ZOOLOGICAL RESEARCH

Potential dual expansion of domesticated donkeys revealed by worldwide analysis on mitochondrial sequences

Xi-Yao Ma[1,#], Tiao Ning[2,3,#], Adeniyi C. Adeola[4,5,#], Jie Li[1,#], Ali Esmailizadeh[4,6], Jacqueline K. Lichoti[7], Bernard R. Agwanda[8], Jainagul Isakova[9], Almaz A. Aldashev[9], Shi-Fang Wu[4], He-Qun Liu[4], Najmudinov Tojiddin Abdulloevich[10], Manilova Elena Afanasevna[10], Khudoidodov Behruz Ibrohimovich[10], Rahamon Akinyele Moshood Adedokun[11], Sunday Charles Olaogun[11], Oscar J. Sanke[12], Godwin F. Mangbon[13], Xi Chen[14,15], Wei-Kang Yang[14,15], Zhe Wang[16,17], Min-Sheng Peng[4,5,18], Sheila C. Ommeh[19,*], Yan Li[1,*], Ya-Ping Zhang[1,4,18,20,*]

[1] *State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan*, *School of Life Sciences*, *Yunnan University*, *Kunming*, *Yunnan* 650091, *China*

[2] *Agriculture College*, *Kunming University*, *Kunming*, *Yunnan* 650214, *China*

[3] *Engineering Research Center for Urban Modern Agriculture of Higher Education in Yunnan Province*, *Kunming University*, *Kunming*, *Yunnan* 650214, *China*

[4] *State Key Laboratory of Genetic Resources and Evolution*, *Yunnan Laboratory of Molecular Biology of Domestic Animals*, *Germplasm Bank of Wild Species*, *Kunming Institute of Zoology*, *Chinese Academy of Sciences*, *Kunming*, *Yunnan* 650223, *China*

[5] *Sino-Africa Joint Research Center*, *Chinese Academy of Sciences*, *Kunming*, *Yunnan* 650223, *China*

[6] *Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman* PB 76169-133, *Iran*

[7] *State Department of Livestock*, *Ministry of Agriculture Livestock and Fisheries*, *Nairobi* 00100, *Kenya*

[8] *Department of Zoology*, *National Museums of Kenya*, *Nairobi* 00100, *Kenya*

[9] *Institute of Molecular Biology and Medicine*, *Bishkek* 720040, *Kyrgyzstan*

[10] *E.N. Pavlovsky Institute of Zoology and Parasitology*, *Academy of Sciences of Republic of Tajikistan*, *Dushanbe* 734025, *Tajikistan*

[11] *Department of Veterinary Medicine*, *University of Ibadan*, *Ibadan* 200284, *Nigeria*

[12] *Taraba State Ministry of Agriculture and Natural Resources*, *Jalingo* 660221, *Nigeria*

[13] *Division of Veterinary Office*, *Serti* 663101, *Nigeria*

[14] *Research Center for Ecology and Environment of Central Asia*, *Chinese Academy of Sciences*, *Urumqi*, *Xinjiang* 830011, *China*

[15] *Key Laboratory of Biogeography and Bioresource in Arid Land*, *Xinjiang Institute of Ecology and Geography*, *Chinese Academy of Sciences*, *Urumqi*, *Xinjiang* 830011, *China*

[16] *Department of Physiology*, *Hebei Medical University*, *Shijiazhuang*, *Hebei* 050017, *China*

[17] *Hebei Key Laboratory of Animal Science*, *Hebei Medical University*, *Shijiazhuang*, *Hebei* 050017, *China*

Received: 03 July 2019; Accepted: 29 October 2019; Online: 29 October 2020

Foundation items: This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA2004010302), Second Tibetan Plateau Scientific Expedition and Research (STEP) Program (2019QZKK0501), National Natural Science Foundation of China (31860305), Sino-Africa Joint Research Center, Chinese Academy of Sciences (SAJC201611), and Research Center for Ecology and Environment of Central Asia, Chinese Academy of Sciences. Y.L. was supported by the Young Academic and Technical Leader Raising Foundation of Yunnan Province. Samples used in this study were provided by the Animal Branch of the Germplasm Bank of Wild Species, Chinese Academy of Sciences (Large Research Infrastructure Funding)

[#]Authors contributed equally to this work

*Corresponding authors, E-mail: sommeh@jkuat.ac.ke; liyan0910@ynu.edu.cn; zhangyp@mail.kiz.ac.cn

