, Tunisia, Iran and western Europe (Fig. 5E).

Next, we tested for the presence of genetic material in the ancient donkeys with f4(*E.a.som*, MV242; Fiumarella1, x) statistics, where x are the ancient donkeys grouped by site according to the Treemix models (Fig. S10). An excess of sharedness with the MV242 lineage was found in the individual Chalow3 as the f4-statistics were positive and significant (Fig. 5F). However, significantly negative f4-statistics showed a deficit in sharedness in a family group of 6 donkeys from Boinville-en-Woëvre (GVA125, GVA347, GVA348, GVA349, GVA353, GVA354) (Fig. 5G), which showed evidence of wild genetic material in ADMIXTURE analysis (Fig. 3A). To determine whether this wild genetic material is derived from a source more divergent than MV242 we tested f4(kiang, MV242; Fiumarella1, x) statistics, where x are the three family groups from Boinville-en-Woëvre. This statistic was negative and significant for family group GVA1 only, which supports restocking in this population from a lineage more divergent than MV242. The f4(kiang, *E.a.som*; Fiumarella1, x) statistics, for the family groups at Boinville-en-Woëvre were balanced, which suggests that this wild genetic material is not from a population more divergent than *E.a.som* (Fig. 5H).

Uniparental markers

To construct the mitochondrial phylogeny, we called variants with "-doHaploCall 1 -minMapQ 25 -minQ 30 -doDepth 5" using ANGSD. Additionally, we included the mitochondrial genomes of three *Equus hemionus hemippus* (accession numbers: ERS7669491, ERS7669492, ERS7669493) (*20*) (*n*=2,805 variants, *n*=256 individuals. We generated a tree with IQ-TREE (version 1.6.12) (*60*), using 100 bootstrap pseudo-replicates for assessing node support (Fig. 5A). The tree was rooted between the zebras and hemiones+ kiangs, as per Jónsson and colleagues (*24*).

To construct the Y-chromosome phylogeny, we called variants using ANGSD with the parameters: "-isHap 1 -baq 1 -remove_bads 1 -uniqueOnly 1 -minMapQ 25 -minQ 30 -rmTriallelic 1e-4-SNP_pval 1e-6 -C 50" for all male equids in our dataset (*n*=125), conditioning on transversions only and including only variants present in more than 90% of individuals, leaving a total of 3,171 variants in the final dataset. We generated a tree with IQ-TREE (version 1.6.12) (*60*), using the same parameters as those used to generate the mitochondrial tree (Fig. 5B).

To estimate the time to the most recent common ancestor (TMRCA), we constructed Bayesian skyline plots using mitochondrial and Y-chromosome variation of domestic donkeys only (*n*=238 and 121 individuals, respectively) using BEAST (version 2.6.5) (*68-70*). We estimated the optimal substitution model for both datasets using the BIC scores estimated from IQ-TREE. we converted the multi-alignment fasta files to BEAST input files using BEAUTi (version 2.5.26) (*68-70*) specifying the following parameters: 1) the optimal model for all three datasets was GTR, with an empirical distribution and a gamma category count of 4. 2) Tips of the ancient individuals were dating in years before present using radiocarbon dates, where available, or the mean of the time period estimated from archaeological context. For ancient donkeys from Shahr-i-Qumis, their age was inferred from the single individual radiocarbon dated at this site (AM805). 3) Selecting the Bayesian skyline demographic model and uncorrelated log-Normal relaced molecular clocks with mean values= [1e-07] per site per year [sampling from a uniform prior between 1e-08 and 1e-05]. BEAST (version 2.5.1) (*68-70*) was

829 run for a total of 500,000,000 iterations for Y-chromosomal and 350,000,000 for mitochondrial
830 reconstructions. The posterior distributions of the tree heights were generated using Tracer
831 (version 1.7.1) (*130*) with 20% as burn-in (Fig. 4B, D).

832

Table S1: Sample information for all modern donkeys and wild equids (*n*=222). The country of origin, short country code, genome depth-of-coverage, the proportion of missing variants after variant calling and accession number are reported. Accessions numbers starting with "SRS" were downloaded from the National Library of Medicine, "ERS" from the European Nucleotide Archive, and those starting with "SAMC" from the Genome Sequence Archive database.

ID	Species	coverage	Proportion missing variants	sex	Country	short country code	Accession
ALG_01	*Equus asinus*	25.378	0.008	M	Algeria	ALG	ERS12239254
IRE_EnglishWpureIrish_1	*Equus asinus*	10.369	0.271	F	Ireland	IRE	SRS3167383
IRE_EnglishWpureIrish_2	*Equus asinus*	9.086	0.35	F	Ireland	IRE	SRS3167384
IRE_pureIrish_3	*Equus asinus*	8.589	0.377	F	Ireland	IRE	SRS3167387
IRE_pureIrish_4	*Equus asinus*	8.33	0.402	F	Ireland	IRE	SRS3167408
IRE_pureIrish_5	*Equus asinus*	9.339	0.344	F	Ireland	IRE	SRS3167409
IRE_pureIrish_6	*Equus asinus*	10.265	0.28	F	Ireland	IRE	SRS3167406
IRE_pureIrish_7	*Equus asinus*	9.055	0.367	F	Ireland	IRE	SRS3167407
IRE_pureIrish_8	*Equus asinus*	9.021	0.369	M	Ireland	IRE	SRS3167410
Aw_1	Asiatic wild ass	11.057	0.178	M	NA	AW	SRS3167373
Aw_2	Asiatic wild ass	11.464	0.163	F	NA	AW	SRS3167374
CHI_dz	*Equus asinus*	150.067	0.003	F	China (plain)	CHI	SRS7835299
CHI_Guangling_3	*Equus asinus*	9.817	0.28	M	China (plain)	CHI	SRS3167352
CHI_Guangling_4	*Equus asinus*	10.497	0.23	F	China (plain)	CHI	SRS3167350
CHI_HetianGray_1	*Equus asinus*	10.168	0.253	M	China (plain)	CHI	SRS3167356
CHI_HetianGray_2	*Equus asinus*	9.831	0.274	F	China (plain)	CHI	SRS3167354
CHI_HetianGray_3	*Equus asinus*	9.527	0.292	F	China (plain)	CHI	SRS3167361
CHI_HetianGray_4	*Equus asinus*	9.993	0.253	F	China (plain)	CHI	SRS3167381
CHI_BY02A	*Equus asinus*	13.477	0.105	F	China (plain)	CHI	SRS3167450
CHI_BY03A	*Equus asinus*	11.341	0.176	F	China (plain)	CHI	SRS3167463
CHI_BY06A	*Equus asinus*	12.584	0.11	M	China (plain)	CHI	SRS3167461
CHI_BY07A	*Equus asinus*	10.626	0.211	M	China (plain)	CHI	SRS3167462
CHI_GL03A	*Equus asinus*	11.589	0.152	F	China (plain)	CHI	SRS3167460
CHI_GL04A	*Equus asinus*	12.789	0.119	F	China (plain)	CHI	SRS3167357
CHI_HL06	*Equus asinus*	5.968	0.641	F	China (plain)	CHI	SAMC048978
CHI_HL28	*Equus asinus*	6.461	0.571	M	China (plain)	CHI	SAMC048979
CHI_HL29	*Equus asinus*	6.229	0.597	M	China (plain)	CHI	SAMC048980
CHI_JM01A	*Equus asinus*	11.482	0.152	F	China (plain)	CHI	SRS3167380
CHI_JM05A	*Equus asinus*	10.701	0.183	M	China (plain)	CHI	SRS3167379
CHI_JM06A	*Equus asinus*	11.65	0.149	F	China (plain)	CHI	SRS3167378
CHI_JM07A	*Equus asinus*	11.037	0.174	M	China (plain)	CHI	SRS3167377
CHI_JM11A	*Equus asinus*	10.004	0.24	M	China (plain)	CHI	SRS3167392
CHI_KL02A	*Equus asinus*	10.818	0.192	F	China (plain)	CHI	SRS3167391
CHI_KL03A	*Equus asinus*	13.283	0.105	F	China (plain)	CHI	SRS3167389

CHI_KL04A	*Equus asinus*	11.804	0.147	F	China (plain)	CHI	SRS3167390
CHI_KL05A	*Equus asinus*	9.689	0.266	F	China (plain)	CHI	SRS3167388
CHI_XJ1	*Equus asinus*	5.939	0.635	F	China (plain)	CHI	SAMC049000
CHI_XJ2	*Equus asinus*	5.798	0.65	F	China (plain)	CHI	SAMC049001
CHI_XJ3	*Equus asinus*	5.891	0.641	F	China (plain)	CHI	SAMC049002
CHI_XJ5	*Equus asinus*	7.545	0.476	F	China (plain)	CHI	SAMC049003
CHI_XJ6	*Equus asinus*	6.481	0.576	F	China (plain)	CHI	SAMC049004
CHI_YM01	*Equus asinus*	6.365	0.591	M	China (plain)	CHI	SAMC049023
CHI_YM04	*Equus asinus*	6.346	0.588	M	China (plain)	CHI	SAMC049024
CHI_YM05	*Equus asinus*	5.749	0.655	F	China (plain)	CHI	SAMC049025
CHI_YM12	*Equus asinus*	5.585	0.676	F	China (plain)	CHI	SAMC049026
CHI_Qingyang_1	*Equus asinus*	10.453	0.228	M	China (plain)	CHI	SRS3167413
CHI_Qingyang_2	*Equus asinus*	10.786	0.224	F	China (plain)	CHI	SRS3167414
CHI_Qingyang_3	*Equus asinus*	10.631	0.239	M	China (plain)	CHI	SRS3167415
CHI_Qingyang_4	*Equus asinus*	10.647	0.251	M	China (plain)	CHI	SRS3167411
CHI_Turfan_1	*Equus asinus*	10.508	0.246	F	China (plain)	CHI	SRS3167412
CHI_Turfan_2	*Equus asinus*	8.709	0.386	F	China (plain)	CHI	SRS3167424
CHI_Turfan_3	*Equus asinus*	7.963	0.421	M	China (plain)	CHI	SRS3167423
CHI_Turfan_4	*Equus asinus*	8.699	0.367	F	China (plain)	CHI	SRS3167422
CHI_Turfan_5	*Equus asinus*	7.969	0.441	F	China (plain)	CHI	SRS3167421
CHI_Xinjiang_1	*Equus asinus*	11.005	0.215	F	China (plain)	CHI	SRS3167454
CHI_Xinjiang_2	*Equus asinus*	9.063	0.336	M	China (plain)	CHI	SRS3167473
CHI_Xinjiang_3	*Equus asinus*	9.652	0.285	M	China (plain)	CHI	SRS3167474
CHI_Xinjiang_4	*Equus asinus*	8.417	0.393	M	China (plain)	CHI	SRS3167475
CHI_Xinjiang_5	*Equus asinus*	9.763	0.3	F	China (plain)	CHI	SRS3167470
CHI_Yunnan_1	*Equus asinus*	10.606	0.247	M	China (plain)	CHI	SRS3167371
CHI_Yunnan_2	*Equus asinus*	10.24	0.271	F	China (plain)	CHI	SRS3167369
CHI_Yunnan_3	*Equus asinus*	10.891	0.212	M	China (plain)	CHI	SRS3167370
Easi_Willy2	*Equus asinus*	28.839	0.025	M	Denmark	Eas	SRS431817
Eboe_0227A	*Equus burchelli*	22.537	0.142	F	NA	Eboe	ERS559290
EGY_1	*Equus asinus*	11.642	0.166	F	Egypt	EGY	SRS3167452
EGY_2	*Equus asinus*	10.872	0.225	F	Egypt	EGY	SRS3167456
EGY_3	*Equus asinus*	10.963	0.221	M	Egypt	EGY	SRS3167455
EGY_4	*Equus asinus*	7.997	0.436	M	Egypt	EGY	SRS3167382
EGY_5	*Equus asinus*	7.275	0.506	F	Egypt	EGY	SRS3167358
Egre_0228A	*Equus grevyi*	18.65	0.126	F	NA	Egre	SRS1208552
EGY_155	*Equus asinus*	14.498	0.136	M	Egypt	EGY	SRS3167349
EGY_161	*Equus asinus*	8.868	0.39	M	Egypt	EGY	SRS3167353
EGY_169	*Equus asinus*	8.949	0.368	M	Egypt	EGY	SRS3167359
EGY_02	*Equus asinus*	27.085	0.007	M	Egypt	EGY	ERS12239300
EGY_14	*Equus asinus*	35.048	0.004	F	Egypt	EGY	ERS12239301
EGY_17	*Equus asinus*	29.792	0.005	M	Egypt	EGY	ERS12239302

Ehar_0229A	*Equus hartmannae*	19.301	0.144	F	NA	Ehar	SRS861660
Ekia_0231A	*Equus kiang*	14.887	0.125	F	NA	Eki	SRS861663
Ekiang_XZYL	*Equus kiang*	7.79	0.587	F	NA	Eki	SAMC049022
Ekiang_YP21	*Equus kiang*	6.723	0.679	F	NA	Eki	SAMC049027
Ekiang_ZYL	*Equus kiang*	6.649	0.667	F	NA	Eki	SAMC049051
Ekiang_kun1	*Equus kiang*	5.825	0.636	M	NA	Eki	SAMC048991
Ekiang_kun2	*Equus kiang*	6.132	0.592	F	NA	Eki	SAMC048992
Eona_0230A	*Equus hemionus*	21.498	0.102	M	NA	Eon	SRS474403
Eona_0261A	*Equus hemionus*	9.497	0.29	F	NA	Eon	SRS693024
Esom_0226A	*Equus africanus somaliensis*	25.869	0.087	F	NA	Esom	SRS861674
ETH_1	*Equus asinus*	9.15	0.291	M	Ethiopia	ETH	SRS3167398
ETH_10	*Equus asinus*	8.783	0.352	M	Ethiopia	ETH	SRS3167432
ETH_2	*Equus asinus*	8.508	0.388	M	Ethiopia	ETH	SRS3167445
ETH_3	*Equus asinus*	8.565	0.379	M	Ethiopia	ETH	SRS3167444
ETH_4	*Equus asinus*	9.482	0.289	M	Ethiopia	ETH	SRS3167442
ETH_5	*Equus asinus*	8.008	0.478	M	Ethiopia	ETH	SRS3167443
ETH_6	*Equus asinus*	8.088	0.41	M	Ethiopia	ETH	SRS3167430
ETH_7	*Equus asinus*	8.006	0.426	F	Ethiopia	ETH	SRS3167433
ETH_8	*Equus asinus*	9.409	0.29	F	Ethiopia	ETH	SRS3167428
ETH_9	*Equus asinus*	8.617	0.362	M	Ethiopia	ETH	SRS3167431
ETH_14B	*Equus asinus*	32.065	0.005	M	Ethiopia	ETH	ERS12239255
ETH_5B	*Equus asinus*	35.773	0.051	M	Ethiopia	ETH	ERS12239256
ETH_6B	*Equus asinus*	28.247	0.037	M	Ethiopia	ETH	ERS12239257
GHA_01	*Equus asinus*	20.988	0.014	M	Ghana	GHA	ERS12239258
GHA_07	*Equus asinus*	29.092	0.005	F	Ghana	GHA	ERS12239259
IRA_D2	*Equus asinus*	10.35	0.423	F	Iran	IRA	SAMC048970
IRA_D7	*Equus asinus*	6.842	0.572	M	Iran	IRA	SAMC048971
IRA_D9	*Equus asinus*	7.274	0.546	M	Iran	IRA	SAMC048972
IRA_D10	*Equus asinus*	10.796	0.338	M	Iran	IRA	SAMC048965
IRA_D11	*Equus asinus*	6.628	0.613	M	Iran	IRA	SAMC048966
IRA_D13	*Equus asinus*	9.359	0.46	F	Iran	IRA	SAMC048967
IRA_D14	*Equus asinus*	12.299	0.267	M	Iran	IRA	SAMC048968
IRA_D16	*Equus asinus*	10.518	0.332	F	Iran	IRA	SAMC048969
KAZ_04	*Equus asinus*	27.202	0.006	F	Kazakhstan	KAZ	ERS12239260
KAZ_07	*Equus asinus*	29.988	0.005	F	Kazakhstan	KAZ	ERS12239261
KEN_YPO86	*Equus asinus*	12.269	0.262	M	Kenya	KEN	SAMC049038
KEN_YPO97	*Equus asinus*	10.148	0.452	F	Kenya	KEN	SAMC049048
KEN_YPO98	*Equus asinus*	8.334	0.491	M	Kenya	KEN	SAMC049049
KEN_YPO101	*Equus asinus*	7.208	0.6	M	Kenya	KEN	SAMC049028
KEN_YPO102	*Equus asinus*	8.117	0.544	M	Kenya	KEN	SAMC049029

KEN_YPO104	*Equus asinus*	8.768	0.391	F	Kenya	KEN	SAMC049030
KEN_YPO106	*Equus asinus*	11.438	0.261	M	Kenya	KEN	SAMC049031
KEN_YPO89	*Equus asinus*	12.803	0.231	M	Kenya	KEN	SAMC049041
KEN_YPO99	*Equus asinus*	8.575	0.494	M	Kenya	KEN	SAMC049050
KEN_YPO87	*Equus asinus*	9.529	0.353	M	Kenya	KEN	SAMC049039
KEN_YPO88	*Equus asinus*	10.558	0.353	M	Kenya	KEN	SAMC049040
KEN_YPO90	*Equus asinus*	12.444	0.275	F	Kenya	KEN	SAMC049042
KEN_YPO91	*Equus asinus*	11.092	0.335	M	Kenya	KEN	SAMC049043
KEN_YPO92	*Equus asinus*	11.048	0.36	M	Kenya	KEN	SAMC049044
KEN_YPO93	*Equus asinus*	12.105	0.191	M	Kenya	KEN	SAMC049045
KEN_YPO95	*Equus asinus*	10.662	0.34	M	Kenya	KEN	SAMC049046
KEN_YPO96	*Equus asinus*	12.505	0.27	M	Kenya	KEN	SAMC049047
Kia_1	*Equus asinus*	27.447	0.067	F	NA	Eki	SRS3167376
KYR_Sdonk3	*Equus asinus*	8.328	0.478	F	Kyrgyzstan	KYR	SAMC048996
KYR_Sdonk6	*Equus asinus*	29.255	0.487	F	Kyrgyzstan	KYR	SAMC048997
KYR_Sdonk7	*Equus asinus*	8.071	0.585	M	Kyrgyzstan	KYR	SAMC048998
KYR_Sdonk9	*Equus asinus*	6.826	0.28	F	Kyrgyzstan	KYR	SAMC048999
KYR_Sdonk12	*Equus asinus*	5.903	0.006	M	Kyrgyzstan	KYR	SAMC048993
KYR_16	*Equus asinus*	9.971	0.406	F	Kyrgyzstan	KYR	ERS12239262
KYR_31	*Equus asinus*	9.053	0.346	F	Kyrgyzstan	KYR	ERS12239263
KYR_Sdonk1	*Equus asinus*	29.519	0.524	M	Kyrgyzstan	KYR	SAMC048994
KYR_Sdonk2	*Equus asinus*	9.799	0.007	M	Kyrgyzstan	KYR	SAMC048995
MAU_2990	*Equus asinus*	23.125	0.009	F	Mauritania	MAU	ERS12239299
MAU_3094	*Equus asinus*	24.434	0.007	F	Mauritania	MAU	ERS12239298
MAU_3261	*Equus asinus*	27.093	0.006	F	Mauritania	MAU	ERS12239297
MON_08	*Equus asinus*	28.569	0.005	F	Mongolia	MON	ERS12239264
MON_10	*Equus asinus*	27.899	0.006	F	Mongolia	MON	ERS12239265
NIG_YPO62	*Equus asinus*	10.673	0.389	M	Nigeria	NIG	SAMC049032
NIG_YPO63	*Equus asinus*	11.36	0.327	F	Nigeria	NIG	SAMC049033
NIG_YPO64	*Equus asinus*	8.866	0.458	M	Nigeria	NIG	SAMC049034
NIG_YPO65	*Equus asinus*	10.422	0.318	M	Nigeria	NIG	SAMC049035
NIG_YPO66	*Equus asinus*	10.574	0.349	F	Nigeria	NIG	SAMC049036
NIG_YPO67	*Equus asinus*	14.534	0.224	M	Nigeria	NIG	SAMC049037
OMA_38	*Equus asinus*	28.817	0.005	F	Oman	OMA	ERS12239266
OMA_39	*Equus asinus*	25.459	0.022	M	Oman	OMA	ERS12239267
OMA_46	*Equus asinus*	28.281	0.005	F	Oman	OMA	ERS12239268
PTGm_02	*Equus asinus*	32.589	0.004	F	Portugal	PTGM	ERS12239269
PTGm_10	*Equus asinus*	23.579	0.008	F	Portugal	PTGM	ERS12239270
SAU_11	*Equus asinus*	28.634	0.007	M	Saudi Arabia	SAU	ERS12239271
SEN_10	*Equus asinus*	28.175	0.005	F	Senegal	SEN	ERS12239272
SOM_01	*Equus asinus*	29.409	0.045	F	Somalia	SOM	ERS12239273
SOM_05	*Equus asinus*	31.333	0.028	F	Somalia	SOM	ERS12239274

SOM_19	*Equus asinus*	31.999	0.005	F	Somalia	SOM	ERS12239275
SOM_20	*Equus asinus*	34.065	0.004	F	Somalia	SOM	ERS12239276
SOM_21	*Equus asinus*	36.165	0.004	F	Somalia	SOM	ERS12239277
ESP_Andalusian_1	*Equus asinus*	9.683	0.36	M	Spain	ESP	SRS3167402
ESP_Basque_10	*Equus asinus*	11.632	0.186	F	Spain	ESP	SRS3167401
ESP_Basque_11	*Equus asinus*	9.909	0.267	F	Spain	ESP	SRS3167400
ESP_Basque_12	*Equus asinus*	10.066	0.268	M	Spain	ESP	SRS3167399
ESP_Basque_13	*Equus asinus*	9.553	0.304	F	Spain	ESP	SRS3167404
ESP_ZamoranoLeones_14	*Equus asinus*	10.128	0.282	F	Spain	ESP	SRS3167403
ESP_ZamoranoLeones_15	*Equus asinus*	12.438	0.192	F	Spain	ESP	SRS3167405
ESP_ZamoranoLeones_16	*Equus asinus*	11.581	0.232	F	Spain	ESP	SRS3167385
ESP_ZamoranoLeones_17	*Equus asinus*	11.841	0.187	F	Spain	ESP	SRS3167386
ESP_Baleares_18	*Equus asinus*	9.217	0.457	F	Spain	ESP	SRS3167372
ESP_Andalusian_2	*Equus asinus*	11.462	0.215	F	Spain	ESP	SRS3167368
ESP_Andalusian_3	*Equus asinus*	9.061	0.455	F	Spain	ESP	SRS3167367
CYK_IslasCanarias_4	*Equus asinus*	10.154	0.248	M	Canary Islands	CYK	SRS3167465
CYK_IslasCanarias_5	*Equus asinus*	12.988	0.15	M	Canary Islands	CYK	SRS3167464
CYK_IslasCanarias_6	*Equus asinus*	9.165	0.353	F	Canary Islands	CYK	SRS3167472
CYK_IslasCanarias_7	*Equus asinus*	11.042	0.203	M	Canary Islands	CYK	SRS3167441
CYK_IslasCanarias_8	*Equus asinus*	9.904	0.273	M	Canary Islands	CYK	SRS3167469
CYK_IslasCanarias_9	*Equus asinus*	9.155	0.332	M	Canary Islands	CYK	SRS3167471
SUD_20	*Equus asinus*	25.12	0.007	F	Sudan	SUD	ERS12239278
SUD_49	*Equus asinus*	30.011	0.005	M	Sudan	SUD	ERS12239279
SUD_55	*Equus asinus*	31.734	0.004	M	Sudan	SUD	ERS12239280
SYR_06	*Equus asinus*	32.131	0.005	F	Syria	SYR	ERS12239281
SYR_19	*Equus asinus*	32.954	0.004	F	Syria	SYR	ERS12239282
TIB_DQFS2	*Equus asinus*	9.408	0.4	F	Tibet	TIB	SAMC048957
TIB_XZCD05	*Equus asinus*	8.972	0.382	F	Tibet	TIB	SAMC048958
TIB_XZCM09	*Equus asinus*	8.816	0.399	M	Tibet	TIB	SAMC048959
TIB_XZCM18	*Equus asinus*	10.206	0.285	M	Tibet	TIB	SAMC048960
TIB_DQFS1	*Equus asinus*	6.518	0.578	F	Tibet	TIB	SAMC048973
TIB_DQFS4	*Equus asinus*	6.837	0.537	F	Tibet	TIB	SAMC048974
TIB_DQFS6	*Equus asinus*	6.658	0.555	F	Tibet	TIB	SAMC048975
TIB_XZCD01	*Equus asinus*	7.04	0.526	F	Tibet	TIB	SAMC049005
TIB_XZCD02	*Equus asinus*	5.867	0.64	M	Tibet	TIB	SAMC049006
TIB_XZCD06	*Equus asinus*	5.985	0.63	F	Tibet	TIB	SAMC049007
TIB_XZCM01	*Equus asinus*	6.331	0.588	F	Tibet	TIB	SAMC049008
TIB_XZCM02	*Equus asinus*	5.934	0.627	M	Tibet	TIB	SAMC049009
TIB_XZCM05	*Equus asinus*	5.968	0.623	M	Tibet	TIB	SAMC049010
TIB_XZCM06	*Equus asinus*	6.112	0.62	M	Tibet	TIB	SAMC049011
TIB_XZCM10	*Equus asinus*	6.77	0.534	M	Tibet	TIB	SAMC049012
TIB_XZCM12	*Equus asinus*	5.808	0.652	F	Tibet	TIB	SAMC049013