DOI: 10.24272/j.issn.2095-8137.2020.007

[18] *Kunming College of Life Science*, *University of Chinese Academy of Sciences*, *Kunming*, *Yunnan* 650204, *China*
[19] *Animal Biotechnology Group*, *Institute of Biotechnology Research*, *Jomo Kenyatta University of Agriculture and Technology*, *Nairobi* 00200, *Kenya*
[20] *Center for Excellence in Animal Evolution and Genetics*, *Chinese Academy of Sciences*, *Kunming*, *Yunnan* 650223, *China*

ABSTRACT

Molecular studies on donkey mitochondrial sequences have clearly defined two distinct maternal lineages involved in domestication. However, domestication histories of these two lineages remain enigmatic. We therefore compared several population characteristics between these two lineages based on global sampling, which included 171 sequences obtained in this study (including Middle Asian, East Asian, and African samples) plus 536 published sequences (including European, Asian, and African samples). The two lineages were clearly separated from each other based on whole mitochondrial genomes and partial non-coding displacement loop (D-loop) sequences, respectively. The Clade I lineage experienced an increase in population size more than 8 000 years ago and shows a complex haplotype network. In contrast, the population size of the Clade II lineage has remained relatively constant, with a simpler haplotype network. Although the distribution of the two lineages was almost equal across the Eurasian mainland, they still presented discernible but complex geographic bias in most parts of Africa, which are known as their domestication sites. Donkeys from sub-Saharan Africa tended to descend from the Clade I lineage, whereas the Clade II lineage was dominant along the East and North coasts of Africa. Furthermore, the migration routes inferred from diversity decay suggested different expansion across China between the two lineages. Altogether, these differences indicated non-simultaneous domestication of the two lineages, which was possibly influenced by the response of pastoralists to the desertification of the Sahara and by the social expansion and trade of ancient humans in Northeast Africa, respectively.

Keywords: Donkey lineage; Domestication history; Population; Expansion

INTRODUCTION

Unlike other species with a similar historical function (e.g., horses), domestic donkeys are underrepresented in the scientific literature (Blench, 2000). Importantly, donkeys remain an essential means of transport in modern society for people living in mountain areas, deserts, and poorer regions of the world (Smith & Pearson, 2005; Starkey, 2000). Both molecular data and archaeological evidence strongly support an African origin for the domestic donkey (*Equus asinus*) (Beja-Pereira et al., 2004; Kimura et al., 2011; Rossel et al., 2008). This domestication, together with that of horses, is believed to have contributed to mobile pastoralism as well as the establishment of ancient overland trade routes and spread of ancient civilizations (Rosenbom et al., 2015). Previous studies on mitochondrial DNA (mtDNA) have revealed that modern donkeys can be clearly separated into two distinct clades: one clustered directly with the Nubian wild ass (*E. africanus africanus*), hereafter termed the Clade I lineage; another derived from a probably extinct wild ass close to the Somali wild ass (*E. africanus somaliensis*), hereafter termed the Clade II lineage (Beja-Pereira et al., 2004; Chen et al., 2006; Kimura et al., 2011). Based on archeological evidence, the "Egyptian hypothesis" proposes that villagers in the Nile Valley domesticated the resident Nubian wild ass approximately 5 000–6 000 years ago (Clutton-Brock, 1992; Epstein, 1971); in contrast, the "pastoralist hypothesis", supported by ethnographic, climatic, and linguistic data, states that pastoralists in northeastern Africa 6 500–7 000 years ago domesticated donkeys in response to the increasing aridity of the Sahara (Marshall, 2007). Additionally, the very high genetic diversity found in the Arabian Peninsula, as based on nuclear microsatellites, indicates an alternative domestication center or melting pot (ancient trade areas or routes) (Rosenbom et al., 2015). Ancient DNA also illustrates the extensive distribution of the Nubian and Somali wild ass in Northeast Africa (Kimura et al., 2011), alluding to potential geographic overlaps between the Nubian wild ass and ancestors of the Clade II donkeys. As such, whether these two distinct lineages were domesticated simultaneously or not remains controversial, even though they probably originated from two mitochondrially distinct wild asses (Jordana et al., 2016; Xia et al., 2019).

If the two lineages were domesticated simultaneously, it is reasonable to assume the existence of similar genetic dynamics corresponding to their rapid expansion with innate mobile characteristics. Therefore, in the present paper, we investigated the genetic characteristics of 84 mtDNA genomes and 707 D-loop sequences to compare the dynamic signatures between these two clades that correspond to migration history. Our results will help to further understand mtDNA diversity in African donkeys and also provide novel insights into their domestication histories.

MATERIALS AND METHODS

Sample information

Blood samples were collected and sequenced from animals living in villages or donkey breeding farms all over the world.

This study was approved by the Ethics and Experimental Animal Committee of the Kunming Institute of Zoology, Chinese Academy of Sciences, China. We selected samples without any discrimination toward age, sex, or location. More detailed information, including geographical location and the US National Center for Biotechnology Information (NCBI) GenBank accession Nos., can be found in Supplementary Table S1. Specifically, 5 mL of blood was drawn from healthy animals by a veterinarian (sampling picture in Supplementary Figure S1) with approval from the above-mentioned Ethics and Experimental Animal Committee. The blood samples were preserved in 95% alcohol for further genomic analysis. Whole DNA was isolated from blood using the standard phenol-chloroform method (Ma, 2018) and subsequently quantified with NanoDrop 2000 (Thermo Fisher Scientific, USA).

We obtained a 371 bp length polymerase chain reaction (PCR) product within the D-loop region for 94 donkeys (Tadzhikistan: n=19; Kyrgyzstan: n=5; Kenya: n=16; Nigeria: n=22; Iran: n =30; China: n =2) using primers designed in previous research (Kefena et al., 2014) (5'-DONK-F: CCC AAGGACTATCAAGGAAG-3';DONK-R:5'-GGAATGGCCCTGA AGAAAG-3'). We strictly followed the reaction system as described in Kefena et al. (2014). Specifically, PCR was carried out in a 10 μL reaction volume containing 1 μL of DNA (50 μg/μL), 1 μL 10×buffer (10×Takara Taq^{TM} Buffer Mg^{2+} plus), 0.2 mmol/L of each dNTP, 1 μmol/L of each primer, and 0.2 U/tube Taq (Taq DNA Polymerase, Takara) under the following conditions: initial denaturation at 94 °C for 15 min followed by 45 cycles of denaturation each at 94 °C for 1 min, hybridization at 56 °C for 1 min, extension at 72 °C for 1 min, and final extension at 72 °C for 20 min. The PCR products were gel-purified as described in Ma (2018) and then sequenced by single-strand PCR using the ABI PRISM^{TM} Dye Terminator Cycle Sequencing kit following protocols recommended by the manufacturer. Sequencing was implemented using ABI-PRISM 3730 standard conditions (Applied Biosystems, 2009). For each PCR product, sequences were determined in both forward and reverse directions for all nucleotide positions to avoid possible artificial variations.

Similarly, we obtained the mtDNA genomes of 56 Chinese donkeys and two wild asses using ABI-PRISM 3730 with primers designed in this study (see Supplementary Table S2). Other mtDNA genomes of 17 donkeys (Tadzhikistan: n=6; Kyrgyzstan: n =5; Kenya: n =4; Nigeria: n =2) and two Tadzhikistan horses were yielded through Ion Torrent technology, as per previous study (Chen et al., 2016). Briefly, we amplified PCR fragments, constructed libraries, performed sequencing using next-generation technology, and *de novo* assembled the reads. For quality control, we followed the caveats mentioned in previous mtDNA genome study of domestic animals (Shi et al., 2014). Variants that differentiated from the GenBank reference sequence under accession No. NC_001788 (Xu et al., 1996) were scored. We then manually checked the bam file exported by Torrent Suite 5.0.2 to confirm the scored variants using Integrative Genomics Viewer (Thorvaldsdóttir et al., 2013).

Collectively, we obtained 94 sequences with a 371 bp length restricted to the D-loop region and 77 almost complete mitochondrial genomes. Sequences generated in this study were deposited in the NCBI archive (GenBank accession Nos.: MK650231–MK650286 for 58 mtDNA genomes obtained by 3730 sequencer; MK619357–MK619411 for 55 D-loop haplotypes of *Equus* yield in this study; SAMN11432793–SAMN11432809 and SRX5702272–SRX5702273 for another 19 mtDNA genomes sequenced through Ion Torrent technology. For specified topology and wider coverage in geographic structure analysis, we combined an additional seven mitochondrial genomes and 536 D-loop sequences downloaded from NCBI website (http://www.ncbi.nlm.nih.gov/), with GenBank accession Nos. listed in Supplementary Table S1. Detailed feature information of all 707 sequences are included in Supplementary Table S1.