TIB_XZCM17	*Equus asinus*	6.1	0.599	M	Tibet	TIB	SAMC049014
TIB_XZSNQS02	*Equus asinus*	6.194	0.607	F	Tibet	TIB	SAMC049015
TIB_XZSNQS03	*Equus asinus*	5.695	0.66	M	Tibet	TIB	SAMC049016
TIB_XZSNQS04	*Equus asinus*	5.636	0.667	M	Tibet	TIB	SAMC049017
TIB_XZSNQS05	*Equus asinus*	5.97	0.632	M	Tibet	TIB	SAMC049018
TIB_XZSNQS06	*Equus asinus*	5.97	0.647	F	Tibet	TIB	SAMC049019
TIB_XZSNQS07	*Equus asinus*	6.327	0.585	M	Tibet	TIB	SAMC049020
TIB_XZSNQS23	*Equus asinus*	7.006	0.516	M	Tibet	TIB	SAMC049021
TKM_107	*Equus asinus*	31.802	0.005	M	Turkmenistan	TKM	ERS12239283
TUK_07	*Equus asinus*	32.266	0.004	M	Turkey	TUK	ERS12239284
TUK_08	*Equus asinus*	31.102	0.005	F	Turkey	TUK	ERS12239285
TUK_26	*Equus asinus*	28.86	0.005	F	Turkey	TUK	ERS12239286
TUN_06	*Equus asinus*	30.823	0.005	M	Tunisia	TUN	ERS12239287
TUN_11	*Equus asinus*	35.539	0.004	F	Tunisia	TUN	ERS12239288
TUN_19	*Equus asinus*	40.152	0.004	F	Tunisia	TUN	ERS12239289
BRA_PegaDonkey	*Equus asinus*	13.706	0.065	F	Brazil	BRA	ERS12239290
YEM_08	*Equus asinus*	29.615	0.005	F	Yemen	YEM	ERS12239291
YEM_11	*Equus asinus*	40.052	0.003	M	Yemen	YEM	ERS12239292
YEM_17	*Equus asinus*	31.904	0.004	F	Yemen	YEM	ERS12239293
YUC_08	*Equus asinus*	27.058	0.007	M	Croatia	YUC	ERS12239294
YUM_08	*Equus asinus*	28.304	0.006	M	Macedonia	YUM	ERS12239295
YUM_13	*Equus asinus*	24.307	0.009	M	Macedonia	YUM	ERS12239296

839

840 **Table S2:** Sample information and naming for each ancient individual. Radiocarbon dates
841 were calibrated using Oxcal online and the IntCal20 calibration curve (*103, 104*). Ages
842 marked with a star are inferred from radiocarbon dates and archaeological context is shown
843 outside the bracket.

ID	Site	Country	Latitude	Longitude	Radiocarbon dated age	Age	Estimated Time period (years)	Period	Sex	Accession number
AC14380	Acemhöyük	Türkiye	38.41123	33.83569	3945±20 (UCIAMS 199621)	2455BCE	2564BCE-2346BCE	Bronze Age	M	ERS12239303
AC14415	Acemhöyük	Türkiye	38.41123	33.83569	3945±20 (UCIAMS 199619)	2455BCE	2564BCE-2346BCE	Bronze Age	M	ERS12239304
MV051	Acemhöyük	Türkiye	38.41123	33.83569	3784±41 (UBA-30288)	2219BCE	2400BCE-2039BCE	Bronze Age	M	ERS12239305
Chalow3	Chalow	Iran	37.10355	56.88528	N/A	2050BCE	2200BCE-	BMAC	F	ERS12239306
DoshanTepe	Doshan Tepe	Iran	35.6833	51.5	2840±15 (UCIAMS 223195)	989BCE	1049BCE-928BCE	Iron Age	F	ERS12239307
Tarquinia214	Tarquinia	Italy	42.0542	11.7576	2445±20 (UCIAMS 224884)	550BCE (*581BCE)	750BCE-	Etruscan (Archaic Period)	M	ERS12239308
Tarquinia501	Tarquinia	Italy	42.0542	11.7576	2515±20 and 2565±20 (UCIAMS 224885 and 224886)	520-500BCE (*680BCE)	803BCE-594BCE and	Etruscan (Archaic Period)	F	ERS12239309
MV242	Nizzana	Israel	30.88569	34.84694	2150±20 (UCIAMS 199283)	204BCE	350BCE-	Hellenistic Period	F	ERS12239310
AM39	Shahr-i-Qumis	Iran	36.5511	54.0175	N/A	800BCE-800CE	800BCE-	Parthian & Sassanian Period	F	ERS12239311
AM44	Shahr-i-Qumis	Iran	36.5511	54.0175	N/A	800BCE-800CE	800BCE-	Parthian & Sassanian Period	M	ERS12239312
AM66	Shahr-i-Qumis	Iran	36.5511	54.0175	N/A	800BCE-800CE	800BCE-	Parthian & Sassanian Period	M	ERS12239313
AM71	Shahr-i-Qumis	Iran	36.5511	54.0175	N/A	800BCE-800CE	800BCE-	Parthian & Sassanian Period	M	ERS12239314
AM805	Shahr-i-Qumis	Iran	36.5511	54.0175	1615±20 and 1585±15 (UCIAMS 223584 and 223188)	481CE	415-538CE and 421-542CE	Parthian & Sassanian Period	M	ERS12239315
AM89	Shahr-i-Qumis	Iran	36.5511	54.0175	N/A	800BCE-800CE	800BCE-	Parthian & Sassanian Period	M	ERS12239316
BourseB	Marseille	France	43.29774	5.374613	N/A	0CE-500CE	0CE-	Late Antiquity	M	ERS12239317
BourseC	Marseille	France	43.29774	5.374613	N/A	0CE-500CE	0CE-	Late Antiquity	M	ERS12239318
GVA125	Boinville-en-Woëvre	France	49.1858	5.6733	N/A	200CE-	200CE-	Roman Period	M	ERS12239319
GVA347	Boinville-en-Woëvre	France	49.1858	5.6733	N/A	200CE-	200CE-	Roman Period	M	ERS12239320
GVA348	Boinville-en-Woëvre	France	49.1858	5.6733	N/A	200CE-	200CE-	Roman Period	M	ERS12239321
GVA349	Boinville-en-Woëvre	France	49.1858	5.6733	N/A	200CE-	200CE-	Roman Period	M	ERS12239322

GVA353	Boinville-en-Woëvre	France	49.1858	5.6733	N/A	200CE-	200CE-	Roman Period	M	ERS12239323
GVA354	Boinville-en-Woëvre	France	49.1858	5.6733	N/A	200CE-	200CE-	Roman Period	F	ERS12239324
GVA355	Boinville-en-Woëvre	France	49.1858	5.6733	N/A	200CE-	200CE-	Roman Period	F	ERS12239325
GVA358	Boinville-en-Woëvre	France	49.1858	5.6733	N/A	200CE-	200CE-	Roman Period	F	ERS12239326
GVA359	Boinville-en-Woëvre	France	49.1858	5.6733	N/A	200CE-	200CE-	Roman Period	M	ERS12239327
Tur168	Yenikapi	Türkiye	40.9997	28.9498	1485±20 (UCIAMS 250285)	596CE	552CE-640CE	Byzantine Period	M	ERS12239328
Tur177	Yenikapi	Türkiye	40.9997	28.9498	1125±15 (UCIAMS 250291)	937CE	887CE-986CE	Byzantine Period	F	ERS12239329
Tur179	Yenikapi	Türkiye	40.9997	28.9498	1140±20 (UCIAMS 250292)	881CE	774CE-987CE	Byzantine Period	M	ERS12239330
Tur277	Yenikapi	Türkiye	40.9997	28.9498	1295±15 (UCIAMS 250363)	721CE	666CE-775CE	Byzantine Period	M	ERS12239331
Albufeira1x1	Albufeira	Portugal	37.0891	-8.2479	765±15 (UCIAMS 208877)	1254CE	1228CE-1280CE	Islamic Period	F	ERS12239332
Fiumarella1	Fiumarella	Italy	39.589	16.8127	165±25 (UCIAMS 229410)	1810CE	1683CE-1936CE	Bronze Age (Intrusive)	M	ERS12239333

844

Table S3: The number of variants identified by GraphTyper (version 2.5.1)(*61*) for modern individuals (*n*=222) before and after filtering for high quality variants.

	Number of variants	**SNPs**	**INDELs**
Raw (all scaffolds)	45,031,411	40,234,452	4,796,959
After filtering (all scaffolds)	13,267,291	11,655,167	1,680,089
After filtering (autosomes only)	13,013,551	11,426,298	1,587,253

Table S4: The number of variants remaining per autosome for the modern individuals (*n*=222) before and after filtering. The length of each autosome in base pairs and centimorgans, and the rate of recombination in cM/Mb is also reported.

Chromosome	Raw variants	Filtered variants	cM	Base pairs (bp)	cM/Mb
JADWZW010000002.1	2169745	637112	53.924	119293623	0.452
JADWZW010000003.1	4493039	1328891	120.098	238843737	0.503
JADWZW010000004.1	3304316	1021216	93.066	183770576	0.506
JADWZW010000005.1	1646778	474536	50.987	92920267	0.549
JADWZW010000006.1	2151743	628929	61.797	112287698	0.55
JADWZW010000007.1	1726889	500863	51.792	93367529	0.555
JADWZW010000008.1	2350964	689712	60.556	123522326	0.49
JADWZW010000009.1	2124123	714987	63.363	104245332	0.608
JADWZW010000010.1	1191724	342829	39.723	64678186	0.614
JADWZW010000011.1	1769988	531435	47.142	90664430	0.52
JADWZW010000012.1	1660121	525478	44.934	85786229	0.524
JADWZW010000013.1	1965805	578347	46.365	106341547	0.436
JADWZW010000014.1	1107198	307520	43.321	64917852	0.667
JADWZW010000015.1	876130	249858	36.895	47664211	0.774
JADWZW010000016.1	991742	284692	33.995	50230352	0.677
JADWZW010000017.1	941418	276091	30.639	50732407	0.604
JADWZW010000018.1	986503	315280	37.799	47651278	0.793
JADWZW010000019.1	670734	202542	20.526	33165847	0.619
JADWZW010000020.1	545920	165635	20.42	26995809	0.756
JADWZW010000021.1	1931507	585299	54.196	100519399	0.539
JADWZW010000022.1	1817414	517089	49.405	98587405	0.501
JADWZW010000023.1	705216	202045	21.852	38358317	0.57
JADWZW010000024.1	875823	262496	25.965	47367245	0.548
JADWZW010000025.1	865649	260065	23.107	46609582	0.496
JADWZW010000026.1	817711	237991	26.279	47151018	0.557
JADWZW010000027.1	581105	182904	22.156	28688911	0.772
JADWZW010000028.1	672875	208766	23.347	32167345	0.726
JADWZW010000029.1	1226740	356584	42.517	63892262	0.665
JADWZW010000030.1	754861	227618	24.809	37452155	0.662
JADWZW010000031.1	623338	196741	22.059	30281758	0.728
		Total	**1293.04**	**2,308,154,633**	**0.599**

Table S5: Ancestral populations and ancestry proportions for hybrid individuals, calculated using qpAdm modelling (version 810) (*64*). Significant *p*-values are indicated with a *.

ID	**Source 1**	**Source 2**	**Weight 1**	**Weight 2**	***p*-value**
ETH_5B	ETH	*E.a.som*	0.818	0.182	<0.001*
SOM_19	IRA	*E.a.som*	0.951	0.049	<0.001*
ETH_6B	ETH	*E.a.som*	0.956	0.044	<0.001*
SOM_20	IRA	*E.a.som*	0.963	0.038	<0.001*
SOM_21	IRA	*E.a.som*	0.968	0.032	<0.001*
ALG_01	ETH	*E.a.som*	0.974	0.026	<0.001*
YEM_08	EGY	*E.a.som*	0.981	0.019	<0.001*
SUD_49	SUD	*E.a.som*	0.981	0.019	<0.001*
YEM_11	EGY	*E.a.som*	0.983	0.017	<0.001*
OMA_46	EGY	*E.a.som*	0.983	0.017	<0.001*
YEM_17	EGY	*E.a.som*	0.984	0.016	<0.001*
OMA_38	EGY	*E.a.som*	0.985	0.015	<0.001*
MAU_3261	NIG	*E.a.som*	0.987	0.013	0.021*
OMA_39	EGY	*E.a.som*	0.988	0.012	0.005*
MAU_2990	NIG	*E.a.som*	0.991	0.009	0.568
GHA_07	NIG	*E.a.som*	0.993	0.007	0.183
SEN_10	NIG	*E.a.som*	0.995	0.005	0.281
CHI_Turfan_2	CHI	*E.kiang*	0.995	0.005	0.003*
MAU_3094	NIG	*E.a.som*	0.996	0.004	0.626
NIG_YPO67	NIG	*E.a.som*	0.997	0.003	0.259

Table S6: Genome coverage, proportion of missing variants and predicted accuracy of imputation based on tests conducted on modern variants for all ancient samples (*n*=31 individuals, *n*=7,161,029 variants, TI/TV=2.17).

Sample	Missingness	Coverage	Predicted Imputation Accuracy (all variants)
AM66	0.558	4.920	0.986
GVA348	0.600	5.050	0.986
Tur168	0.604	4.640	0.986
Tur179	0.612	4.693	0.986
Tur177	0.727	3.693	0.985
GVA349	0.731	4.090	0.985
AC14415	0.785	3.370	0.985
Tarquinia501	0.799	2.900	0.985
AM805	0.812	2.870	0.985
Tur277	0.827	2.951	0.984
BourseC	0.858	2.240	0.984
GVA125	0.862	2.060	0.984
Tarquinia214	0.864	2.250	0.984
MV242	0.879	2.510	0.984
GVA347	0.883	2.100	0.984
AC14380	0.892	2.120	0.984
Chalow3	0.895	2.110	0.984
MV051	0.901	1.270	0.984
AM89	0.901	1.570	0.984
AM44	0.903	1.670	0.984
AM71	0.910	1.540	0.984
BourseB	0.915	1.600	0.984
Albufeira1x1	0.919	1.620	0.983
Fiumarella1	0.930	1.440	0.983
DoshanTepe	0.938	1.540	0.983
GVA359	0.944	1.230	0.983
GVA355	0.953	1.210	0.983
GVA353	0.963	1.060	0.982
GVA358	0.963	1.105	0.982
AM39	0.973	0.770	0.981
GVA354	0.973	0.953	0.981

Table S7: The depth of reads and variants 5 base pairs either side of the missense mutation (G>A) at position JADWZW010000004.1:161390091 in *FGF5* (*46*) for 31 ancient donkeys.

sample	depth of reads	C	C	A	G	T	G/A	G	A	G	C	C
AC14380	3	.	.	.	.	.	.	.	.	.	.	.
AC14415	4	.	.	.	.	.	.	.	.	.	.	.
Albufeira1x1	1	.	.	.	.	.	.	.	.	.	.	.
AM39	2	.	.	.	.	.	.	.	.	.	.	.
AM44	1	.	.	.	.	.	.	.	.	.	.	.
AM66	5	.	.	.	.	.	.	.	.	.	.	.
AM71	4	.	.	.	.	.	.	.	.	.	.	.
AM805	6	.		.	.	.	.	.	.	.	.	.
AM89	3	.	.	.	.	.	A(1)	.	.	.	.	.
BourseB	0	.	.	.	.	.	.	.	.	.	.	.
BourseC	0	.	.	.	.	.	.	.	.	.	.	.
Chalow3	2	.	.	.	.	.	.	.	.	.	.	.
DoshanTepe	1	.	.	.	.	.	.	.	.	.	.	.
Fiumarella1	1	.	.	.	.	.	.	.	.	.	.	.
GVA125	0	.	.	.	.	.	.	.	.	.	.	.
GVA347	3	.	.	.	.	.	.	.	.	.	.	.
GVA348	4	.	.	.	.	.	.	.	.	.	.	.
GVA349	3	.	.	.	.	.	.	.	.	.	.	.
GVA353	0	.	.	.	.	.	.	.	.	.	.	.
GVA354	0	.	.	.	.	.	.	.	.	.	.	.
GVA355	0	.	.	.	.	.	.	.	.	.	.	.
GVA358	1	.	.	.	.	.	.	.	.	.	.	.
GVA359	4	.	.	.	.	.	.	.	.	.	.	.
MV051	0	.	.	.	.	.	.	.	.	.	.	.
MV242	3	.	.	.	.	.	.	.	.	.	.	.
Tarquinia214	1	.	.	.	.	.	.	.	.	.	.	.
Tarquinia501	4	.	.	.	.	.	.	.	.	.	.	.
Tur168	3	.	.	.	.	.	.	.	.	.	.	.
Tur177	2	.	.	.	.	.	.	.	.	.	G (1)	.
Tur179	3	.	.	.	.	.	.	.	.	.	.	.
Tur277	2	.	.	.	.	.	.	.	.	.	.	.

Table S8: The depth of reads and variants 5 base pairs either side of the frameshift deletion (delAT) at position (JADWZW010000004.1:161397694) in *FGF5* (*46*) for 31 ancient donkeys.

Sample	depth of reads	T	A	G	C	G	A/-	T/-	G	T	C	A	A
AC14380	2	.	.	.	.	.	.	.	.	.	.	.	.
AC14415	6	.	.	.	.	.	.	.	.	.	.	.	.
Albufeira1x1	0	.	.	.	.	.	.	.	.	.	.	.	.
AM39	1	.	.	.	.	.	.	.	.	.	.	.	.
AM44	1	.	.	.	.	.	.	.	.	.	.	.	.
AM66	8	.	.	.	.	A(1)	.	.	.	.	.	.	.
AM71	0	.	.	.	.	.	.	.	.	.	.	.	.
AM805	5	.	.	.	.	.	.	.	.	.	.	.	.
AM89	2	.	.	.	.	.	.	.	.	.	.	.	.
BourseB	2	.	.	.	T(1)	.	.	.	.	.	.	.	.
BourseC	4	.	.	.	.	.	.	.	.	.	.	.	.
Chalow3	3	.	.	.	.	A(1)	.	.	.	.	.	.	.
DoshanTepe	1	.	.	.	.	.	.	.	.	.	.	.	.
Fiumarella1	1	.	.	.	.	.	.	.	.	.	.	.	.
GVA125	5	.	.	.	.	A(1)	.	.	.	A(1)	G(1)	G(1)	.
GVA347	3	.	.	.	.	.	.	.	.	.	.	.	.
GVA348	10	.	.	.	.	.	.	.	.	.	.	.	.
GVA349	2	.	.	.	.	.	.	.	.	.	.	.	.
GVA353	3	.	.	.	.	.	.	.	.	.	.	.	.
GVA354	1	.	.	.	.	.	.	.	.	.	.	.	.
GVA355	0	.	.	.	.	.	.	.	.	.	.	.	.
GVA358	1	.	.	.	T(1)	.	.	.	.	.	.	.	.
GVA359	1	.	.	.	.	.	.	.	.	.	.	.	.
MV051	1	.	.	.	.	.	.	.	.	.	.	.	.
MV242	4	.	.	.	.	.	.	.	.	.	.	.	.
Tarquinia214	4	.	.	.	.	A(1)	.	.	.	.	.	.	.
Tarquinia501	5	.	.	.	.	.	.	.	.	.	.	.	.
Tur168	4	.	.	.	.	.	.	.	.	.	.	.	.
Tur177	6	.	.	.	T(1)	.	.	.	.	.	.	.	.
Tur179	3	.	.	.	-(1)	-(1)	-(1)	-(1)	.	.	.	.	.
Tur277	5	.	.	.	.	.	.	.	.	.	.	.	.

Table S9: The depth of reads and variants 5 base pairs either side of the T>A splice site mutation in *KIT* at position JADWZW010000004.1:139925278 for 31 ancient donkeys (*45*).

Sample	depth of reads	G	A	G	G	T/A	A	A	A	G	C
AC14380	0	.	.	.	.	.	.	.	.	.	.
AC14415	5	.	T(1)	.	.	.	.	.	.	T(1)	T(1)
Albufeira1x1	2	.	.	.	.	.	.	.	.	.	.
AM39	0	.	.	.	.	.	.	.	.	.	.
AM44	1	.	.	.	.	.	.	.	.	.	.
AM66	3	.	.	.	.	.	.	.	.	.	.
AM71	2	.	.	.	.	.	.	.	.	.	.
AM805	3	.	.	.	.	.	.	.	.	.	.
AM89	5	.	.	.	.	.	.	.	.	.	.
BourseB	1	.	.	.	.	.	.	.	.	.	.
BourseC	8	.	.	.	.	.	.	.	.	.	.
Chalow3	1	.	.	.	.	.	.	.	.	.	.
DoshanTepe	1	.	.	.	.	.	.	.	.	.	.
Fiumarella1	1	.	.	.	.	.	.	.	.	.	.
GVA125	1	.	.	.	.	.	.	.	.	.	.
GVA347	1	.	.	.	.	.	.	.	.	.	.
GVA348	3	.	.	.	.	.	.	.	.	.	.
GVA349	3	.	.	.	.	.	.	.	.	.	.
GVA353	0	.	.	.	.	.	.	.	.	.	.
GVA354	2	.	.	.	.	.	.	.	.	.	.
GVA355	0	.	.	.	.	.	.	.	.	.	.
GVA358	2	.	.	.	.	.	.	.	.	.	.
GVA359	0	.	.	.	.	.	.	.	.	.	.
MV051	2	.	.	.	.	.	.	.	.	.	.
MV242	3	.	.	.	T(1)	.	.	.	.	.	.
Tarquinia214	4	.	.	.	.	.	.	.	.	.	.
Tarquinia501	2	.	.	.	.	.	.	.	.	.	.
Tur168	6	.	.	.	.	.	.	.	.	.	.
Tur177	6	.	.	.	.	.	.	.	.	.	.
Tur179	5	.	.	.	.	.	.	.	.	.	.
Tur277	2	.	.	.	.	.	.	.	.	.	.

Table S10: Levels of relatedness between ancient individuals estimated using KING (version 2.2.7) (*66*) with the imputed variant panel, conditioning on transversions only (*n*=31 individuals, *n*=619,981 transversions). Only relationships between individuals inferred to show genetic relatedness are shown.