Phylogenetic construction and estimation of genetic parameters

We reconstructed the phylogenetic tree of equid mtDNA sequences using neighbor-joining and maximum likelihood analyses with MEGA 6 software (Tamura et al., 2013). For mtDNA genomes, 1 000 bootstrap replicates were conducted with the Kimura 2-parameter model, assuming Gamma distributed mutation rates for neighbor-joining (NJ) analysis and Gamma distributed with invariant sites (G+I) distributed mutation rates for maximum likelihood (ML) analysis. Additionally, we constructed a Bayesian tree using Bayesian evolutionary analysis by sampling trees (BEAST) v1.6.1 (Drummond & Rambaut, 2007) for phylogenetic confirmation with the Gamma distributed HKY substitution model under a strict clock rate, assuming constant size coalescence for tree prior. The non-coding region from 16 129 nt to 16 360 nt was excluded due to the short tandem repeats in both horse and donkey (repeat unit: GTGCACCT in horse; CACACCCACAC ACCCATGCGCGCA in donkey).

We truncated the 77 mitochondrial genomes into the 371 bp length of the D-loop region to ensure combined analysis with the 94 D-loop sequences, which yielded 55 haplotypes. The downloaded 536 D-loop sequences were from different research, which covered different fragments within the D-loop region. Therefore, it was inevitable that a very short overlapping fragment (235 bp) would be yielded as we retrieved as many samples as possible. For genetic diversity analysis, because the downloaded 536 D-loop sequences were haplotype data, we selected only haplotypes within each area from our sequenced samples for parallel comparison. A reduced median-joining network was generated using NETWORK (Bandelt et al., 1999). Average pairwise differences were estimated using Arlequin v3.5 (Excoffier & Lischer, 2010). The nucleotide diversity (π), mismatch distribution, Tajima's D test, and Fu's Fs test were calculated using DnaSP v6 (Rozas et al., 2017).

We assessed the population dynamics for the mtDNA genomes of the two lineages using Bayesian Skyline Plot

(BSP) implemented in BEAST (Drummond & Rambaut, 2007). We estimated the evolution rate using the previously estimated divergence time of 2 million years between the horse and donkey from paleontological data (Lindsay et al., 1980). Migration time to regions distant from the domestication center was inferred by estimating the evolution time of haplotypes only found in the region to the major haplotype.

A geographic distribution map was plotted using the "rworldmap" package (South, 2011) implemented in R (R-Core-Team, 2019).

RESULTS AND DISCUSSION

Two clearly separated domesticated donkey clades

A clearly defined phylogenetic relationship can allow comparison of genetic characters between two lineages. Thus, we constructed a highly supported phylogenetic tree based on the 77 mtDNA genome sequences from the current study (including 73 donkeys from China, Tadzhikistan, Kyrgyzstan, Kenya, and Nigeria, two Asiatic wild asses, and two Tadzhikistan indigenous horses) combined with seven mtDNA genomes downloaded from NCBI (including two donkeys, two Somali wild asses, and two Asian wild asses) (Supplementary Table S1). Using the horse as an outgroup, the Somali wild ass formed a sister clade with all donkeys, a topology highly differentiated from that based on D-loop sequences, which was highly supported by the neighbor-joining, maximum likelihood, and Bayesian methods analyzed here (Figure 1). This topology was also consistent with previous study that concentrated on the major mitochondrial coding regions and placed the Somali wild ass as a sister clade outside all domestic donkeys with high confidence, although they did not label the lineage information (Sun et al., 2016). Here, we obtained consistent topologies for the major clades according to both whole mitochondrial genomes and coding regions only (Figure 1; Supplementary Figure S2). With a view that coding regions experience restricted selective sweep, whereas selective force is potentially relaxed on the D-loop regions (Endicott et al., 2009), it is reasonable to assume that the phylogenetic relationship revealed by mitochondrial genome data would be much closer to reality.

The two donkey lineages were clearly separated from each other, with high support (Figure 1), further confirming the existence of two lineages. Considering the lack of geographic representation from mtDNA genomes (58 of 75 domestic donkeys were from China), we reconstructed the phylogenetic tree using both neighbor-joining and maximum likelihood methods with D-loop sequences from the 171 sequences obtained in this study and 536 sequences publicly available on NCBI (see sample information in Supplementary Table S1 and phylogenetic tree in Supplementary Figure S3). Similarly, the donkeys were divided into two groups. Although the clades showed lower support, which is common in analysis of short segments like the D-loop region, all sequences with published lineage information clustered into their corresponding clades. In addition, the 73 donkey D-loop sequences obtained in this study showed consistent lineage sorting, as observed in the whole mtDNA genome phylogeny, further supporting the D-loop region phylogenetic relationship. Our phylogenetic tree based on D-loop sequences also revealed a sister relationship between the Somali wild ass and all domesticated donkeys (Supplementary Figure S3), consistent with the topology revealed by the mitochondrial genome, but inconsistent with previous studies showing different topology using D-loop sequences (Beja-Pereira et al., 2004; Han et al., 2014). The number of haplotypes used for phylogenetic construction may account for this discrepancy, as more haplotypes represent greater genetic diversity, and subsequently more statistical credibility. To test this, we randomly selected donkey haplotypes of equal sample size, as previously described in Han et al. (2014), and constructed a phylogenetic tree with the same parameters in this study. We obtained similar phylogenetic topology (Supplementary Figure S4), showing a closer relationship shared by the Somali wild ass and domesticated Clade II lineage. Therefore, a large sample size for phylogenetic analysis can compensate for the limited information obtained from short segments.

Different demographic dynamics of two lineages

As the donkey samples were credibly sorted, we subsequently compared several population characteristics between the two lineages. First, we scanned sequence variation within the mitochondrial genome for each lineage. The Clade I lineage showed a total of 288 varied sites: 258 in the coding region and 30 in the non-coding regions (D-loop). The Clade II lineage showed a total of 188 varied sites: 168 in the coding region and 20 in the non-coding regions (D-loop). Figure 2A illustrates the distribution of nucleotide diversity (π) within each lineage along the genome based on assessment of 200 nt windows (step size=100 nt) centered at the midpoint. Although the highest diversity was observed in the D-loop region presenting the short tandem repeats of CACACCCACACACCCATGCGCGCA (donkey reference NC_001788: from 16 173 nt to 16 340 nt) in both lineages, the Clade I lineage possessed significantly more segregation sites (average segregation sites per window=1.300) in the coding region than the Clade II lineage (average segregation sites per window=1.091) (Wilcox test: P=0.016), whereas no significant differences were found in segregation counts in the non-coding region in the two lineages (average segregation sites per window in Clade I and Clade II were 3.057 and 2.303, respectively; Wilcox test: P=0.628). This discrepancy may be due to the saturation effect in the D-loop region, demographics, selective forces, and/or other factors suggestive of potential independent domestic histories between the two lineages. To further address this issue, we used mtDNA genomes to reconstruct ancestral population dynamics with BSP for each lineage (Figure 2B). A similar demographic history would be expected under the assumption of simultaneous domestication. However, we observed a constant effective population size in the Clade II lineage during most history, whereas the Clade I lineage experienced an apparent population increase approximately 8 000 years