ID1	ID2	Proportion of IBD	Degree of relatedness
GVA355	GVA358	0.705	1st
GVA125	GVA353	0.250	2nd
GVA347	GVA353	0.238	2nd
GVA125	GVA348	0.201	2nd
GVA348	GVA353	0.189	2nd
GVA348	GVA354	0.188	2nd
GVA125	GVA354	0.178	2nd
GVA347	GVA354	0.166	3rd
GVA125	GVA347	0.162	3rd
GVA347	GVA348	0.159	3rd
GVA353	GVA354	0.154	3rd
GVA348	GVA349	0.138	3rd
GVA347	GVA349	0.125	3rd
GVA349	GVA353	0.124	3rd
GVA125	GVA349	0.108	3rd
GVA349	GVA354	0.101	3rd
Tarquinia214	Tarquinia501	0.081	4th
GVA348	GVA355	0.052	4th
GVA348	GVA358	0.051	4th

Table S11: Sample information and accession numbers for each modern (*n*=79) and ancient horse (*n*=75) used for estimating inbreeding levels (Fig. 4D, E, F). Whole-genome sequence data and metadata on the site, country and age (inferred from the radiocarbon dates) for the ancient horses were obtained from (*21, 41, 50, 131*). Whole-genome sequence data for modern horses were obtained from (*132-137*)

ID	Site	Country	Age	age2	label	Accession number
Akha_0248A_AKT001	NA	NA	2000CE	2000	M1	ERS1246351
Akha_0302A_AKT003	NA	NA	2000CE	2000	M2	ERS1246352
Arabian_UFL_948_ERR3465834	NA	NA	2000CE	2000	M3	ERS3631438
Bava_0183A_BY01	NA	NA	2000CE	2000	M4	ERS1263371
CDM12_Chaidamu_DulanQinghai	NA	NA	2000CE	2000	M5	SRS4251825
CDM15_Chaidamu_DulanQinghai	NA	NA	2000CE	2000	M6	SAMN28422840
CuTr_0137A_CU_COL163706	NA	NA	2000CE	2000	M7	ERS1560528
CuTr_0138A_CU_COL163725	NA	NA	2000CE	2000	M8	ERS1560529
DB35_DeBa_Debao_Guangxi	NA	NA	2000CE	2000	M9	SAMN28422841
DT12_DaTo_Datong_QilianQinghai	NA	NA	2000CE	2000	M10	SAMN28422842
DT3_DaTo_Datong_QilianQinghai	NA	NA	2000CE	2000	M11	SAMN28422843
Dutc_0308A	NA	NA	2000CE	2000	M12	ERS1246371
ELC21_Erlunchun_InnerMongolia	NA	NA	2000CE	2000	M13	SRS4251811
Fjor_0142A_Fjord	NA	NA	2000CE	2000	M14	SRS438157
Frie_0298A_SAMEA3951220	NA	NA	2000CE	2000	M15	ERS1138354
Frie_0300A_SAMEA3951222	NA	NA	2000CE	2000	M16	ERS1138356
FrMo_0041A_FM0001	NA	NA	2000CE	2000	M17	ERS1246356
FrMo_0065A_FM1798	NA	NA	2000CE	2000	M18	ERS1246364
Hafl_0309A_HF0002	NA	NA	2000CE	2000	M19	ERS1982326
Hafl_0310A_HF0003	NA	NA	2000CE	2000	M20	ERS1982327
Hano_0172A_HAN01	NA	NA	2000CE	2000	M21	ERS1263372
Hano_0312A_HN001	NA	NA	2000CE	2000	M22	ERS1982322
Hols_0173A_HOL01	NA	NA	2000CE	2000	M23	ERS1263373
Icel_0144A_P5782	NA	NA	2000CE	2000	M24	SRS309532
Icel_0247A_IS074	NA	NA	2000CE	2000	M25	ERS709890
JC5_JiCh_Jianchang_SW	NA	NA	2000CE	2000	M26	SAMN28422844
Jeju_0274A_SAMN01057171	NA	NA	2000CE	2000	M27	SRS346578
Jeju_0275A_SAMN01057172	NA	NA	2000CE	2000	M28	SRS346579
JZ3_JiZi_JiangziTibet	NA	NA	2000CE	2000	M29	SRS4251838
JZ4_JiZi_JiangziTibet	NA	NA	2000CE	2000	M30	SRS4251838
Lipi_0187A_CSess113	NA	NA	2000CE	2000	M31	SRS1818811
Lipi_0188A_FRal169	NA	NA	2000CE	2000	M32	SRS1818795
LKZ22_Langkazi_Tibet	NA	NA	2000CE	2000	M33	SAMN28422845
LKZ28_Langkazi_Tibet	NA	NA	2000CE	2000	M34	SRS4251851
Marw_0239A_SRR1275408	NA	NA	2000CE	2000	M35	SRS603966
Mixd_0314A_UKH4	NA	NA	2000CE	2000	M36	ERS1076964
Mong_0153A_KB7754	NA	NA	2000CE	2000	M37	ERS805731
Mong_0216A_TG1111D2629	NA	NA	2000CE	2000	M38	SRS543625
Morg_0096A_EMS595	NA	NA	2000CE	2000	M39	ERS806987

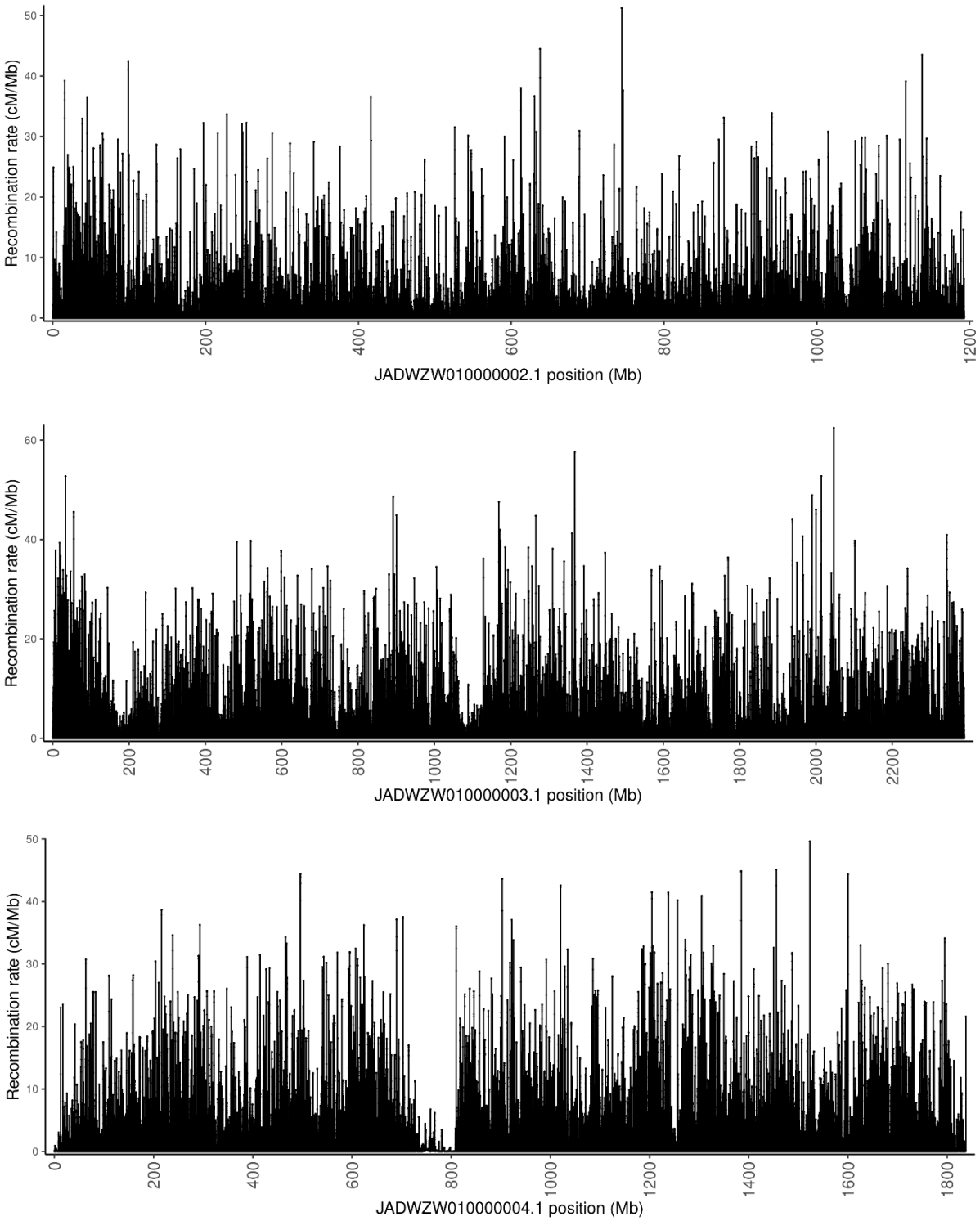
Morg_0315A_EQ053	NA	NA	2000CE	2000	M40	ERS1982319
MZH22_Mozhu_MozhugongTibet	NA	NA	2000CE	2000	M41	SAMN28422846
MZH24_Mozhu_MozhugongTibet	NA	NA	2000CE	2000	M42	SAMN28422847
NM2_NiMu_Tibet	NA	NA	2000CE	2000	M43	SAMN28422848
NM20_NiMu_Tibet	NA	NA	2000CE	2000	M44	SAMN28422849
Nori_0316A_NO180	NA	NA	2000CE	2000	M45	ERS1982325
NQ9916_NiQi_NingqiangShaanxi	NA	NA	2000CE	2000	M46	SAMN28422850
Olde_0176A_OLD01	NA	NA	2000CE	2000	M47	ERS1263375
Olde_0177A_OLD02	NA	NA	2000CE	2000	M48	ERS1263376
Pain_0319A_UKH16	NA	NA	2000CE	2000	M49	ERS1076966
Pain_0320A_UKH29	NA	NA	2000CE	2000	M50	ERS1076967
Quar_0321A_QH070	NA	NA	2000CE	2000	M51	ERS1246372
Quar_0322A_QH225	NA	NA	2000CE	2000	M52	ERS1246374
Reit_0323A_DR011	NA	NA	2000CE	2000	M53	ERS1982318
Reit_0324A_DR033	NA	NA	2000CE	2000	M54	ERS1982315
Shet_0249A_SPH020	NA	NA	2000CE	2000	M55	ERS715262
Shet_0250A_SPH041	NA	NA	2000CE	2000	M56	ERS715261
Sorr_0236A_SAMN02439778	NA	NA	2000CE	2000	M57	SRS513153
Sorr_0270A_SAMN03955413	NA	NA	2000CE	2000	M58	SRS1022305
Stan_0325A_AS002	NA	NA	2000CE	2000	M59	ERS1230234
Standardbred_UFL_CU1406_ERR3465842	NA	NA	2000CE	2000	M60	ERS3631446
Standardbred_UFL_CU2446_ERR3465843	NA	NA	2000CE	2000	M61	ERS3631447
Swis_0326A_RAO310_2	NA	NA	2000CE	2000	M62	ERS1263382
Swis_0327A_RAO441_2	NA	NA	2000CE	2000	M63	ERS1263383
Thor_0290A_SAMN01047706	NA	NA	2000CE	2000	M64	SRS345336
Thoroughbred_UFL_CU3903_ERR3465845	NA	NA	2000CE	2000	M65	ERS3631449
Trak_0178A_TRA01	NA	NA	2000CE	2000	M66	ERS1263377
Trak_0179A_TRA02	NA	NA	2000CE	2000	M67	SRS1818810
UFL_QH140147_ERR3465848	NA	NA	2000CE	2000	M68	ERS3631452
Wels_0330A_WP006	NA	NA	2000CE	2000	M69	ERS1982316
Wels_0331A_WP007	NA	NA	2000CE	2000	M70	ERS1982323
West_0180A_WF01	NA	NA	2000CE	2000	M71	ERS1263379
West_0181A_WF02	NA	NA	2000CE	2000	M72	ERS1263380
WMG8_Mongolian_Mongolia	NA	NA	2000CE	2000	M73	SAMN28422851
Wurt_0182A_BW01	NA	NA	2000CE	2000	M74	ERS1263370
WZ6_MoGo_InnerMongolia	NA	NA	2000CE	2000	M75	SRS4251803
Yaku_0164A_Yak2	NA	NA	2000CE	2000	M76	ERS849387
Yaku_0169A_Yak7	NA	NA	2000CE	2000	M77	ERS849392
YL2_YiLi_Zhaosu_Pair	NA	NA	2000CE	2000	M78	SAMN28422852
YQ29_YaQi_Yanqi_Xinjiang	NA	NA	2000CE	2000	M79	SAMN28422853
ARUS_0222A_CGG101397	Tumeski	Russia	1825CE	1825	A1	SRS497178, SRS497177
WitterPlace_UK17_267	Witter Place	United Kingdom	1750CE	1750	A2	ERS3213633
Beauvais_GVA375_467	Beauvais, VIlliers-de-l'Isle Adam	France	1550CE	1550	A3	ERS3213470
TavanTolgoi_GEP13_730	Tavan Tolgoi	Mongolia	1287CE	1287	A4	ERS3213603
TavanTolgoi_GEP14_730	Tavan Tolgoi	Mongolia	1287CE	1287	A5	ERS3213604

TavanTolgoi_GEP21_730	Tavan Tolgoi	Mongolia	1287CE	1287	A6	ERS3213605
Yenikapi_Tur150_1443	Yenikapi	Turkey	961CE	961	A7	ERS3213646
Yenikapi_Tur145_1156	Yenikapi	Turkey	951CE	951	A8	ERS3213642
Saadjarve_Saa1_1117	Saadjärve	Estonia	900CE	900	A9	ERS3213583
Nustar_5_1187	Nuštar	Croatia	830CE	830	A10	ERS3213573
Marvele_18_1189	Marvele cemetery	Lithuania	829CE	829	A11	ERS3213561
Marvele_32_1144	Marvele cemetery	Lithuania	829CE	829	A12	ERS3213565
Yenikapi_Tur229_1443	Yenikapi	Turkey	827CE	827	A13	ERS3213660
Grigorevka4_PAVH2_1192	Gregorevka	Kazakhstan	825CE	825	A14	ERS1892698
Yenikapi_Tur193_1443	Yenikapi	Turkey	792CE	792	A15	ERS3213657
Yenikapi_Tur140_1289	Yenikapi	Turkey	777CE	777	A16	ERS3213638
Khotont_UCIE2012x85_1291	Khotont	Mongolia	725CE	725	A17	ERS3213547
BozAdyr_KYRH10_1267	Boz-Adyr	Kyrgyzstan	700CE	700	A18	ERS3213485
BozAdyr_KYRH8_1267	Boz-Adyr	Kyrgyzstan	700CE	700	A19	ERS3213486
Yenikapi_Tur172_1695	Yenikapi	Turkey	674CE	674	A20	ERS1892707
Yenikapi_Tur194_1360	Yenikapi	Turkey	657CE	657	A21	ERS3213658
Yenikapi_Tur142_1396	Yenikapi	Turkey	648CE	648	A22	ERS3213640
Yenikapi_Tur141_1430	Yenikapi	Turkey	640CE	640	A23	ERS3213639
Yenikapi_Tur170_1443	Yenikapi	Turkey	601CE	601	A24	ERS3213649
SharIQumis_AM115_1557	Shar-I-Qumis	Iran	472CE	472	A25	ERS3213596
Yenikapi_Tur146_1730	Yenikapi	Turkey	350CE	350	A26	ERS3213643
Yenikapi_Tur171_1689	Yenikapi	Turkey	336CE	336	A27	ERS3213650
FrankfurtHeddernheim_Fr1_1863	Frankfurt-Heddenheim	Germany	180CE	180	A28	ERS3213533
Chartres_GVA26_1917	Chartres, boulevard de la Courtille	France	110CE	110	A29	ERS3213502
Chartres_GVA4_1917	Chartres, boulevard de la Courtille	France	110CE	110	A30	ERS3213506
Chartres_GVA43_1917	Chartres, boulevard de la Courtille	France	110CE	110	A31	ERS3213507
Chartres_GVA81_1917	Chartres, boulevard de la Courtille	France	110CE	110	A32	ERS3213518
GolModII_Mon24_1993	Gol Mod II	Mongolia	40CE	40	A33	ERS3213535
GolModII_Mon23_2007	Gol Mod II	Mongolia	35CE	35	A34	ERS3213534
GolModII_Mon26_1999	Gol Mod II	Mongolia	27CE	27	A35	ERS3213537
GolModII_Mon28_1988	Gol Mod II	Mongolia	27CE	27	A36	ERS1892697
GolModII_Mon25_2011	Gol Mod II	Mongolia	17CE	17	A37	ERS3213536
GolModII_Mon27_2011	Gol Mod II	Mongolia	17CE	17	A38	ERS3213538
SaintJust_GVA242_2250	Saint-Just-en-Chaussée	France	75BCE	-75	A39	ERS3213589
Actiparc_GVA308_2312	Actiparc	France	210BCE	-210	A40	ERS3213454
AC7970_AMIS-1-00131_Tur_m290	Acemhoyuk	Turkey	290BCE	-290	A41	ERS7255955
OlonKurinGol_OKG2_2367	Olon Kurin Gol	Mongolia	350BCE	-350	A42	ERS3213577
Fetusx9m_CGG-1-022147_Spa_m475	Els Vilars	Spain	475BCE	-475	A43	ERS7256018
SV2019x18_AMIS-1-02382_Tun_m581	Althiburos	Tunisia	581BCE	-581	A44	ERS7256181
18ELTu18_AMIS-1-01102_Spa_m588	El Turuñuelo	Spain	588BCE	-588	A45	ERS7255954
SV2019x19_AMIS-1-02383_Tun_m643	Althiburos	Tunisia	643BCE	-643	A46	ERS7256182
UE4618_CGG_1_020962	Els Vilars	Spain	655BCE	-655	A47	ERS3213526
Hasanlu1140_CGG-1-019998_Ira_m663	Tepe Hasanlu	Iran	663BCE	-663	A48	ERS7256042
UE11080x11082_CGG-1-020973_Spa_m664	Els Vilars	Spain	664BCE	-664	A49	ERS7256188
Rid1_CGG_1_018468	Ridala	Estonia	700BCE	-700	A50	ERS7256148

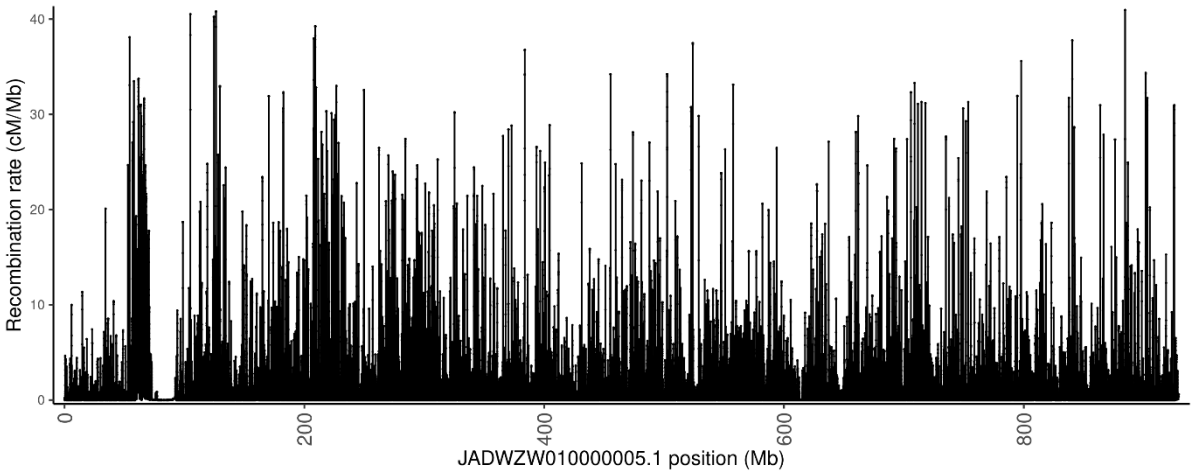
Rid2_CGG_1_018469	Ridala	Estonia	700BCE	-700	A51	ERS7256149
Hasanlu2327_CGG-1-019995_Ira_m768	Tepe Hasanlu	Iran	768BCE	-768	A52	ERS7256043
Hasanlu3398_CGG-1-019986_Ira_m768	Tepe Hasanlu	Iran	768BCE	-768	A53	ERS7256046
HasanluV31E_CGG-1-021461_Ira_m768	Tepe Hasanlu	Iran	768BCE	-768	A54	ERS7256049
Hasanlu3394_CGG-1-019997_Ira_m790	Tepe Hasanlu	Iran	790BCE	-790	A55	ERS7256045
Fen4_CGG-1-018396_Chi_m800	Fengtai	China	800BCE	-800	A56	ERS7256017
Hasanlu2405_CGG-1-019992_Ira_m868	Tepe Hasanlu	Iran	868BCE	-868	A57	ERS7256044
Hasanlu368_CGG-1-019994_Ira_m878	Tepe Hasanlu	Iran	878BCE	-878	A58	ERS7256048
Hasanlu3461_CGG-1-020003_Ira_m913	Tepe Hasanlu	Iran	913BCE	-913	A59	ERS7256047
CD5203_AMIS-1-00107_Tur_m985	Çadır Höyük	Turkey	985BCE	-985	A60	ERS7255998
UushgiinUvur_Mon45_3080	Uushgiin Uvur	Mongolia	1065BCE	-1065	A61	ERS3213624
UushgiinUvur_Mon37_3085	Uushgiin Uvur	Mongolia	1075BCE	-1075	A62	ERS3213617
UushgiinUvur_Mon39_3085	Uushgiin Uvur	Mongolia	1075BCE	-1075	A63	ERS3213618
UushgiinUvur_Mon84_3123	Uushgiin Uvur	Mongolia	1075BCE	-1075	A64	ERS1892705
UushgiinUvur_Mon86_3039	Uushgiin Uvur	Mongolia	1075BCE	-1075	A65	ERS1892706
SAGxS27_CGG-1-019559_Ira_m1102	Sagzabad	Iran	1102BCE	-1102	A66	ERS7256175
UushgiinUvur_Mon87_3117	Uushgiin Uvur	Mongolia	1103BCE	-1103	A67	ERS3213626
Mon43_CGG_1_018079	Uushgiin Uvur	Mongolia	1106BCE	-1106	A68	ERS3213622
UushgiinUvur_Mon42_3130	Uushgiin Uvur	Mongolia	1110BCE	-1110	A69	ERS3213621
CD1819_AMIS-1-00115_Tur_m1299	Çadır Höyük	Turkey	1299BCE	-1299	A70	ERS7255996
Bateni_Rus16_3350	Bateni	Russia	1336BCE	-1336	A71	ERS3213468
TP4_CGG-1-018394_Geo_m1578	Tachti Perda	Georgia	1578BCE	-1578	A72	ERS7256186
AC9016_AMIS-1-00134_Tur_m1900	Acemhoyuk	Turkey	1900BCE	-1900	A73	ERS7255957
Sintashta_NB46_4023	Sintashta	Russia	2009BCE	-2009	A74	ERS821436
AC8811_AMIS-1-00133_Tur_m2125	Acemhoyuk	Turkey	2125BCE	-2125	A75	ERS7255956

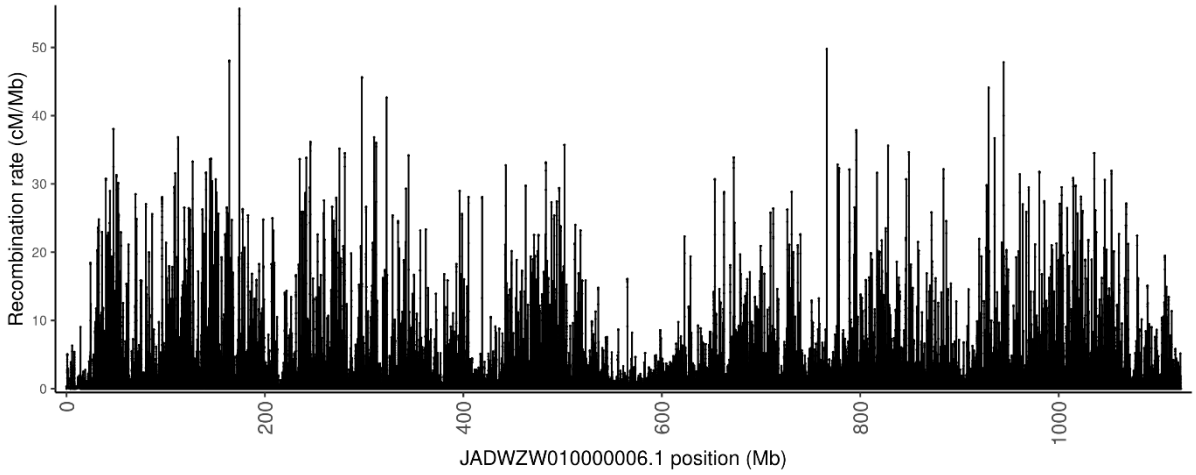
882

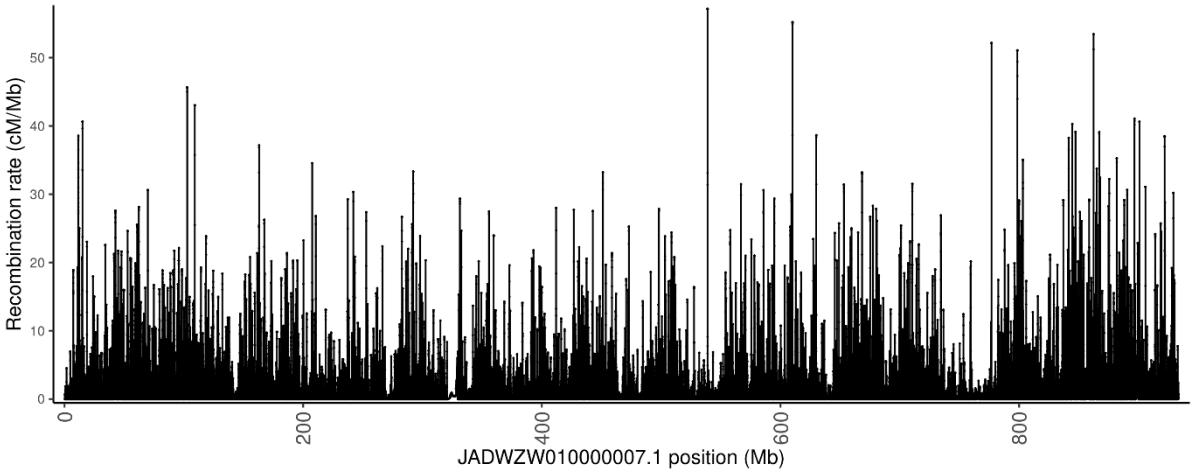
883


Recombination rate (cM/Mb)
JADWZW010000002.1 position (Mb)
Recombination rate (cM/Mb)
JADWZW010000003.1 position (Mb)
Recombination rate (cM/Mb)
JADWZW010000004.1 position (Mb)

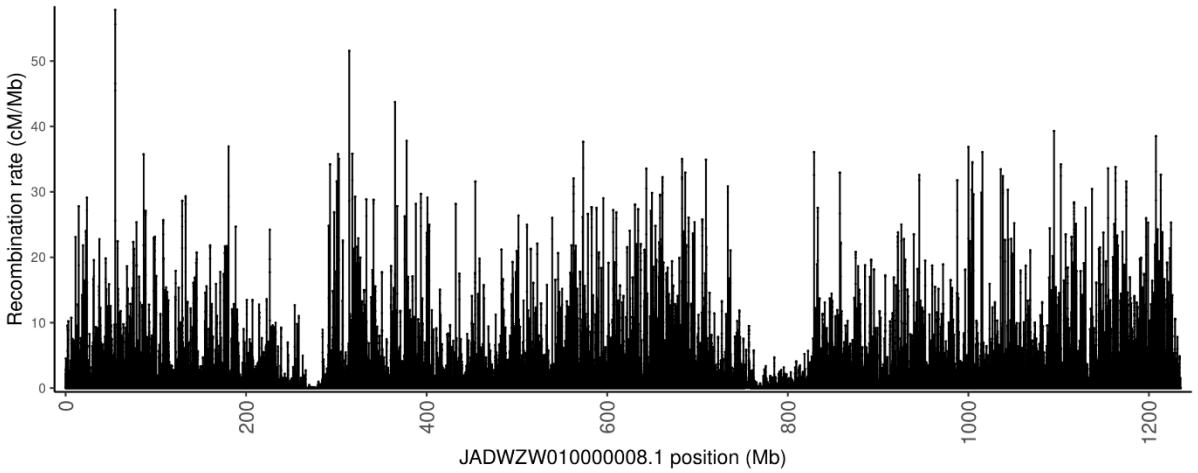
884

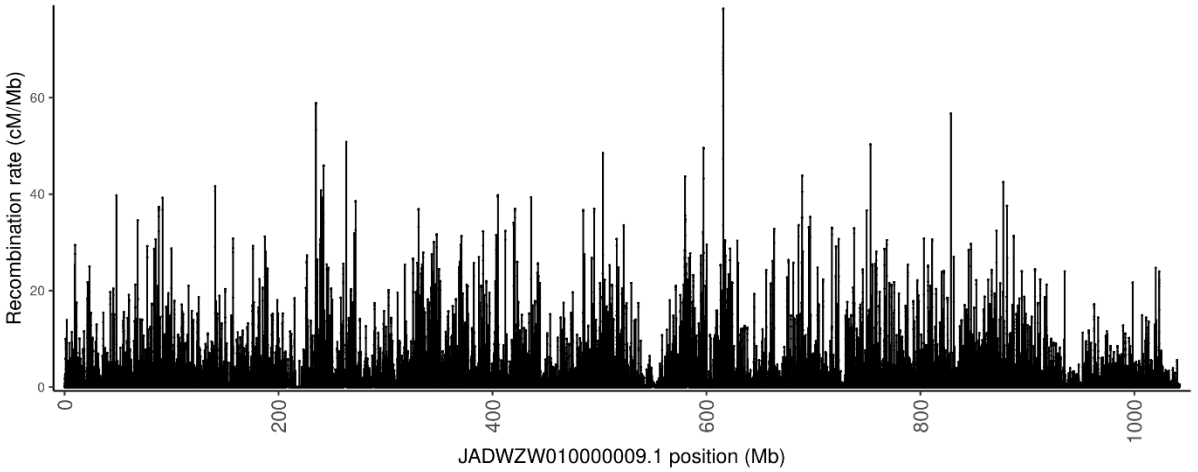

Recombination rate (cM/Mb)
0
10
20
30
40
0
200
400
600
800
JADWZW010000005.1 position (Mb)

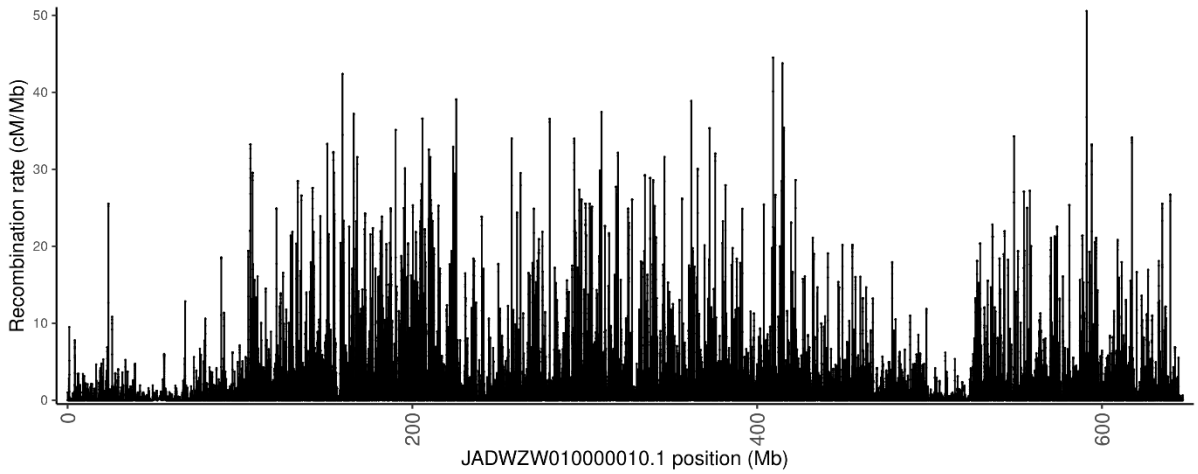

Recombination rate (cM/Mb)
0
10
20
30
40
50
0
200
400
600
800
1000
JADWZW010000006.1 position (Mb)


Recombination rate (cM/Mb)
0
10
20
30
40
50
0
200
400
600
800
JADWZW010000007.1 position (Mb)

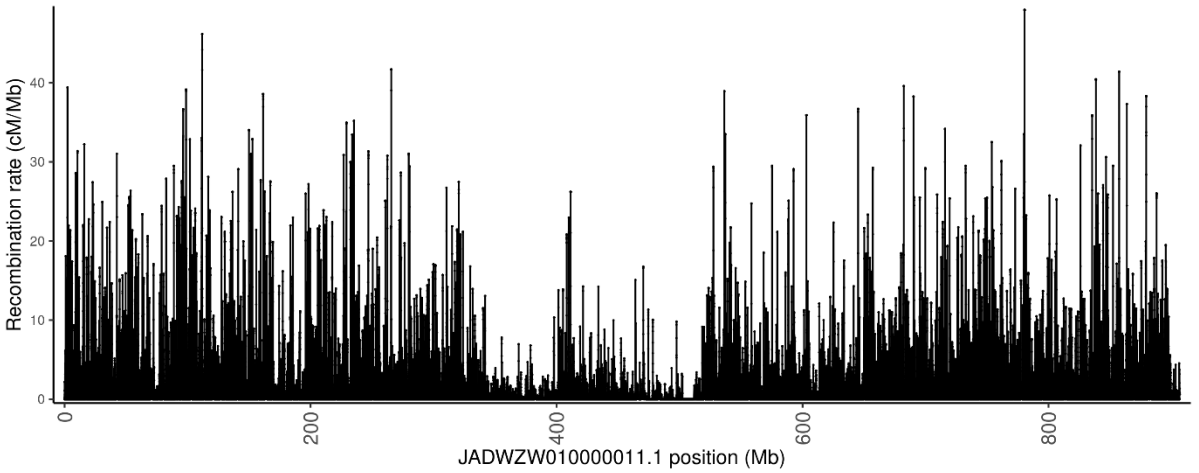
885

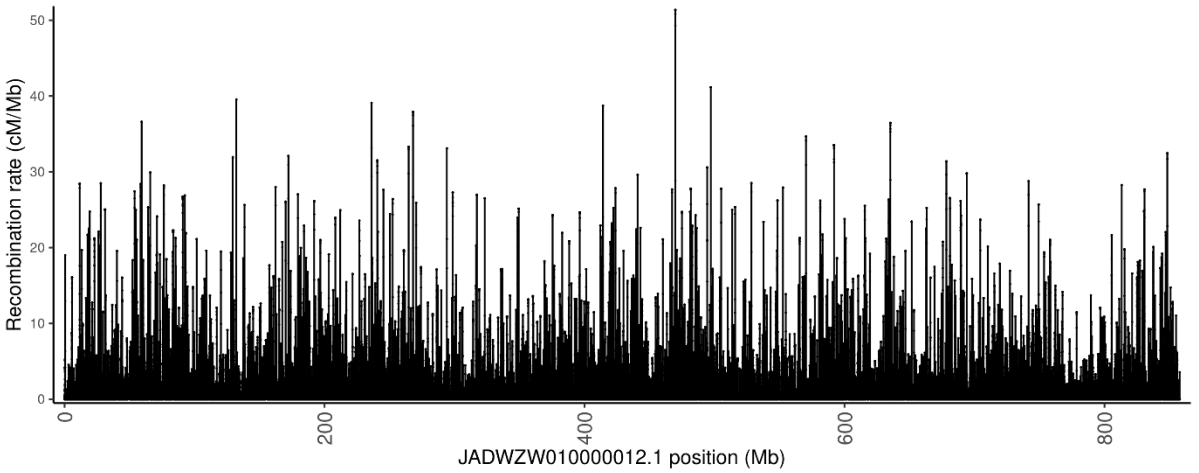

Recombination rate (cM/Mb)
0
10
20
30
40
50
0
200
400
600
800
1000
1200
JADWZW010000008.1 position (Mb)

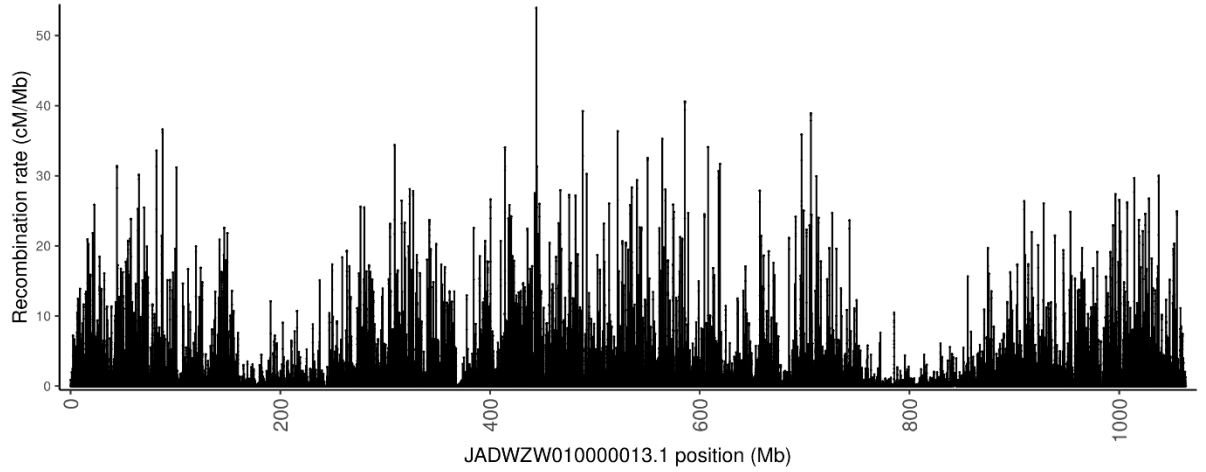

Recombination rate (cM/Mb)
0
20
40
60
0
200
400
600
800
1000
JADWZW010000009.1 position (Mb)


Recombination rate (cM/Mb)
0
10
20
30
40
50
0
200
400
600
JADWZW010000010.1 position (Mb)

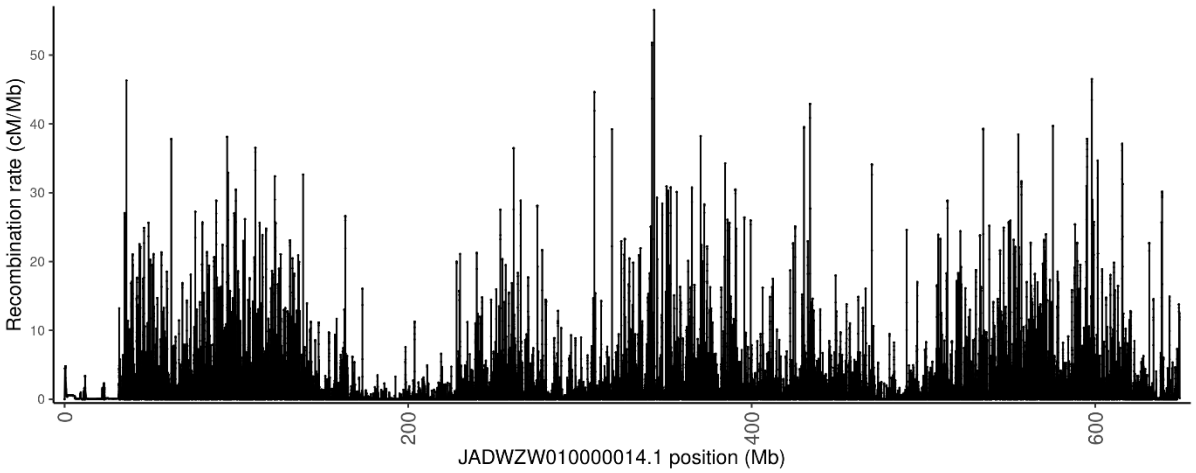
886

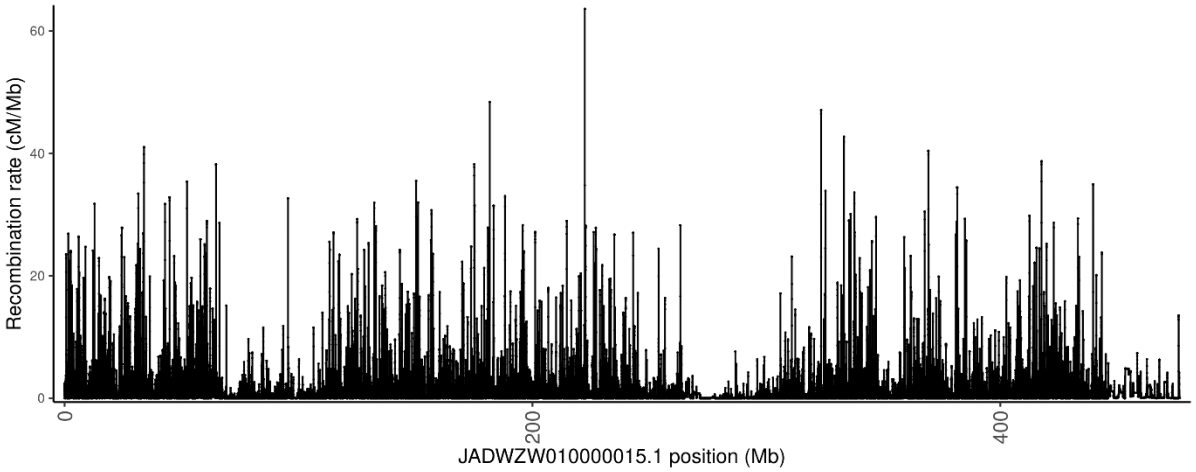

Recombination rate (cM/Mb)
0
10
20
30
40
0
200
400
600
800
JADWZW010000011.1 position (Mb)

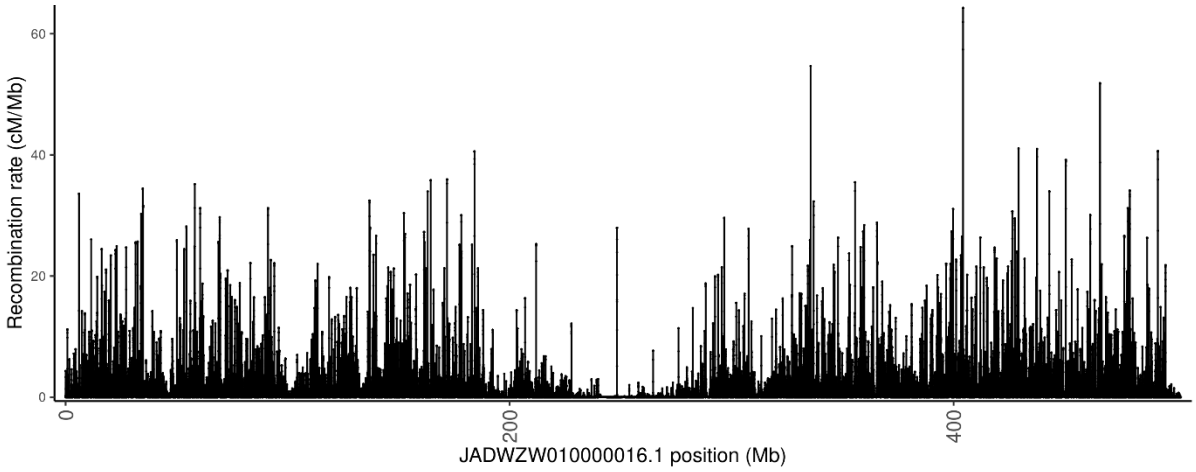

Recombination rate (cM/Mb)
0
10
20
30
40
50
0
200
400
600
800
JADWZW010000012.1 position (Mb)


Recombination rate (cM/Mb)
0
10
20
30
40
50
0
200
400
600
800
1000
JADWZW010000013.1 position (Mb)

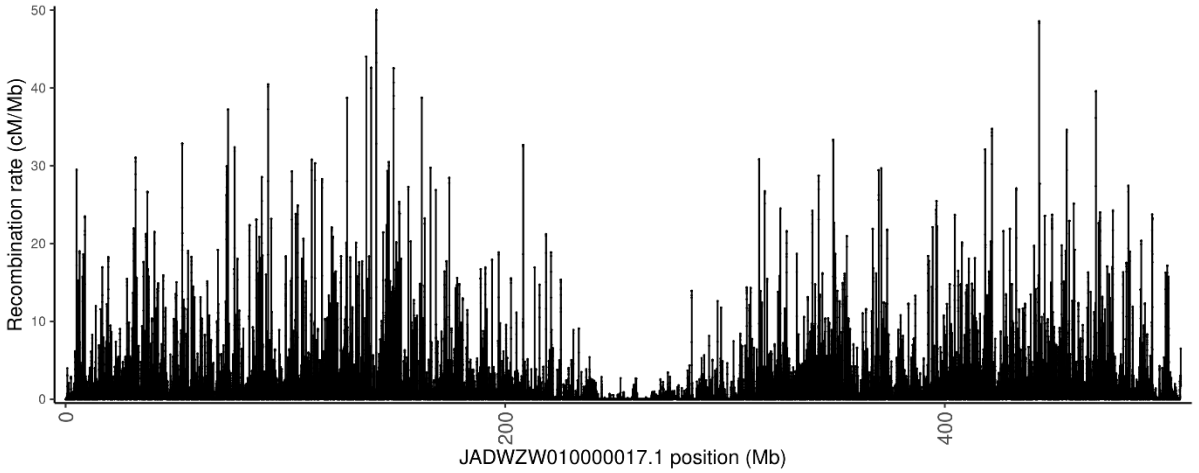
887

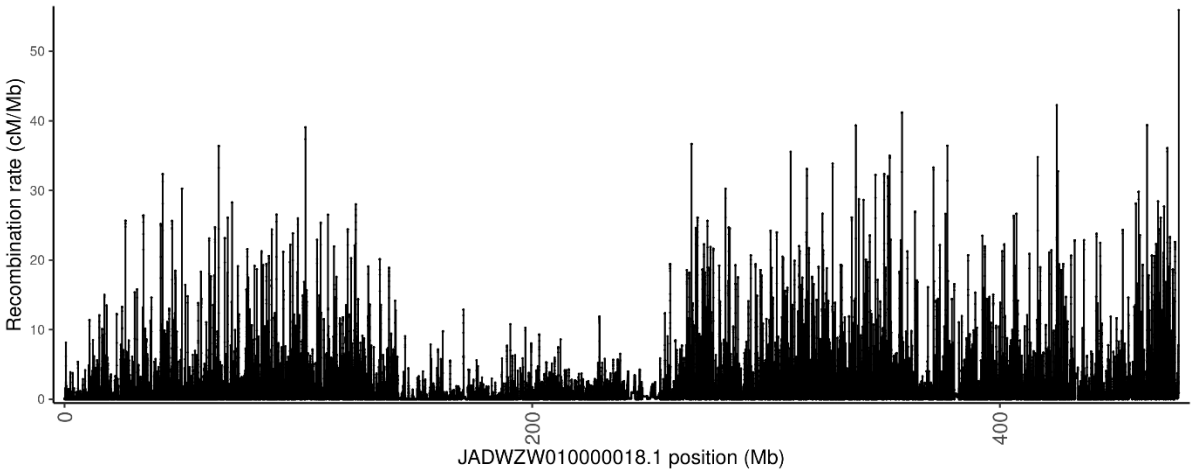

Recombination rate (cM/Mb)
0
10
20
30
40
50
0
200
400
600
JADWZW010000014.1 position (Mb)

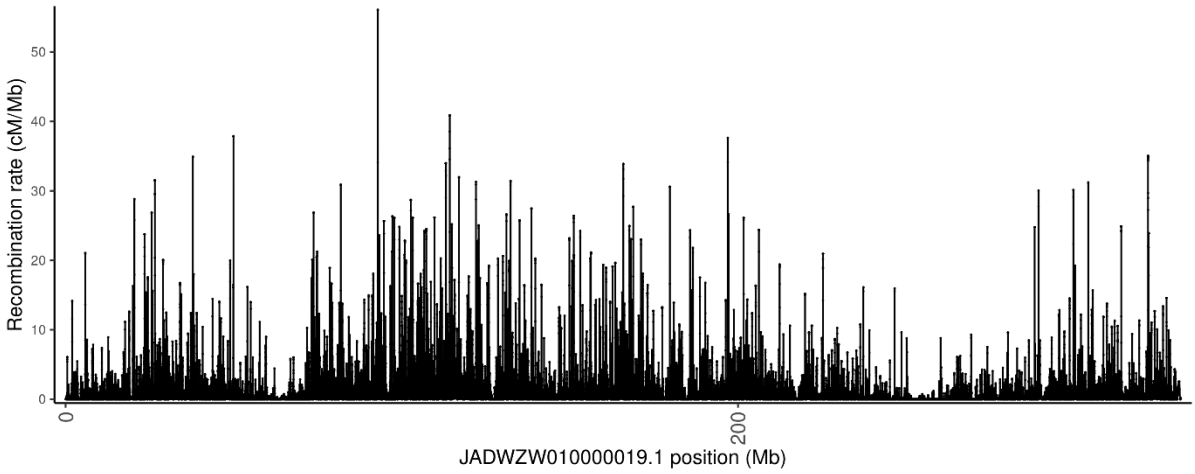

Recombination rate (cM/Mb)
0
20
40
60
0
200
400
JADWZW010000015.1 position (Mb)


Recombination rate (cM/Mb)
0
20
40
60
0
200
400
JADWZW010000016.1 position (Mb)