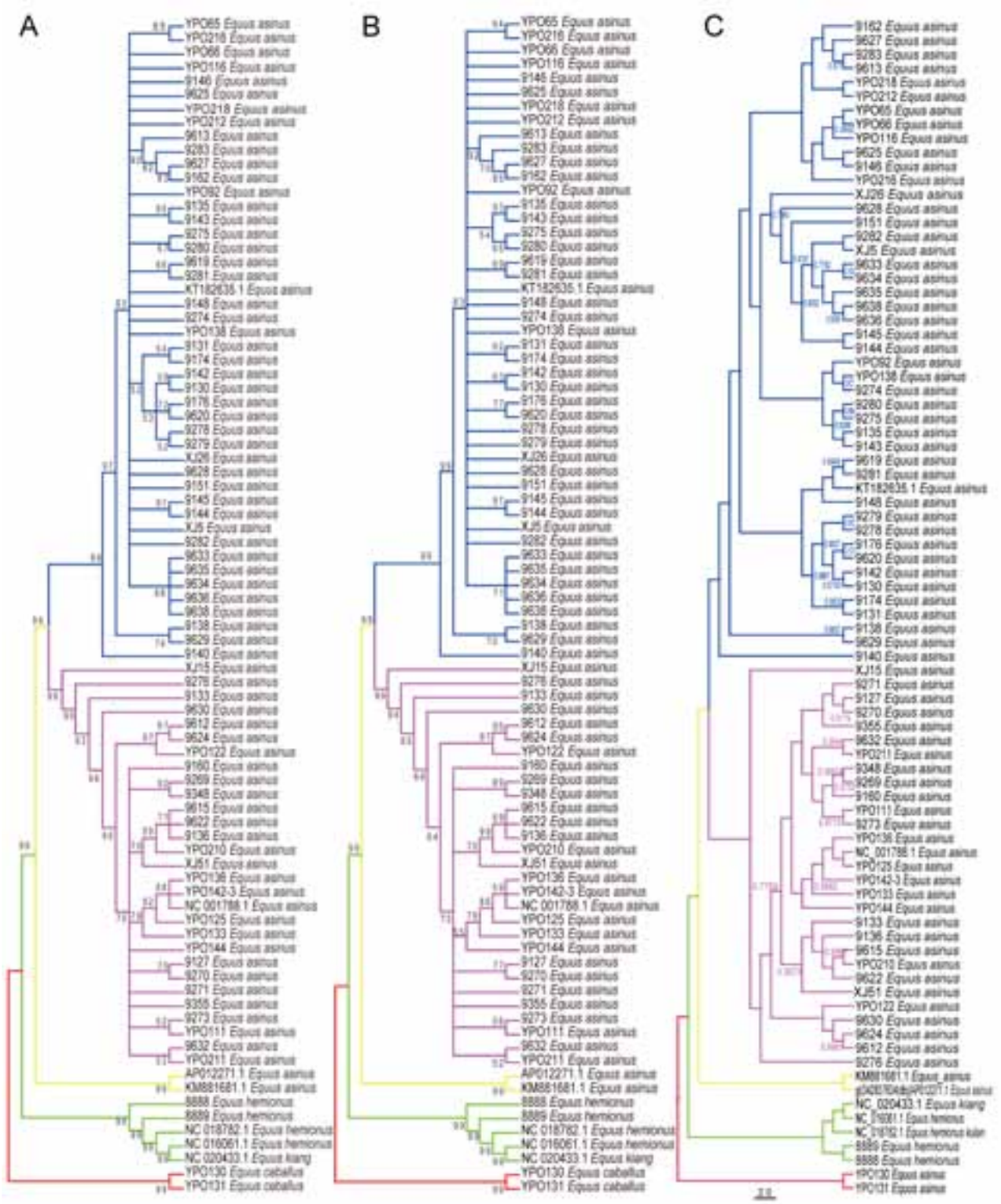

Figure 1 mtDNA genome phylogeny of *Equus*

A: Neighbor-joining tree; B: Maximum likelihood tree; C: Bayesian tree. Colored branches denote different categories. Red: Horse; Green: Asiatic wild ass; Yellow: Somali wild ass; Blue: Clade I lineage; Magenta: Clade II lineage.

ago, coinciding with the archeological date 7 770±95 before present (BP) (Marshall, 2007). Noticeably, both lineages exhibited a marked decrease in recent years, probably due to the occurrence of the industrial revolution. To further confirm this discrepancy in population dynamics, we assessed the demographic history of the two lineages based on D-loop sequences, with nucleotide mismatch distribution (Rogers & Harpending, 1992), Fu's *F* s test, and Tajima's *D* test. The nucleotide mismatch curve showed a single peak in the Clade I lineage and double peaks in the Clade II lineage (Figure 2C). Moreover, results showed that both Fu's *F*s test and Tajima's *D* test significantly deviated from neutrality in the Clade I lineage (Tajima's *D*=–2.310, *P*<0.01; Fu's *F*s=–33.909) but not in the Clade II lineage (Tajima's *D* = –0.993, *P* >0.10; Fu's *F*s= –17.464). These results were in accordance with the expansion in the Clade I lineage and constant size in the Clade II lineage, confirming the demographic dynamics revealed by mitochondrial genomes.