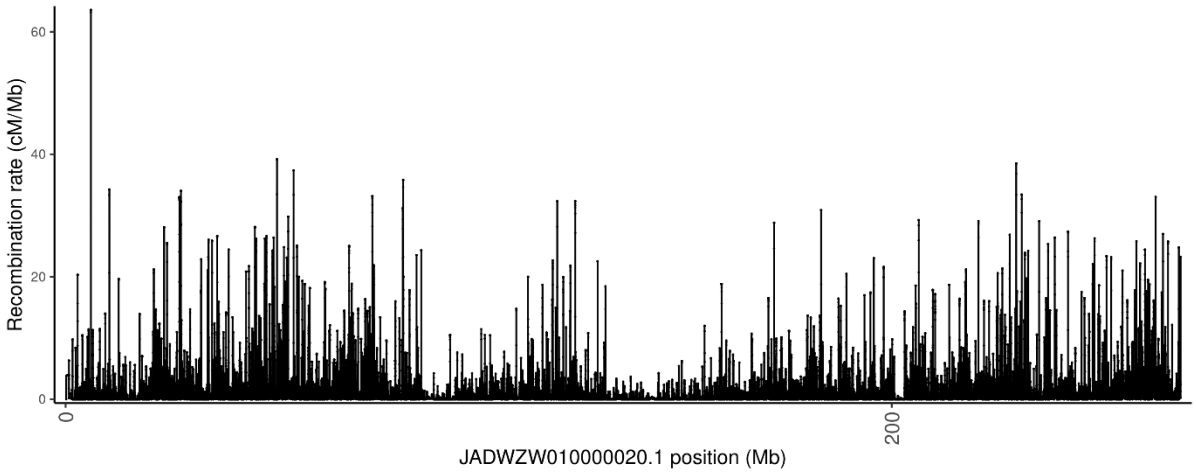
888

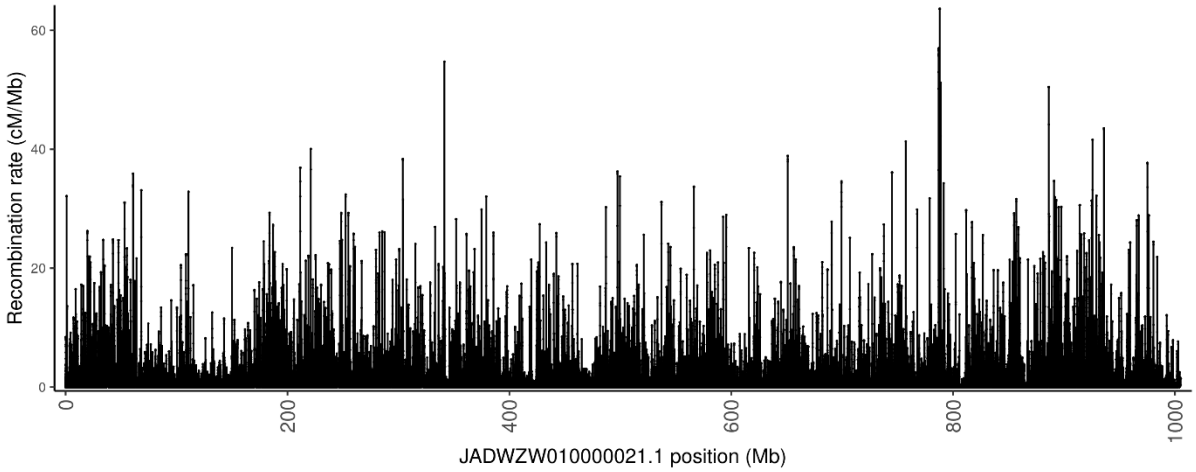

Recombination rate (cM/Mb)
0
10
20
30
40
50
0
200
400
JADWZW010000017.1 position (Mb)

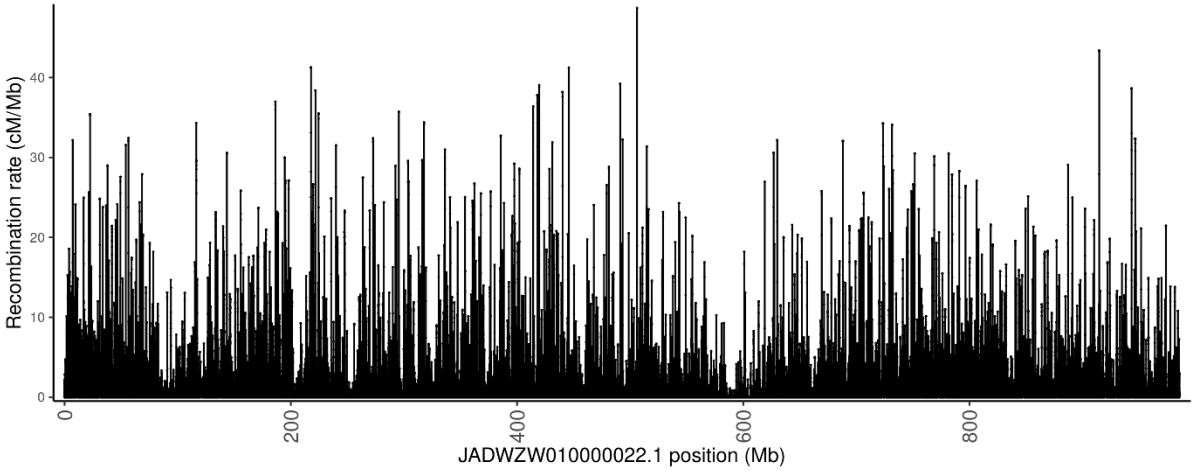

Recombination rate (cM/Mb)
0
10
20
30
40
50
0
200
400
JADWZW010000018.1 position (Mb)


Recombination rate (cM/Mb)
0
10
20
30
40
50
0
200
JADWZW010000019.1 position (Mb)

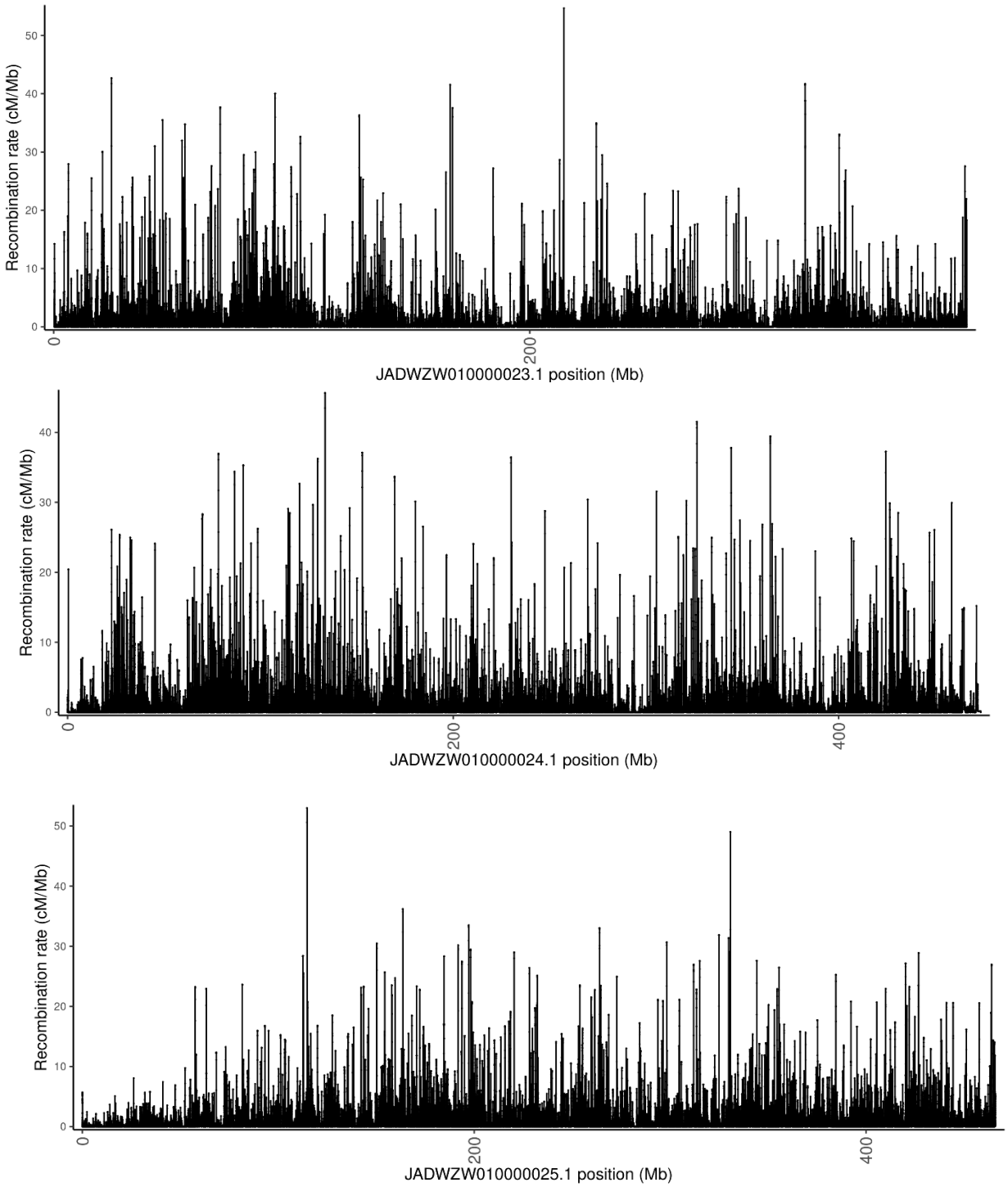
889


Recombination rate (cM/Mb)
0
20
40
60
0
200
JADWZW010000020.1 position (Mb)

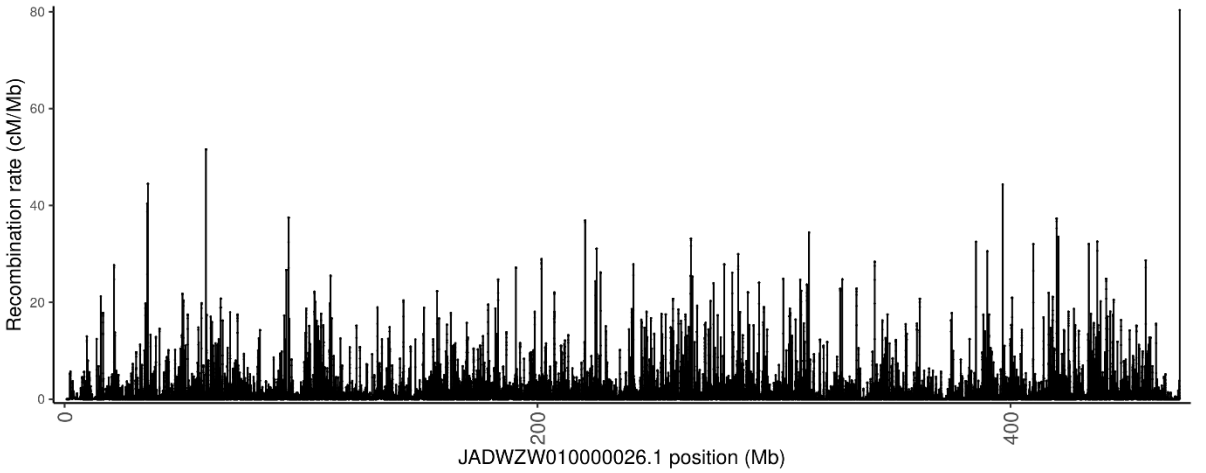

Recombination rate (cM/Mb)
0
20
40
60
0
200
400
600
800
1000
JADWZW010000021.1 position (Mb)

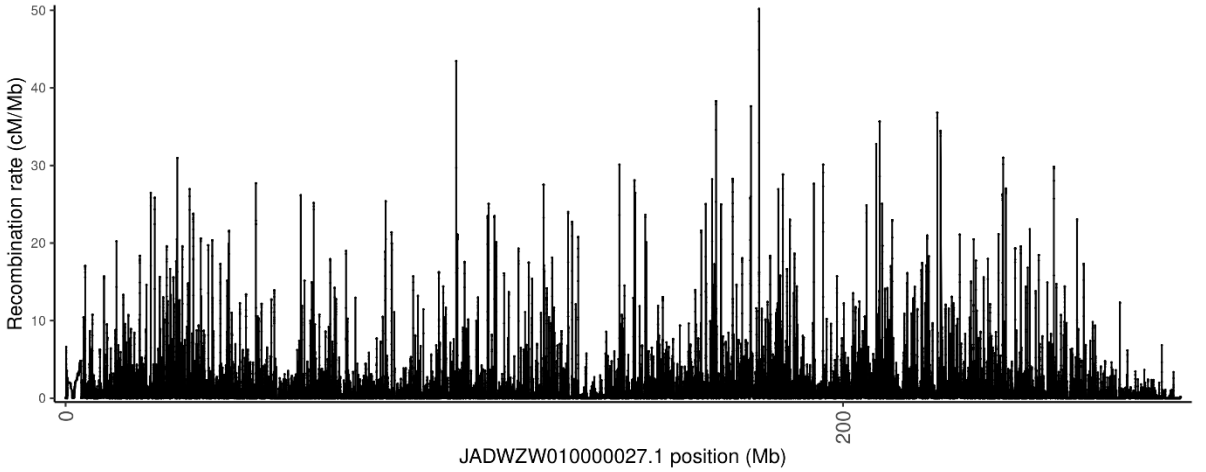

Recombination rate (cM/Mb)
0
10
20
30
40
0
200
400
600
800
JADWZW010000022.1 position (Mb)

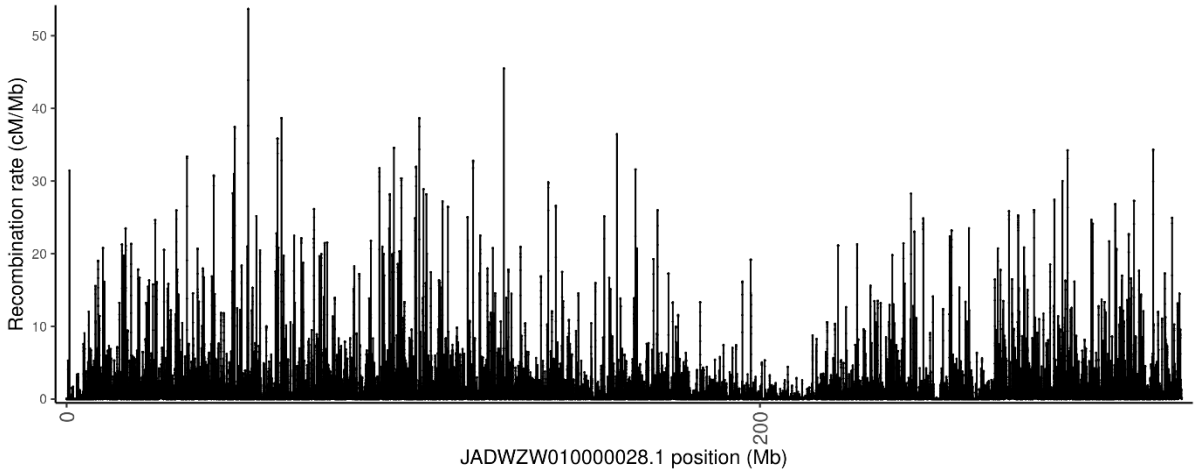
890


Recombination rate (cM/Mb)
0
10
20
30
40
50
0
200
JADWZW010000023.1 position (Mb)
Recombination rate (cM/Mb)
0
10
20
30
40
0
200
400
JADWZW010000024.1 position (Mb)
Recombination rate (cM/Mb)
0
10
20
30
40
50
0
200
400
JADWZW010000025.1 position (Mb)

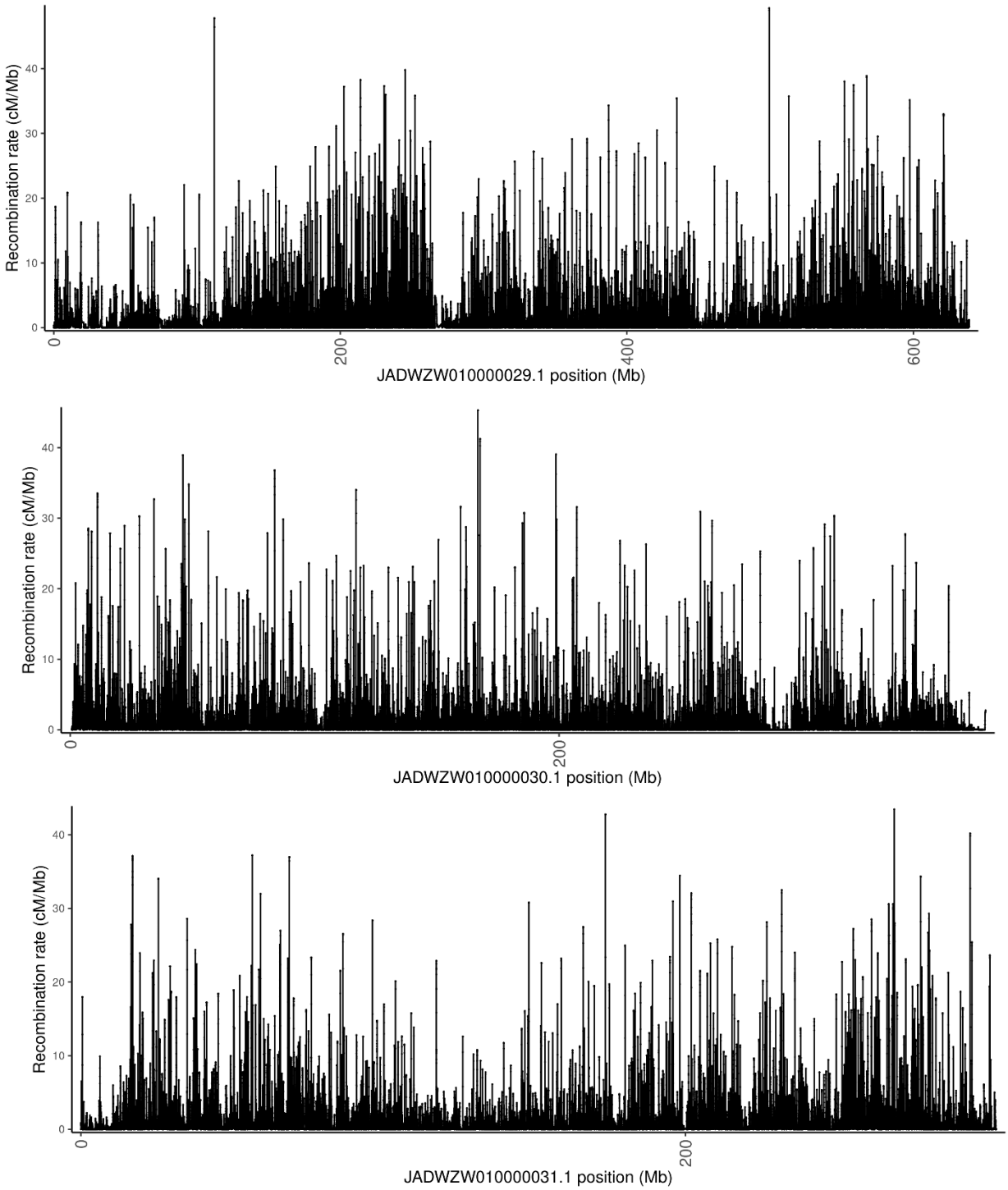
891


Recombination rate (cM/Mb)
0
20
40
60
80
0
200
400
JADWZW010000026.1 position (Mb)


Recombination rate (cM/Mb)
0
10
20
30
40
50
0
200
JADWZW010000027.1 position (Mb)


Recombination rate (cM/Mb)
0
10
20
30
40
50
0
200
JADWZW010000028.1 position (Mb)

892



893 **Fig. S1:** Recombination rates across all 30 donkey autosomes as estimated using LDHat
894 (version 2.2) (*62*).

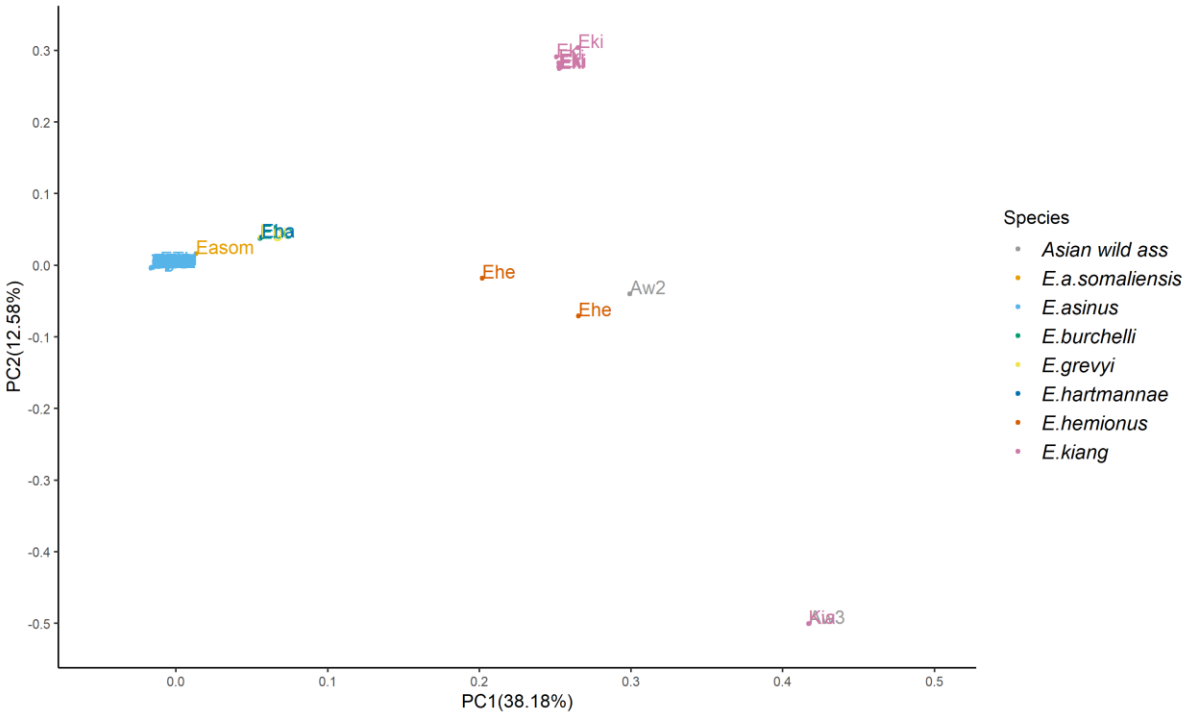


Fig. S2: PCA of domestic donkeys and wild ass species using the phased variant panel (*n*=222 individuals, *n*=13,013,551 variants, TI/TV=2.18) using PLINK (version 1.9) (*63*).

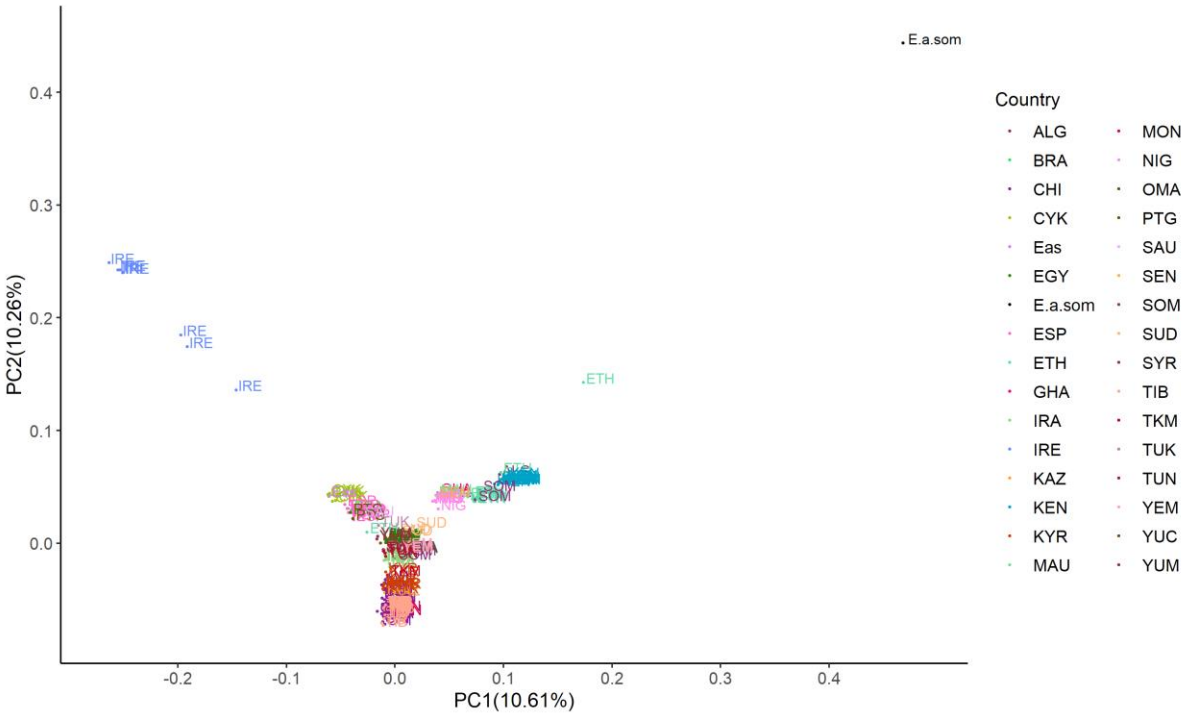
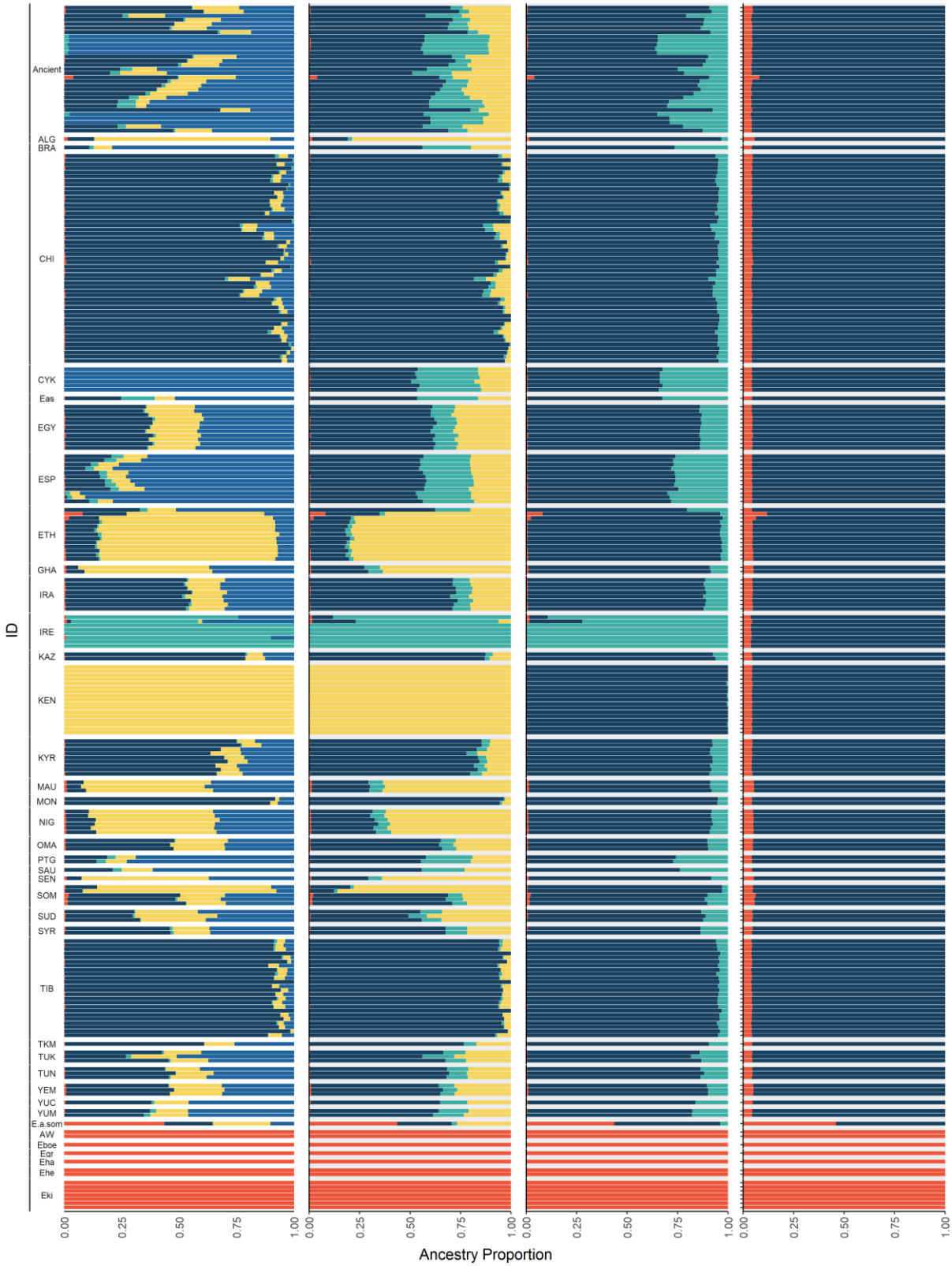


Fig. S3: PCA of domestic donkeys and wild ass species (*E.a.som*) using the phased variant panel (*n*=208 individuals, *n*=13,013,551 variants, TS/TV=2.18) using PLINK (version 1.9) (*63*).



902 **Fig. S4:** ADMIXTURE (version 1.3.0) (*56*) analysis for all modern equids (donkeys and wild
903 asses) and ancient donkeys using the imputed variant panel for K values of 2-5 (*n*=222
904 modern equids, *n*=31 ancient donkeys, *n*=494,050 variants, TI/TV= 2.18, optimal K=4).

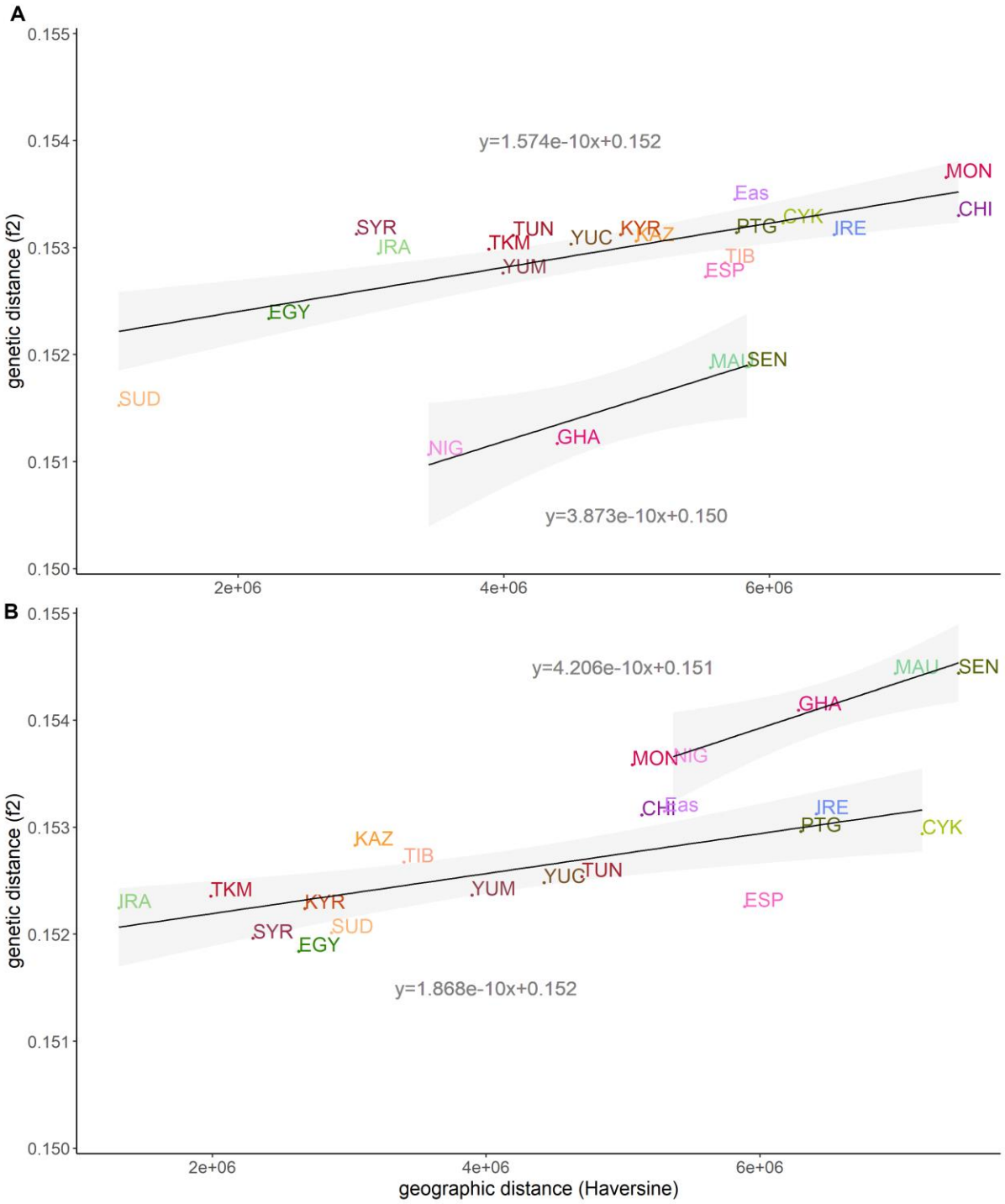


Fig. S5: Genetic distance (f2, estimated using ADMIXTOOLS2) (*124*, *125*) verses geographical distance (estimated as haversine distance) from: A) donkeys from Ethiopia, and B) donkeys from Yemen. Two separate linear regressions were fitted for each dataset: one for subpopulations from western Africa only, and another for all other subpopulations. F2 statistics were estimated for all phased SNPs, but masking regions that were attributed to wild ancestry as estimated using PCAdmix (n=11,577,531 variants, TI/TV=2.18).

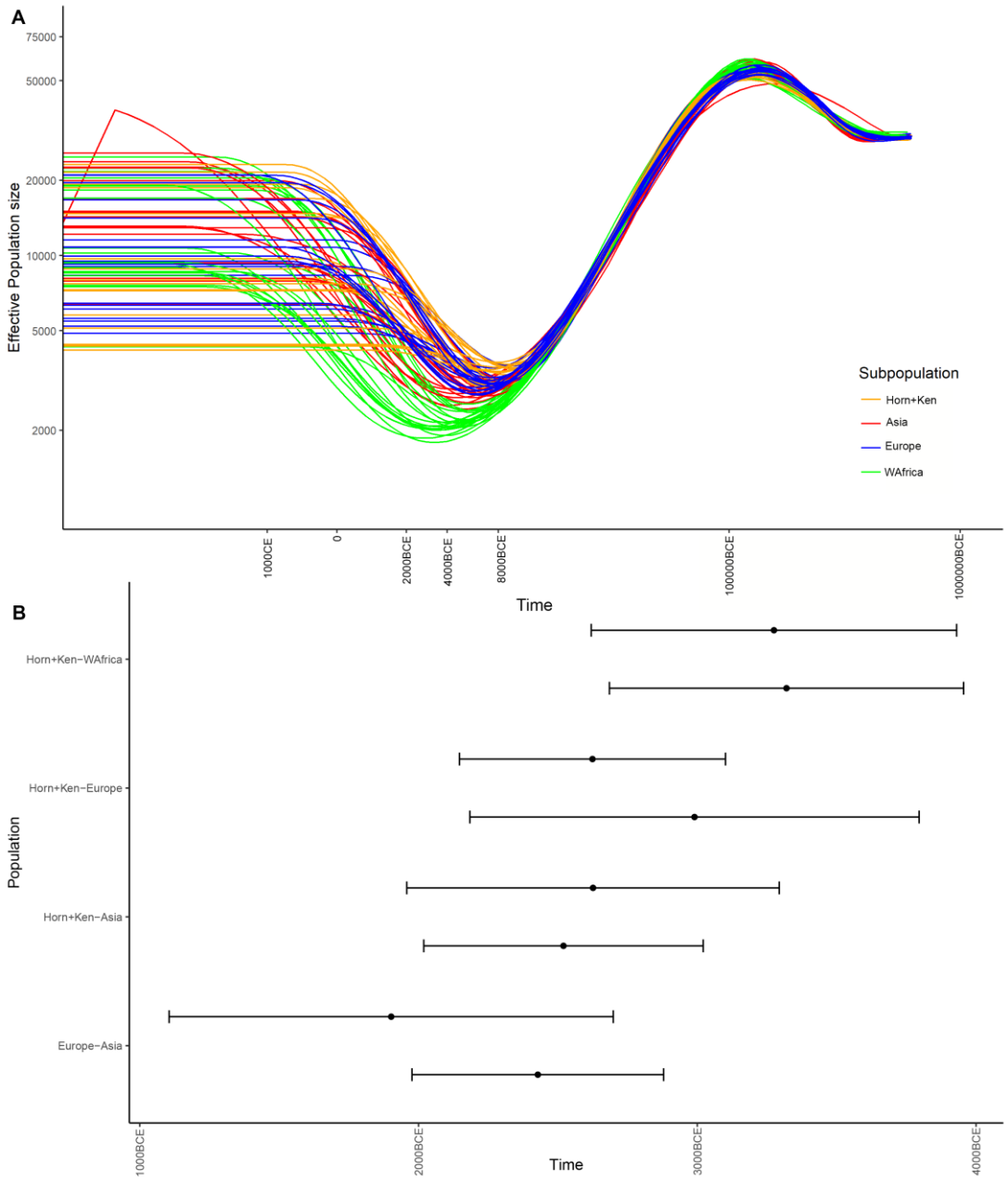


Fig. S6: SMC++ (version 1.15.4) (*28*) population models dating splits from Horn of Africa + Kenya (Horn+Ken), western Africa (WAfrica), Asia (Asia) and Europe (Europe) with an assumed generational time interval of 8 years. Three donkeys from each subpopulation were used, with 10 bootstrap pseudo-replicates (resampling 90% of each chromosome) for two different datasets. Samples used for the first dataset were Horn+Ken: KEN_YPO90, ETH_4, SOM_01, WAfrica: SEN_10, GHA_01, NIG_YPO62, Asia: CHI_KL02A, CHI_GL04A, TIB_DQFS1, Europe: PTGm10, ESP_Andalusian_1, CYK_IslasCanarias_4. The samples used for the second dataset were Horn+Ken: KEN_YPO89, SOM_05, ETH_5, WAfrica: NIG_YPO63, NIG_YPO65, NIG_YPO66, Asia: CHI_JM05A, CHI_XJ6, TIB_XZSNQS07, Europe: PTGm02, ESP_Andalusian_3, ESP_Basque_10. A) Estimated effective population sizes over time (the second dataset is shown in semi transparency). B) Estimated population split times between the subpopulations for the two datasets with standard deviation bars.

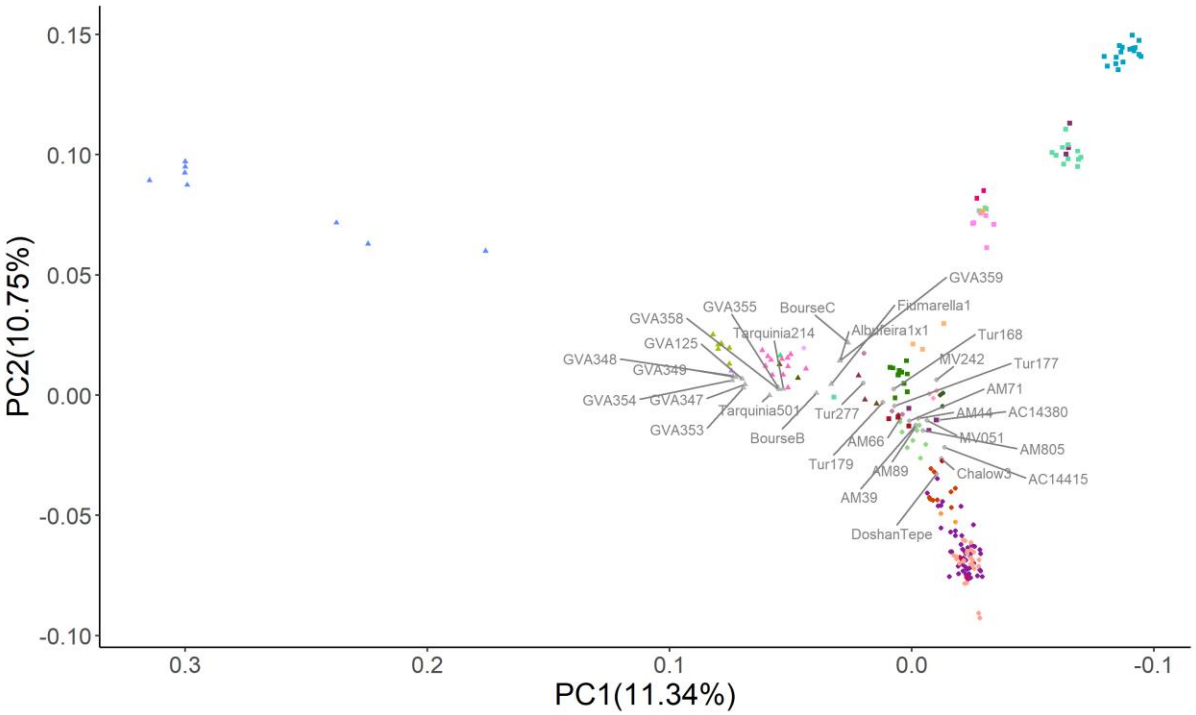


Fig. S7: PCA of ancient imputed donkeys (black, n=31) and modern donkeys (coloured, n=206) using the smartpca program from the EIGENSOFT package (version 6.1.4) (75, 76). The pseudo-haploidized genomes of the ancient donkeys (n=31) were projected onto the PCA and labelled and colored in grey.

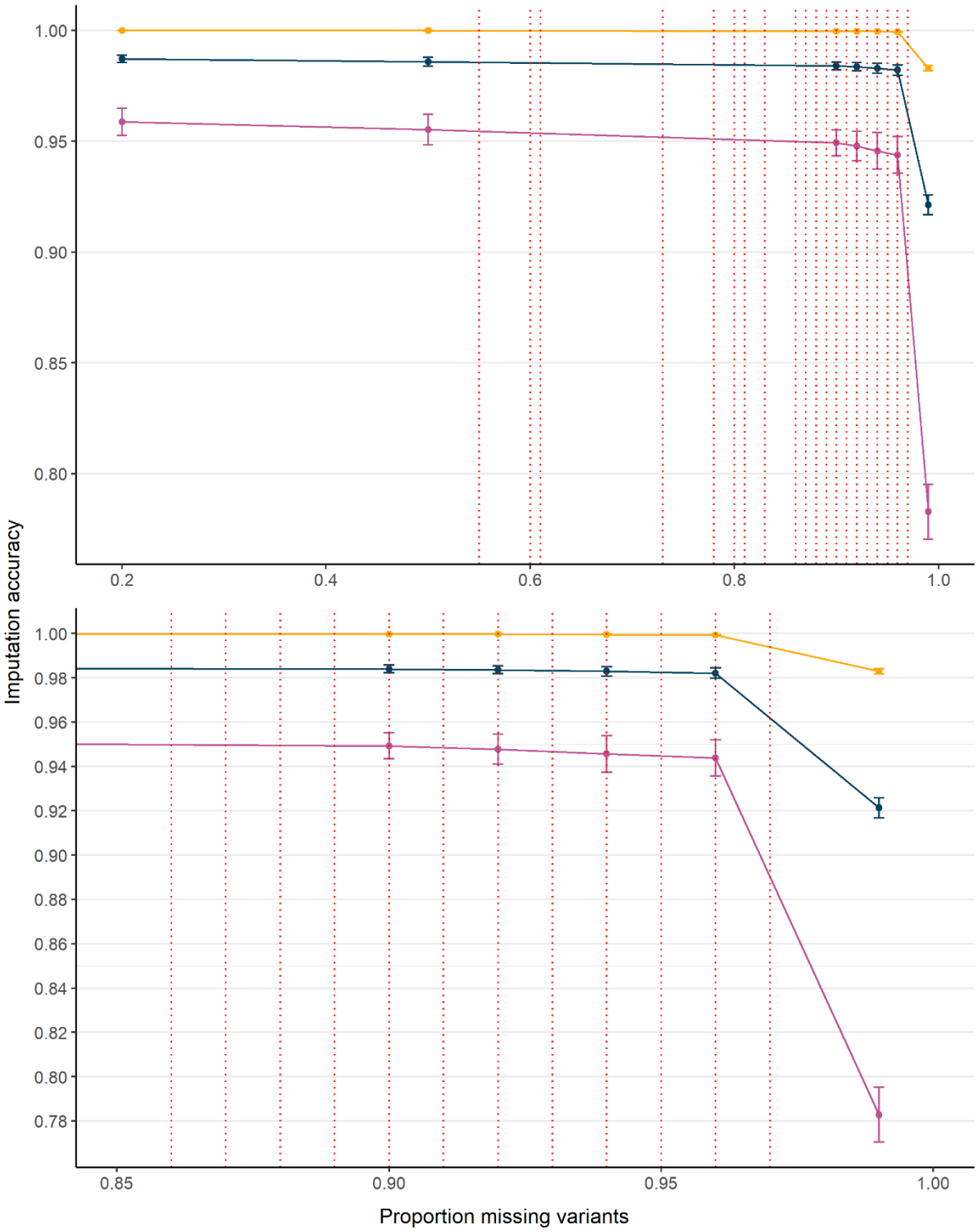


Fig. S8: The imputation accuracy verses the proportion of missing variants before imputation as estimated by downscaling modern donkey variants (*n*=10 individuals). Accuracy of all variants (blue), homozygotes only (yellow) and heterozygotes only (purple) and plotted separately. The same imputation pipeline was used as that to impute the ancient donkey genomes. The proportion of missing variants for each ancient sample (*n*=31) are shown as red dotted lines.