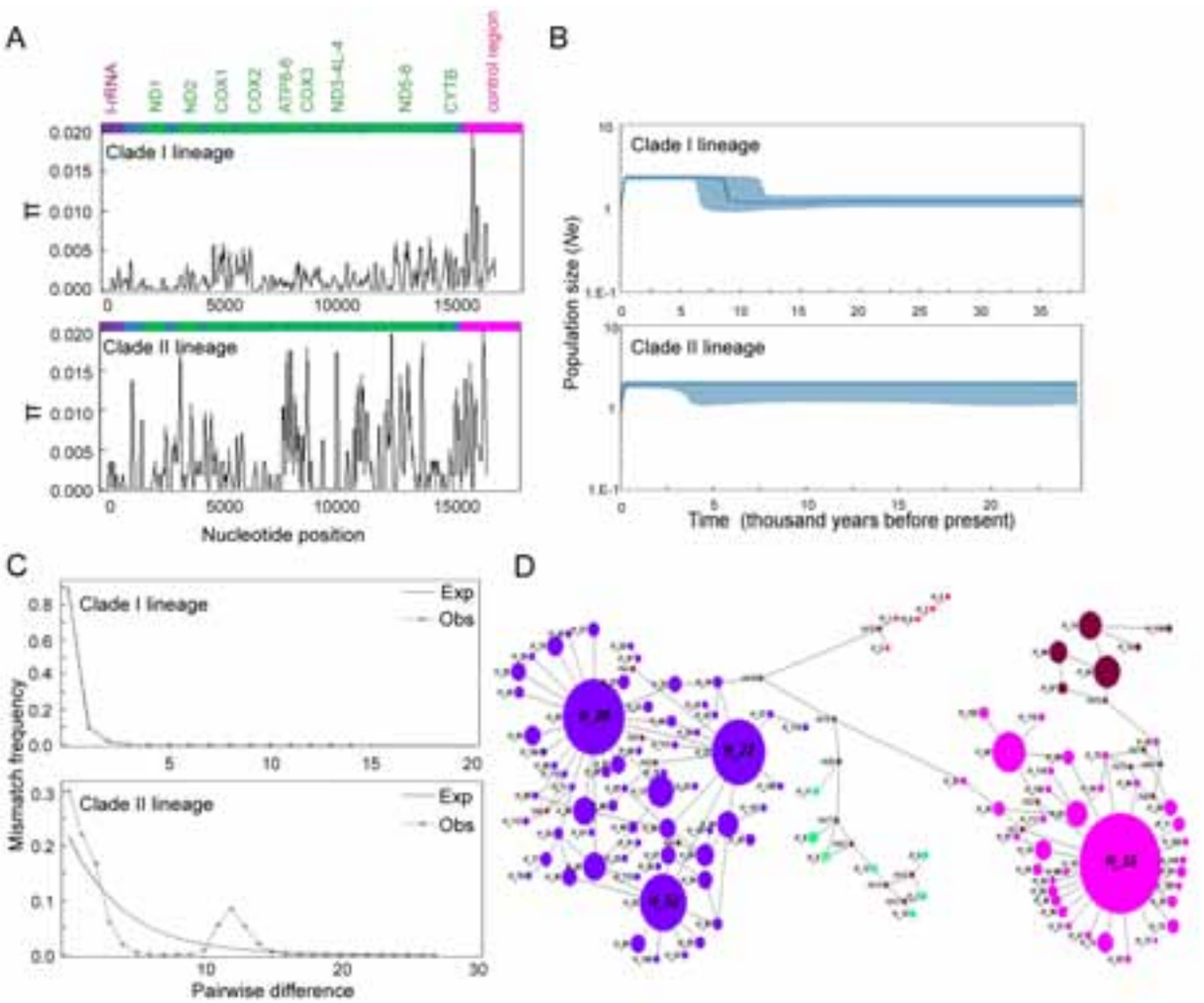

Figure 2 Population characteristics between Clade I and Clade II lineages

A: Sequence variation of mtDNA genome within each lineage. Nucleotide diversity (π) using 200 nt windows (step size=100 nt) centered at midpoint. Schematic (linearized) genetic map of mitochondrial genome is presented. B: Demographic histories for each lineage estimated using Bayesian skyline plot from mitochondrial genomes. Shadow means 95% highest posterior density intervals of effective number of females. C: Mismatch distribution for each mtDNA lineage. D: Median-joining network based on D-loop region. Red diamonds are haplotypes of horse outgroup. Green circles are haplotypes of Asiatic wild ass. Dark red circles are haplotypes of Somali wild ass. Indigo circles are haplotypes of Clade I lineage. Magenta circles are haplotypes of Clade II lineage. Circle size is proportional to frequencies of haplotypes.

Discrepancies of haplotype structure

In addition to demographic estimation, we also investigated several other genetic characters to assess whether they showed similar evolution patterns between the two lineages. We first constructed a reduced median network based on D-loop sequences (Figure 2D). Similar to previous study (Beja-

Pereira et al., 2004; Kimura et al., 2011), the Clade II haplotypes were mainly derived from a single major haplotype (H16), with a simple star-like shape, whereas the genetic architecture of the Clade I haplotypes was more complicated, with more universally occurring haplotypes (e.g., H20, H52, H22). This was consistent with the much higher genetic distance and nucleotide diversity within the Clade I lineage (average pairwise distance=1.932 7, *SD*=1.100 7; *Pi*=0.008 4, *SD*=0.000 3) than within the Clade II lineage (average pairwise distance=1.074 4, *SD*=0.711 2; *Pi*=0.004 6, *SD*= 0.000 5), implying that the Clade I lineage involved many more individuals at the beginning of domestication compared with the Clade II lineage (Table 1).