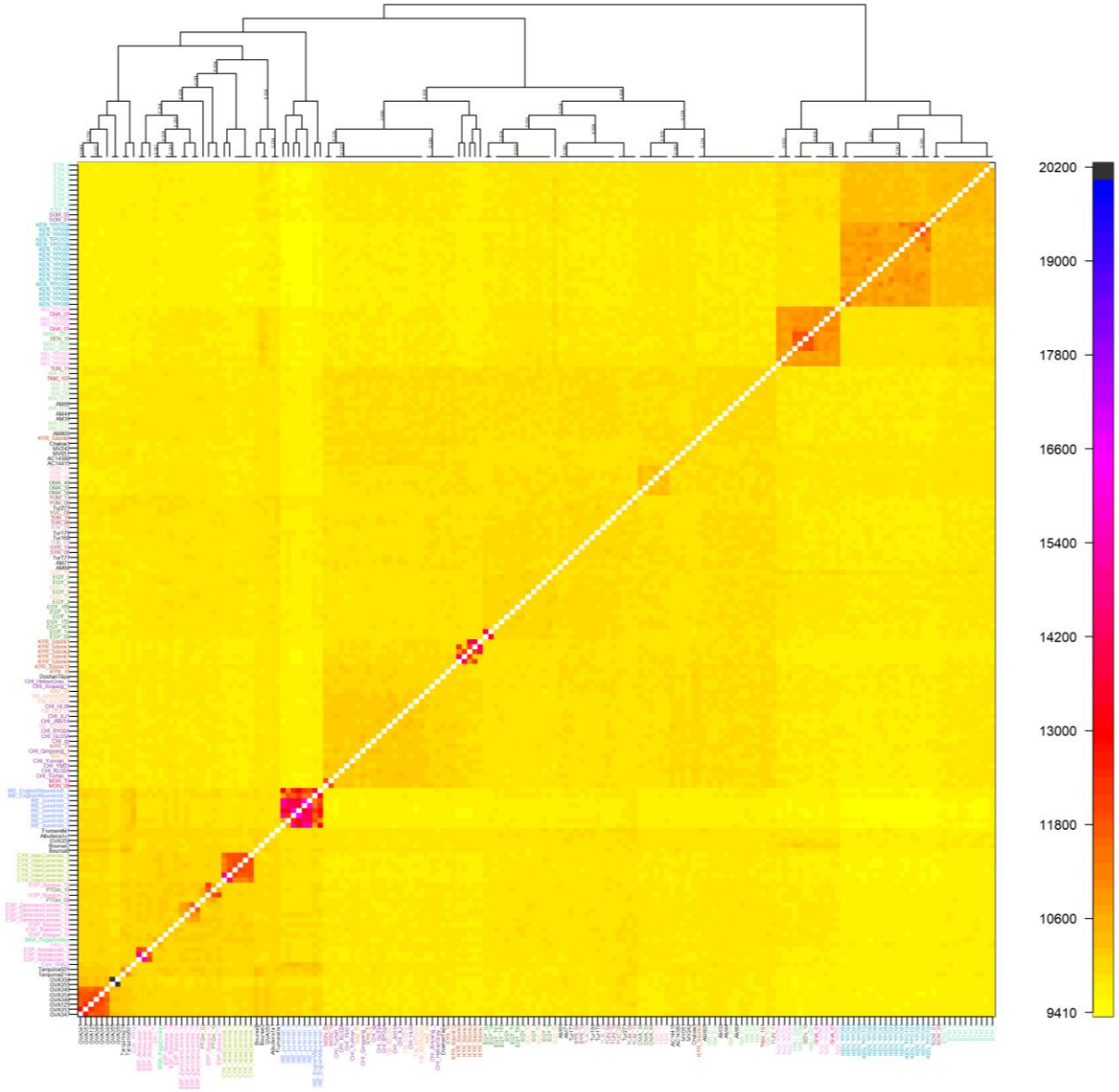
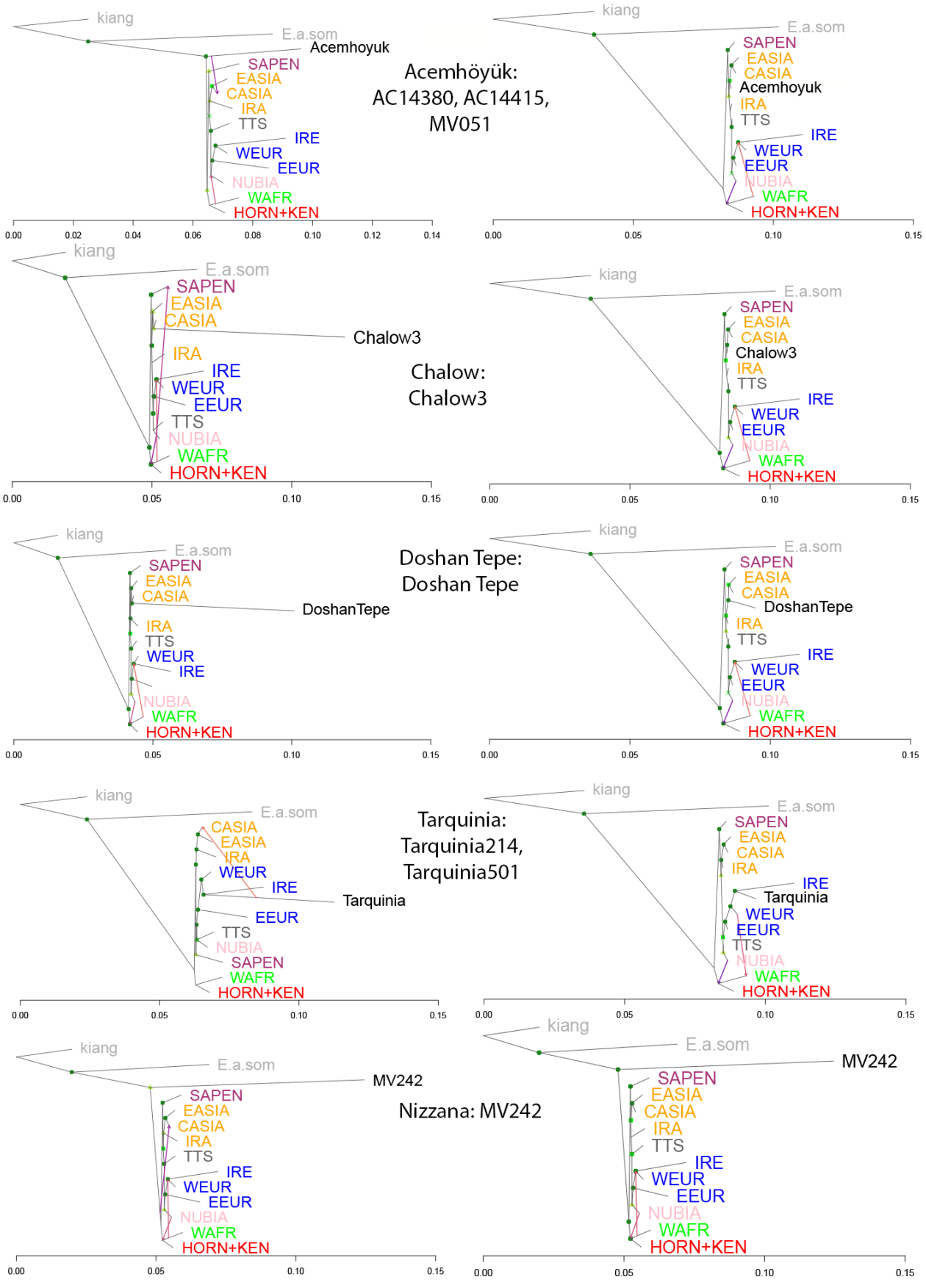
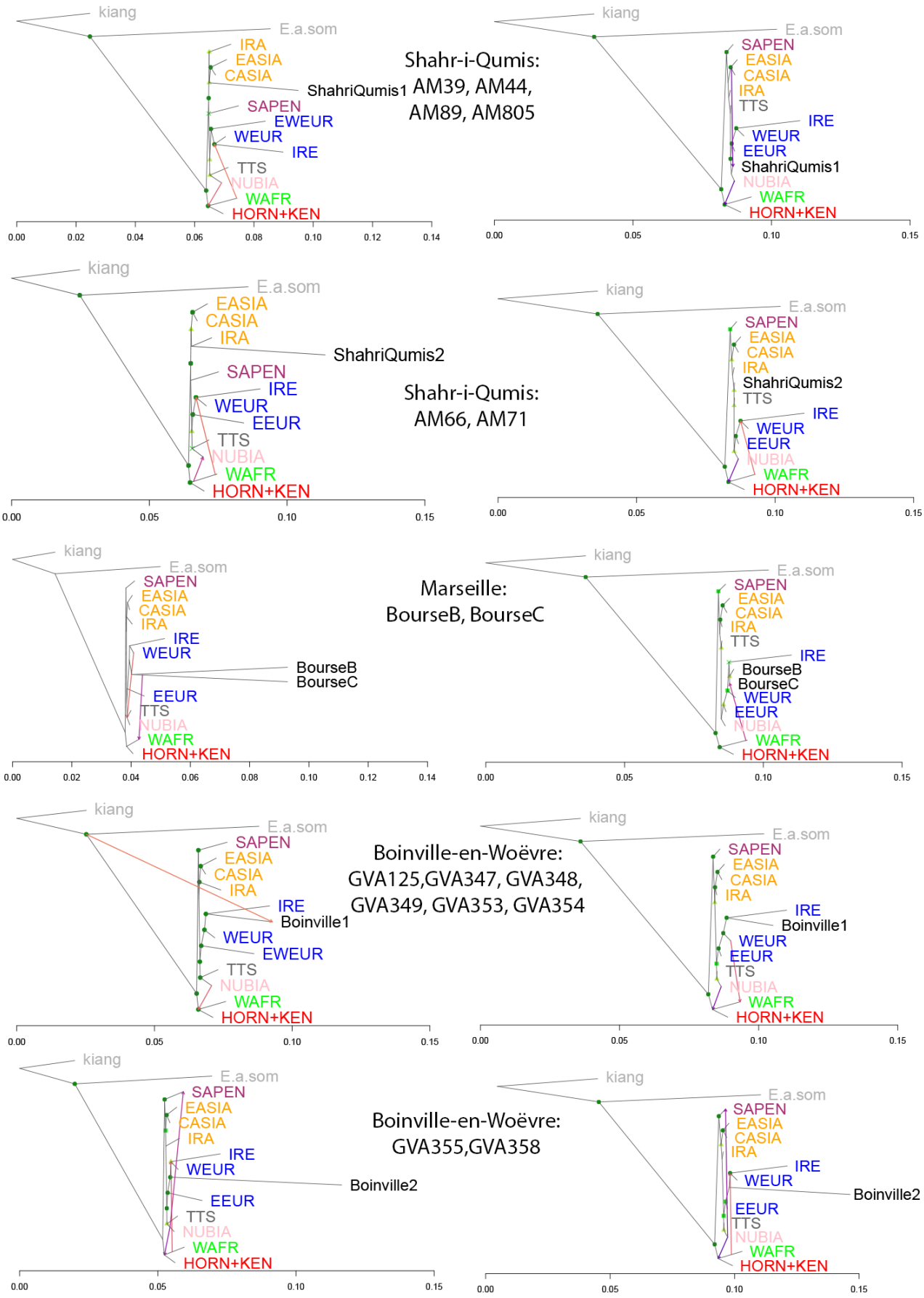


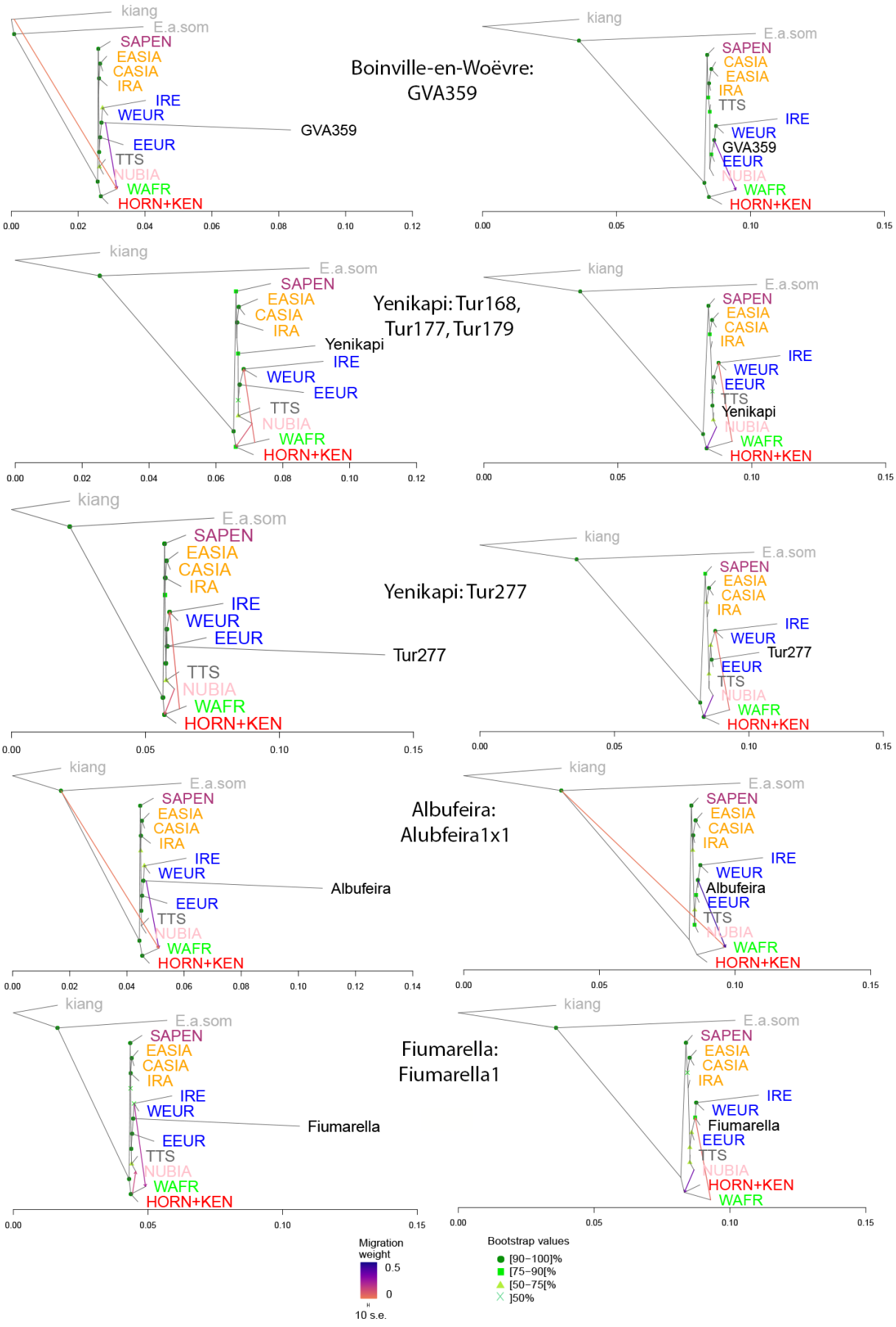
Fig. S9: Maximum likelihood tree and heatmap generated from haplotype sharedness estimated using fineSTRUCTURE (version 4.1.1) for donkeys (n=141 modern and 31 ancient individuals) using imputed variants (*n*=2,245,992, TI/TV= 2.21) (*35*). Only node support values less than 1 are shown on the tree. The heatmap is colour coded according to the number of shared haplotype chunks in the genome.


kiang
E.a.som
Acemhoyuk
SAPEN
EASIA
CASIA
IRA
TTS
IRE
WEUR
EEUR
NUBIA
WAFR
HORN+KEN
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
Acemhöyük:
AC14380, AC14415,
MV051
0.00
0.05
0.10
0.15
Chalow3
Chalow:
Chalow3
DoshanTepe
Doshan Tepe:
Doshan Tepe
Tarquinia
Tarquinia:
Tarquinia214,
Tarquinia501
MV242
Nizzana: MV242

942

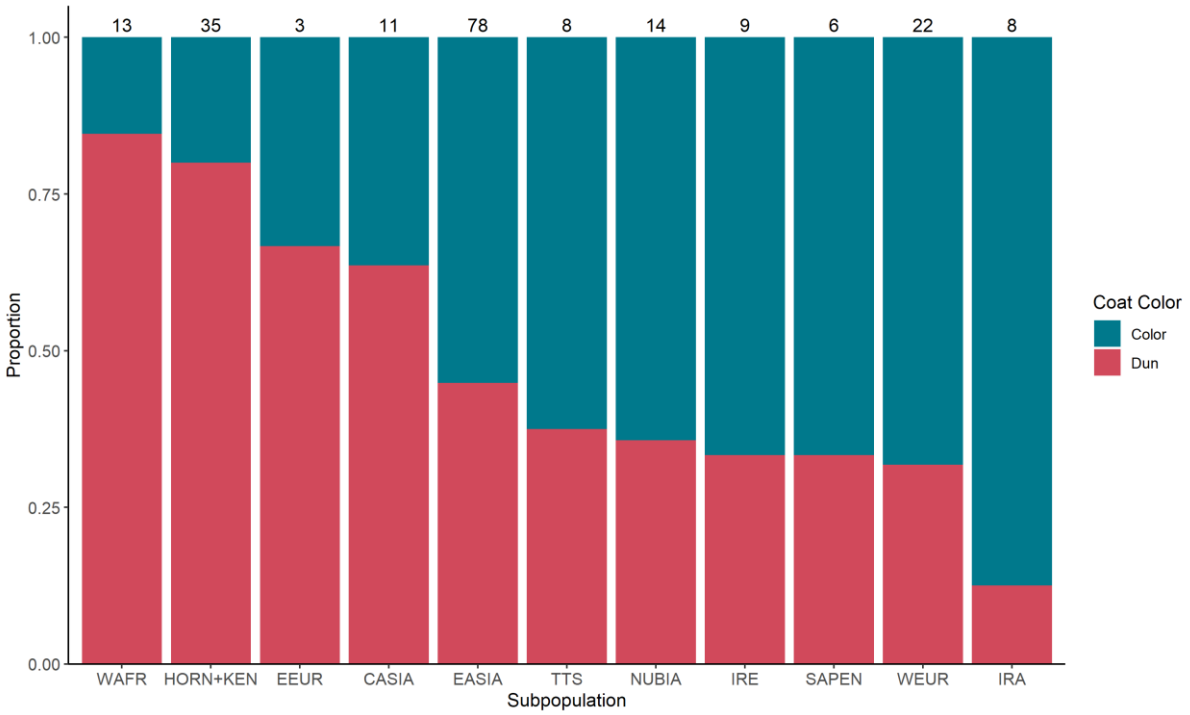

Shahr-i-Qumis:
AM39, AM44,
AM89, AM805
Shahr-i-Qumis:
AM66, AM71
Marseille:
BourseB, BourseC
Boinville-en-Woëvre:
GVA125,GVA347, GVA348,
GVA349, GVA353, GVA354
Boinville-en-Woëvre:
GVA355,GVA358
kiang
E.a.som
IRA
EASIA
CASIA
ShahriQumis1
SAPEN
EWEUR
WEUR
IRE
TTS
NUBIA
WAFR
HORN+KEN
ShahriQumis2
BourseB
BourseC
Boinville1
Boinville2
EEUR
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.05
0.15

943


Boinville-en-Woëvre: GVA359
Yenikapi: Tur168, Tur177, Tur179
Yenikapi: Tur277
Albufeira: Alubfeira1x1
Fiumarella: Fiumarella1
kiang
E.a.som
SAPEN
EASIA
CASIA
IRA
IRE
WEUR
EEUR
TTS
NUBIA
WAFR
HORN+KEN
GVA359
Yenikapi
Tur277
Albufeira
Fiumarella
Migration weight
0.5
0
10 s.e.
Bootstrap values
[90−100]%
[75−90[%
[50−75[%
]50%

944

Fig. S10: Treemix (version 1.13)(*27*) phylogenies for modern donkeys grouped into populations according to Fig. 1C, with kiang as an outgroup. The left column shows the Treemix inferred from pseudo-haploidized variants (*n*=496,697) and the right from imputed variants (*n*=175,093 variants). The trees on each row are from the same site of ancient donkeys, with the site and individuals labelled in the centre of the row. The optimal number of migration edges are shown for each tree, and nodes coloured according to support values from 100 bootstrap replicates.



952 **Fig. S11:** The proportion of modern donkeys with dun and derived coat colors from each
953 subpopulation (*n*=207). The total number of donkeys from each subpopulation is shown
954 above each bar.

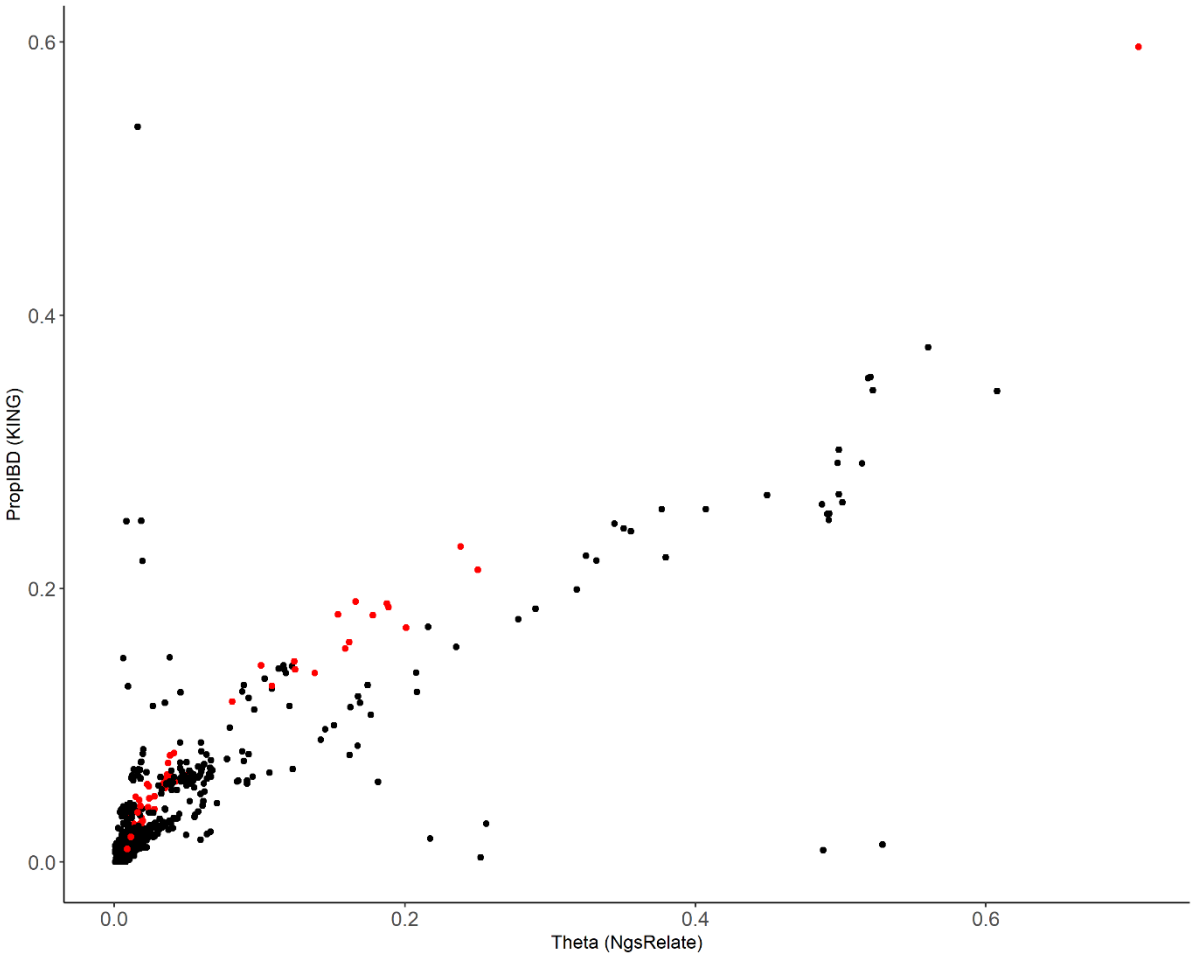


Fig. S12: The relationship between relatedness coefficients calculated using phased and imputed variants in KING (version 2.2.7) (*n*=2,245,992 variants) (*66*) and unimputed variants using NgsRelate (version 2) (*n*=473,263, variants, transversions only) (*67*). Only pairs modern donkey from the same country and ancient donkeys from the same site were included in the analysis (n=2096 pairs). Pairs of ancient donkeys were coloured in red and modern donkeys in black (r=0.871, r^2=0.759).

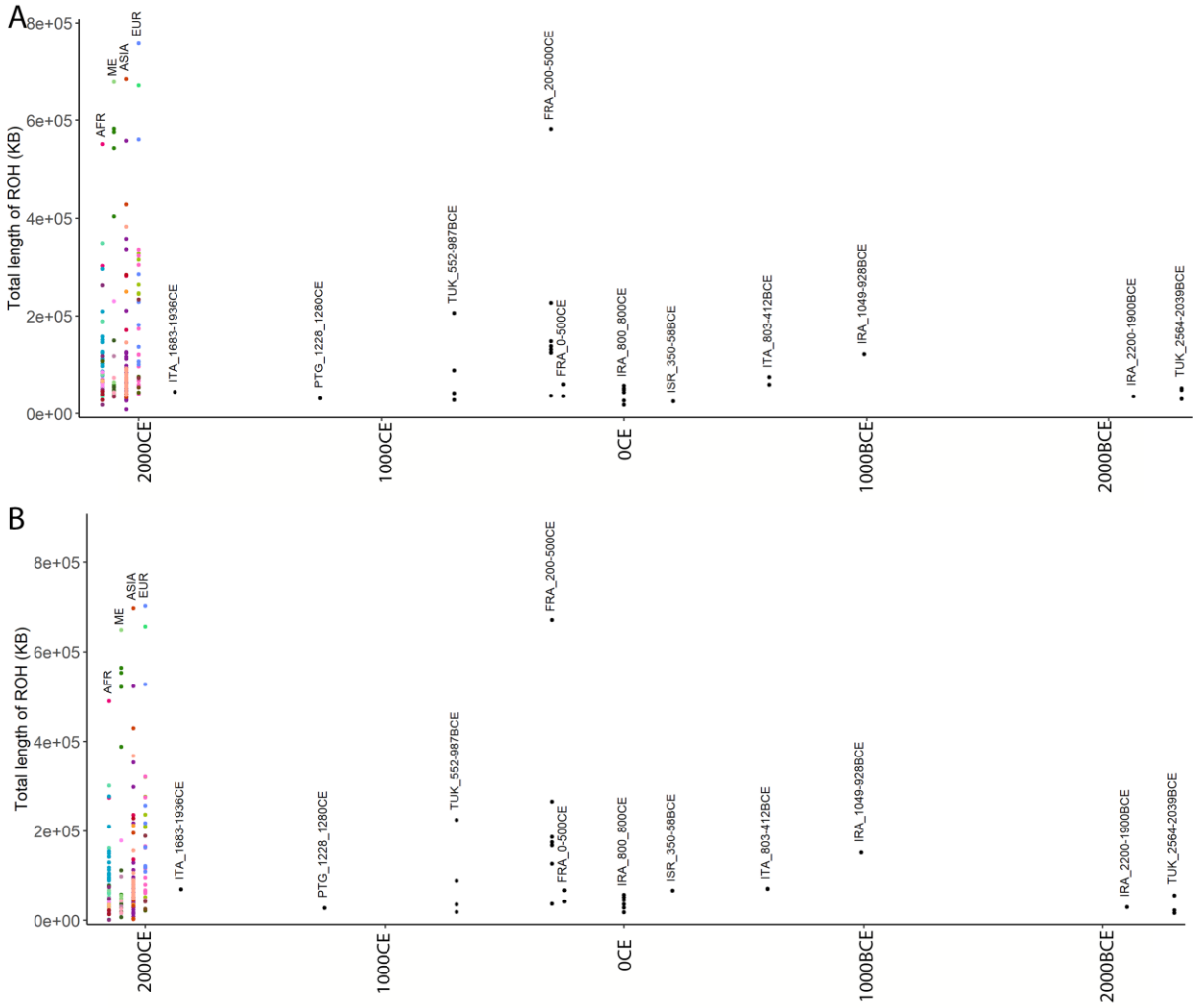
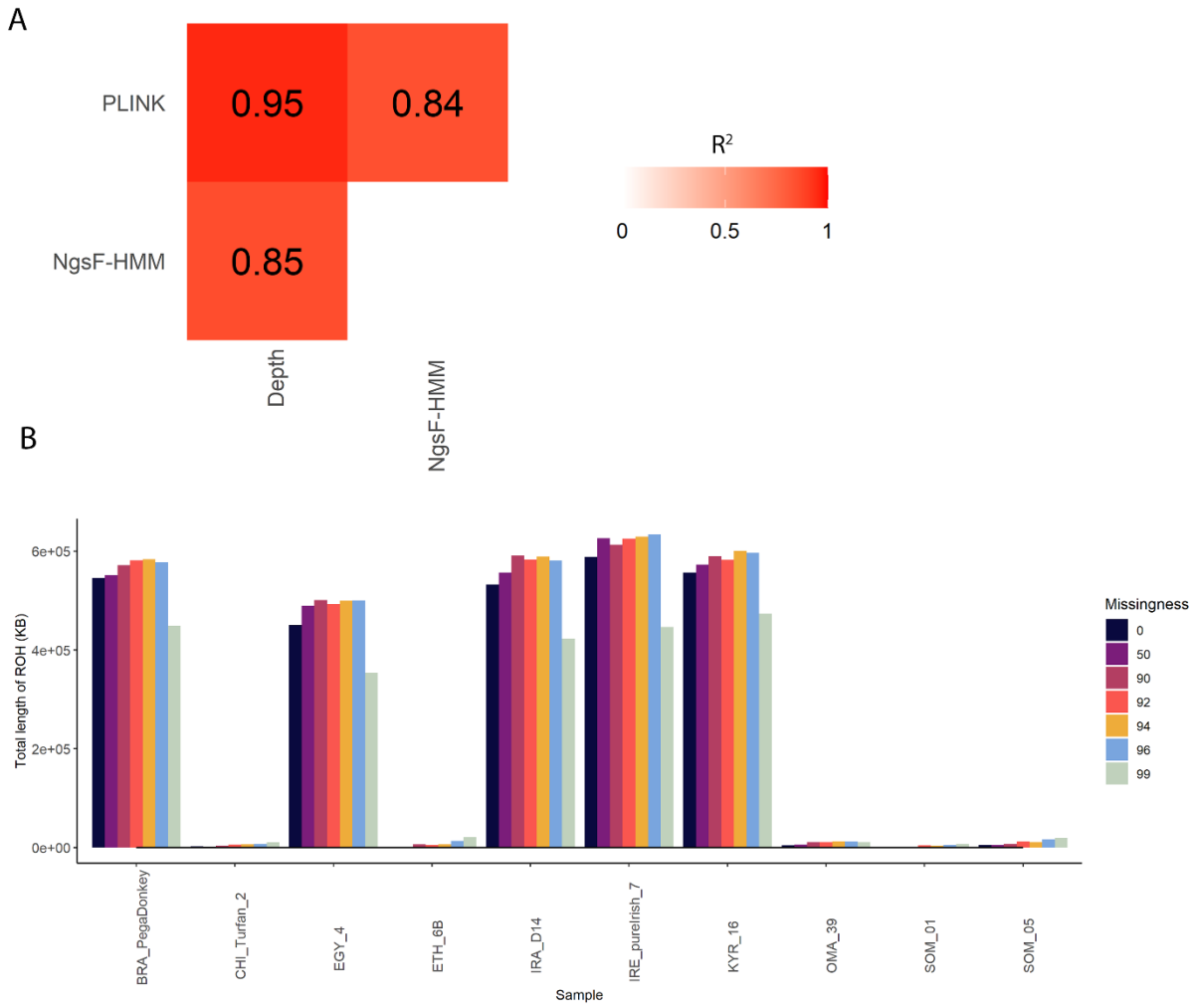
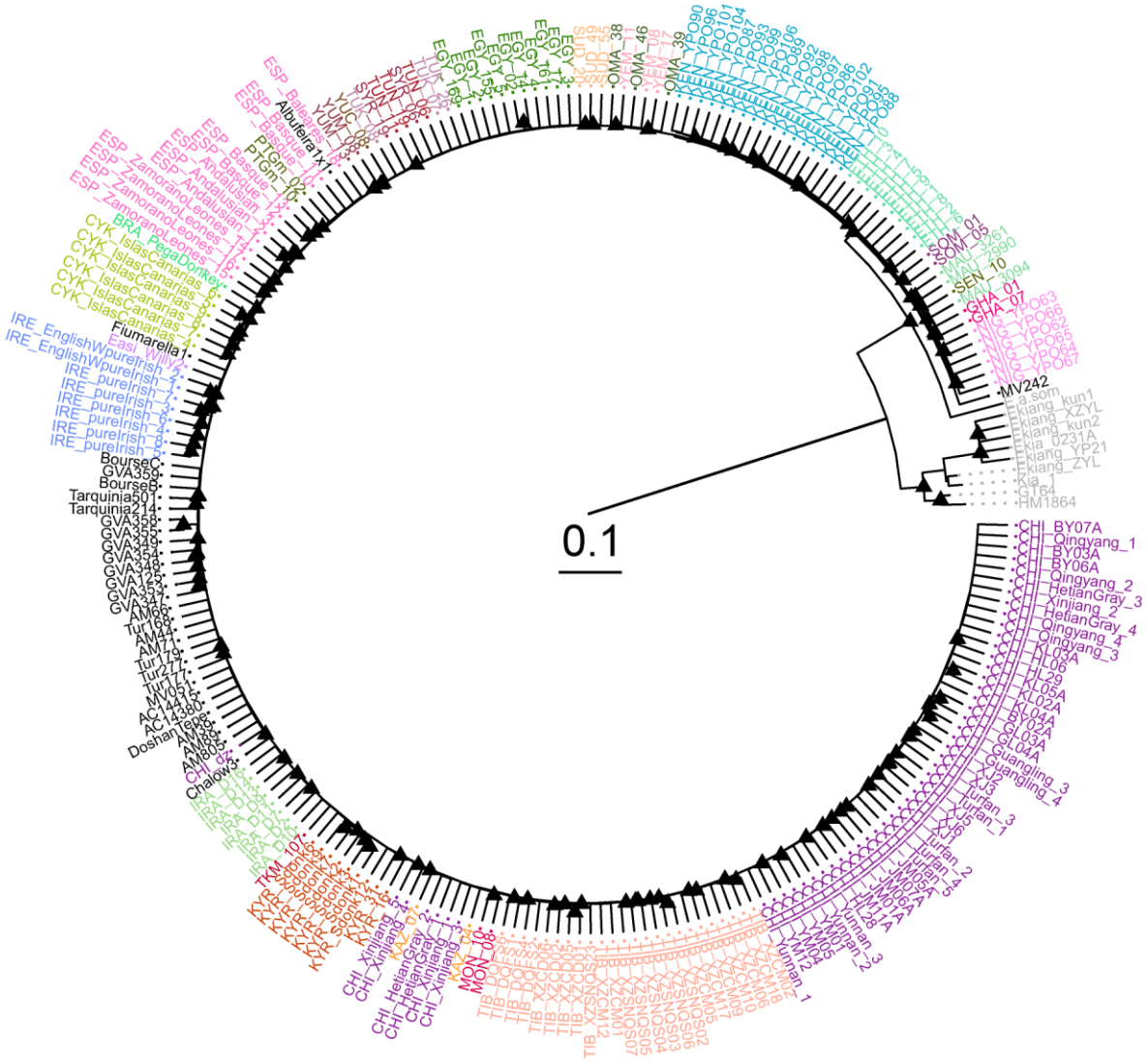


Fig. S13: A) Total length of runs of homozygosity in kilobases, estimated using PLINK (version 1.9) (*63*) plotted as a function of time for all modern and ancient donkeys (*n*=238 individuals), conditioning on transversions only (*n*=1,949,850 variants). B) Total length of runs of homozygosity in kilobases, from depth-based estimated using variants called by ANGSD (version 0.930) (*114*) counts plotted as a function of time for all modern and ancient donkeys (*n*=238 individuals).



969 **Fig. S14:** A) Semi-matrix of squared correlation coefficients between three measures of ROH
970 (PLINK (version 1.9) (*63*), ngsF-HMM (version 1) (*59*) and from depth-based estimated
971 using variants called by ANGSD (version 0.930) (*114*)). B) The total length of ROH
972 estimated in PLINK for 10 modern donkeys after down-sampling and re-imputing variants



973

974 **Fig. S15:** Neighbour joining tree constructed using FastME (version 2.1.4) (*129*) with 100
975 bootstrap pseudo-replicates of modern donkeys, ancient donkeys and kiangs which were
976 included in the Treemix analysis. Two ancient hemippes with coverage over 1X were also
977 included. Bootstrap support values over 90% are labelled with a black triangle.

References

61. H. P. Eggertsson, H. Jonsson, S. Kristmundsdottir, E. Hjartarson, B. Kehr, G. Masson *et al.*, Graphtyper enables population-scale genotyping using pangenome graphs. *Nat. Genet.* **49**, 1654-1660 (2017).
62. A. Auton, G. McVean, Recombination rate estimation in the presence of hotspots. *Gen. Res.* **17**, 1219-1227 (2007).
63. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender *et al.*, PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575 (2007).
64. W. Haak, I. Lazaridis, N. Patterson, N. Rohland, S. Mallick, B. Llamas *et al.*, Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207-211 (2015).
65. B. M. Peter, Admixture, population structure, and f-statistics. *Genetics* **202**, 1485 (2016).
66. A. Manichaikul, J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale, W. M. Chen, Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-2873 (2010).
67. K. Hanghøj, I. Moltke, P. A. Andersen, A. Manica, T. S. Korneliussen, Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *GigaScience* **8**, (2019).
68. R. Bouckaert, T. G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina *et al.*, BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).
69. A. J. Drummond, S. Y. W. Ho, M. J. Phillips, A. Rambaut, Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).
70. A. J. Drummond, A. Rambaut, B. Shapiro, O. G. Pybus, Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185-1192 (2005).
71. S. Rosenbom, V. Costa, N. Al-Araimi, E. Kefena, A. S. Abdel-Moneim, M. A. Abdalla *et al.*, Genetic diversity of donkey populations from the putative centers of domestication. *Anim. Genet.* **46**, 30-36 (2015).
72. D. Cook, S. Brooks, R. Bellone, E. Bailey, Missense mutation in exon 2 of SLC36A1 responsible for champagne dilution in horses. *PLoS Genet.* **4**, e1000195 (2008).
73. L. Zeng, H. Q. Liu, X. L. Tu, C. M. Ji, X. Gou, A. Esmailizadeh *et al.*, Genomes reveal selective sweeps in kiang and donkey for high-altitude adaptation. *Zool Res* **42**, 450-460 (2021).
74. S. W. Manning, L. Wacker, U. Büntgen, C. Bronk Ramsey, M. W. Dee, B. Kromer *et al.*, Radiocarbon offsets and old world chronology as relevant to Mesopotamia, Egypt, Anatolia and Thera (Santorini). *Sci. Rep.* **10**, 13785 (2020).
75. N. Özgüç, "Seal impressions from the palaces at Acemhöyük" in *Ancient Art in Seals*, E. Porada, Ed. (Princeton University Press, Princeton, 1980), pp. 61-99.
76. Y. Kamis, "Acemhöyük Buluntuları Işığında Erken Tunç Çağı'nda Orta Anadolu'nun Güneyinde Çark Yapımı Seramiğin Ortaya Çıkışı" in *Adalya* (Umran Savaş İnan, Istanbul, 2018), vol. 21, pp. 59-84.
77. A. Öztan, 2010 Yılı Acemhöyük Kazıları. *Kazı Sonuçları Toplantısı* 393-412 (2012).
78. A. Öztan, 2013 Yılı Acemhöyük Kazıları ve Sonuçları. *Kazı Sonuçları Toplantısı* **36**, 61-72 (2014).
79. Y. S. Erdal, K. Özdemir, Ö. D. Erdal, "Acemhöyük'ten Bir İnsan İskeletinde Saptanan Yaralanmaların Adli Antropolojik Açıdan İncelenmesi. " in *Samsat'tan*

Acemhöyük'e Eski Uygarlıkların Izinde: Prof. Dr Aliye Öztan'a Armağan, S. Özkan, H. Hüryılmaz, A. Türker, Eds. (Ege Üniversitesi Basımevi, Izmir, 2017), pp. 105-119.
80. Y. Kamış, A. Öztan, 2018 Yılı Acemhöyük Kazıları. Kazı Sonuçları Toplantısı *Kazı Sonuçları Toplantısı* **41**, 147-160 (2020).
81. A. Vahdati, R. Biscione, R. Le Farina, M. Mashkour, M. Tengberg, H. Fathi *et al.*, "Preliminary report on the first season of excavations at Tepe Chalow: New GKC (BMAC) finds in the plain of Jajarm, NE Iran." in *The Iranian Plateau during the Bronze Age. Development of Urbanisation, Production and Trade,* J. W. Meyer, E. Vila, M. Mashkour, M. Casanova, R. Vallet, Eds. (MOM Éditions, 2019), pp. 179-200.
82. M. Mashkour, A. F. Mohaseb, "Hunting and husbandry in the Ozbaki archaeological Zone (Savojbolagh plain) from the 6th millennium until the Iron Age: Archaeozoological study of Jeiran Tepe, Maral Tepe, Doshan Tepe and Tepe Ozbaki" in *The Archaeological Excavation of Ozbaki. Vol. 1. Art and Architecture* (ICHTO Editions, 2011), pp. 273−302, 597−601.
83. Bagnasco Gianni G., A. Garzulino, M. Marzullo, "The last ten years of research at Tarquinia" in *Knowledge, Analysis and Innovative Methods for the Study and the Sissemination of Ancient Urban Areas, Proceedings of the KAINUA 2017 International Conference in Honour of Professor Giuseppe Sassatelli's 70th Birthday (Bologna, 18-21 Aprile 2017),* S. Garagnani, A. Gaucci, Eds. (CNR - Istituto di Scienze del Patrimonio Culturale, 2017), vol. Archeologia e Calcolatori 28.2, pp. 211-221.
84. A. Negev, "Nessana" in *The New Enclyclopedia of Archaeological Excavations in the Holy Land,* E. Stern, Ed. (Israel Exploration Society Jerusalem, 1993), pp. 1145–1149.
85. D. Urman, *Nessana: Excavations and Studies* (Ben Gurion University, Beer Sheva, 2004), vol. 7.
86. G. Avni, *The Byzantine–Islamic transition in Palestine* (Oxford University Press, Oxford, 2014).
87. J. Hansman, D. Stronach, Excavations at Shahr-i Qūmis, 1967. *J. R. Asiat. Soc.* **102**, 29-62 (1970).
88. J. Hansman, D. Stronach, A Sasanian Repository at Shahr-i Qūmis. *J. R. Asiat. Soc.*, 142-155 (1970).
89. M. Mashkour, H. Davoudi, F. A. Mohaseb, D. S. Beizaee, R. Khazaeli, S. Amiri *et al.*, "Human and animal interactions in the Iranian Plateau. Research conducted by the Osteology Department of Iran National Museum. " in *Iran National Museum publications and Institut Français de Recherche en Iran, Bibliothèque iranienne N°85* (2021), pp. 86-99.
90. L. Jourdan, *La faune du site gallo-romain et paléo-chrétien de la Bourse (Marseille)* (Editions du CNRS, France, 1976), vol. 1.
91. V. Onar, H. Alpak, G. Pazvant, A. Armutak, A. Chrószcz, Byzantine horse skeletons of Theodosius harbour: 1. Paleopathology. *Rev. Med. Vet.* **163**, 139-146 (2012).
92. V. Onar, G. Pazvant, H. Alpak, N. G. Ince, A. Armutak, Z. Kiziltan, Animal skeletal remains of the Theodosius harbor: general overview. *Turkish J. Vet. Anim. Sci.* **37**, 81-85 (2013).
93. V. Onar, G. Pazvant, E. Pasicka, A. Armutak, H. Alpak, Byzantine horse skeletons of Theodosius Harbour: 2. Withers height estimation. *Rev. Med. Vet.* **166**, (2015).
94. M. T. Antunes, A. C. Balbino, P. M. Callapez, E. Crespo, P. Legoinha, P. R. Mein *et al., Silo Islâmico de Albufeira (Rua Henrique Calado). Estudos Arqueozoológicos e*

Arqueobotânicos. (Instituto de Arqueologia e Paleociências (IAP). Universidade Nova de Lisboa, Lisbon, Portugal, 2012).
95. V. Tinè, "Gli scavi nel Riparo della Fiumarella di Tortora (Cosenza)" in *Proceedings of the XXXVII Scientific Meeting "Prehistory and Protohistory of Calabria" (Scalea, Papasidero, Praia a Mare, Tortora, 29 September - 4 October 2002)* (IIPP - Italian Institute of Prehistory and Protohistory, Florence, 2004), pp. 781-789.
96. A. Curci, I resti faunistici dell'insediamento dell'età del Bronzo di Madonna del Petto, scavi 1977. *Taras* **XV**, 204-215 (1995).
97. P. Farello, I reperti faunistici. *Primi Insediamenti sul Monte Titano. Scavi e Ricerche (1997-2004)*, 87-95 and 135-140 (2009).
98. A. Riedel, Notizie preliminari sullo studio della fauna di Spina. *Atti dell'Accademia delle Scienze di Ferrara* **55**, 1-7 (1978).
99. G. Siracusano, La fauna del Bronzo tardo del sito stratificato di Coppa Nevigata: una visione d'insieme. *Origini XV*, 201-217 (1992).
100. A. Seguin-Orlando, R. Donat, C. Der Sarkissian, J. Southon, C. Thèves, C. Manen *et al.*, Heterogeneous hunter-gatherer and steppe-related ancestries in late Neolithic and bell beaker genomes from present-day France. *Curr. Biol.* **31**, 1072-1083.e1010 (2021).
101. C. Gamba, K. Hanghøj, C. Gaunitz, A. H. Alfarhan, S. A. Alquraishi, K. A. Al-Rasheid *et al.*, Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Mol. Ecol. Resour.* **16**, 459-469 (2016).
102. N. Rohland, E. Harney, S. Mallick, S. Nordenfelt, D. Reich, Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **370**, 20130624 (2015).
103. P. J. Reimer, W. E. N. Austin, E. Bard, A. Bayliss, P. G. Blackwell, C. Bronk Ramsey *et al.*, The IntCal20 northern hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* **62**, 725-757 (2020).
104. C. Bronk Ramsey, Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**, 337-360 (2016).
105. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
106. M. Schubert, H. Jónsson, D. Chang, C. Der Sarkissian, L. Ermini, A. Ginolhac *et al.*, Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E5661-5669 (2014).
107. M. Poullet, L. Orlando, Assessing DNA sequence alignment methods for characterizing ancient genomes and methylomes. *Front. Ecol. Evol.* **8**, (2020).
108. P. Skoglund, B. H. Northoff, M. V. Shunkov, A. P. Derevianko, S. Pääbo, J. Krause *et al.*, Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2229-2234 (2014).
109. H. Jónsson, A. Ginolhac, M. Schubert, P. L. Johnson, L. Orlando, mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682-1684 (2013).
110. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer *et al.*, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
111. E. Garrison, Z. N. Kronenberg, E. T. Dawson, B. S. Pedersen, P. Prins, Vcflib and tools for processing the VCF variant call format. *bioRxiv*, 2021.2005.2021.445151 (2021).
112. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Gen. Res.* **20**, 1297-1303 (2010).

113. L. Orlando, A. Ginolhac, G. Zhang, D. Froese, A. Albrechtsen, M. Stiller *et al.*, Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**, 74-78 (2013).
114. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: analysis of next generation sequencing data. *BMC Bioinform.* **15**, 356 (2014).
115. S. K. Beeson, J. R. Mickelson, M. E. McCue, Equine recombination map updated to EquCab3.0. *Anim Genet* **51**, 341-342 (2020).
116. F. Alhaique, F. Marshall, Preliminary report on the Jebel Gharbi fauna from site SJ-00-56 (2000 and 2002 excavations). *Africa* **54**, 498-507 (2009).
117. F. Marshall, L. Weissbrod, Domestication processes and morphological change: through the lens of the donkey and African pastoralism. *Curr. Anthropol.* **52**, S397-S413 (2011).
118. L. Shackelford, F. Marshall, J. Peters, Identifying donkey domestication through changes in cross-sectional geometry of long bones. *J. Archaeol. Sci.* **40**, 4170-4179 (2013).
119. Y. X. Zhao, J. Yang, F. H. Lv, X. J. Hu, X. L. Xie, M. Zhang *et al.*, Genomic reconstruction of the history of native sheep reveals the peopling patterns of nomads and the expansion of early pastoralism in East Asia. *Mol. Biol. Evol.* **34**, 2380-2395 (2017).
120. J. E. Decker, S. D. McKay, M. M. Rolf, J. Kim, A. Molina Alcalá, T. S. Sonstegard *et al.*, Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet.* **10**, e1004254 (2014).
121. M. P. Verdugo, V. E. Mullin, A. Scheu, V. Mattiangeli, K. G. Daly, P. Maisano Delser *et al.*, Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. *Science* **365**, 173-176 (2019).
122. M. Milanesi, S. Capomaccio, E. Vajana, L. Bomba, J. F. Garcia, P. Ajmone-Marsan *et al.*, BITE: an R package for biodiversity analyses. *bioRxiv*, 181610 (2017).
123. A. Brisbin, K. Bryc, J. Byrnes, F. Zakharia, L. Omberg, J. Degenhardt *et al.*, PCAdmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Hum. Biol.* **84**, 343-364 (2012).
124. D. Reich, K. Thangaraj, N. Patterson, A. L. Price, L. Singh, Reconstructing Indian population history. *Nature* **461**, 489-494 (2009).
125. G. Bhatia, N. Patterson, S. Sankararaman, A. L. Price, Estimating and interpreting Fst: the impact of rare variants. *Gen. Res.* **23**, 1514-1521 (2013).
126. S. Schiffels, K. Wang, MSMC and MSMC2: The multiple sequentially markovian coalescent. *Methods Mol. Biol.* **2090**, 147-166 (2020).
127. Z. Zheng, X. Wang, M. Li, Y. Li, Z. Yang, X. Wang *et al.*, The origin of domestication genes in goats. *Sci. Adv.* **6**, eaaz5216 (2020).
128. C. Michel, The Old Assyrian trade in the light of Recent Kültepe Archives. *Journal of the Canadian Society for Mesopotamian Studies* **3**, 71-82 (2008).
129. V. Lefort, R. Desper, O. Gascuel, FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* **32**, 2798-2800 (2015).
130. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901-904 (2018).
131. P. Librado, C. Gamba, C. Gaunitz, C. Der Sarkissian, M. Pruvost, A. Albrechtsen *et al.*, Ancient genomic changes associated with domestication of the horse. *Science* **356**, 442-445 (2017).

132. X. Liu, Y. Zhang, Y. Li, J. Pan, D. Wang, W. Chen *et al.*, EPAS1 gain-of-function mutation contributes to high-altitude adaptation in Tibetan horses. *Mol. Biol. Evol.* **36**, 2591-2603 (2019).
133. V. Jagannathan, V. Gerber, S. Rieder, J. Tetens, G. Thaller, C. Drögemüller *et al.*, Comprehensive characterization of horse genome variation by whole-genome sequencing of 88 horses. *Anim. Genet.* **50**, 74-77 (2019).
134. L. S. Andersson, M. Larhammar, F. Memic, H. Wootz, D. Schwochow, C.-J. Rubin *et al.*, Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. *Nature* **488**, 642-646 (2012).
135. B. Wallner, N. Palmieri, C. Vogl, D. Rigler, E. Bozlak, T. Druml *et al.*, Y Chromosome Uncovers the Recent Oriental Origin of Modern Stallions. *Curr. Biol.* **27**, 2029-2035.e2025 (2017).
136. C. Der Sarkissian, L. Ermini, M. Schubert, M. A. Yang, P. Librado, M. Fumagalli *et al.*, Evolutionary genomics and conservation of the endangered Przewalski's horse. *Curr. Biol.* **25**, 2577-2583 (2015).
137. J. Metzger, M. Karwath, R. Tonda, S. Beltran, L. Águeda, M. Gut *et al.*, Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genom.* **16**, 764 (2015).