Table 1 Genetic distance and nucleotide diversity of two clades

	Pairwise distance	*SD* (distance)	*Pi*	*SD* (*Pi*)
Clade I lineage	1.932 7	1.100 7	0.008 4	0.000 3
Clade II lineage	1.074 4	0.711 2	0.004 6	0.000 5

Biased distribution of two lineages in Africa

Independent migration events may result in a geographical structure. Therefore, we estimated the migration routes of the two lineages by referring to their proportion across the world. D-loop sequences sampled from the Balkans and microsatellites sampled from northeast Africa, the Near East, and the Arabian Peninsula indicated the absence of a geographical structure (Pérez-Pardal et al., 2014; Rosenbom et al., 2015). Indeed, samples collected from Middle Asia (Pakistan, Kazakhstan) and major areas of China demonstrated an almost equal proportion of the two lineages (Figure 3A). The Arabian Peninsula is assumed to be the melting pot from where domestic donkeys migrated to the world (Rosenbom et al., 2015). As this area showed a nearly equal proportion of the two lineages, it is reasonable that a similar pattern is commonly observed across the Eurasian mainland (Beja-Pereira et al., 2004). Nevertheless, all 20 samples collected from Nigeria belonged to the Clade I lineage. When we focused on the distribution in Africa, an apparent spatial structure was detected: donkeys from sub-Saharan Africa (e.g., Ghana, Guinea, Benin, Mali, Senegal, South Africa, and Burkina Faso) tended to be descended from the Clade I lineage, whereas the Clade II lineage was dominant along the East (e.g., Eritrea, Somalia, Swaziland, and Zambia) and North coasts (e.g., Libya, Tunisia, and Morocco) (Figure 3A).

The time of migration to areas distant from the domestication center can be inferred by dating the time to the major haplotype for the derived haplotypes only found in that area. In this way, we inferred that the Clade I lineage migrated into sub-Sahara 4 874.28±1 817.92 years ago, around the commencement of desertification in the Sahara (5 000 to 7 000 BP) (Marshall, 2000). Due to the excellent tolerance of donkeys for deserts, it is reasonable to assume that donkeys, rather than horses, were the major means of transport across the Sahara during the initial period of desertification. Given the biased Clade I lineage distribution in the sub-Sahara, this migration time provides possible evidence for the "pastoralist hypothesis": i. e., pastoralists in northeastern Africa domesticated Clade I lineage donkeys in response to the increasing aridity in the Sahara.

Donkeys are thought to be have been brought into Europe by the second millennium BC, possibly through viticulture introduction, as the donkey is associated with the Syrian god of wine, Dionysus (Meutchieye et al., 2017). Interestingly, the majority of sequences from the Iberian Peninsula belonged to the Clade II lineage, much different from those collected in other parts of Europe (Figure 3A). Additionally, the estimated time of arrival in the Iberian Peninsula was 5 336.8±1 652.56 years ago, much earlier than the known history of ~4 000 years (Cardoso et al., 2013). Considering the dominance of the Clade II lineage in Morocco, it could be assumed that donkeys migrated into the Iberian Peninsula directly through the Strait of Gibraltar. Furthermore, the unequal distribution of Clade I and Clade II in Europe may account for the pronounced footprint in American following the complex process of colonization, where an apparent geographic structure has been observed (Jordana et al., 2016; Xia et al., 2019). The Clade II lineage also migrated dominantly along East Africa, implying a potential association with ancient social expansion, such as the Bantu expansion (Hiernaux, 1968; Herrera & Garcia-Bertrand, 2018), and ancient trade routes of empires such as the Punt and Aksum (Andrews, 2017; Mark, 2011; Mukhtār, 1981).

Different routes of two lineages during expansion to China

As populations expand from centers of origin, genetic diversity is lost as a consequence of the limited numbers of individuals involved in the expansionist movement. Although no apparent geographic structure was detected across China, where the two lineages show an almost equal proportion (Figure 3A), discrepancies in diversity decay may exist between the two lineages if non-simultaneous expansion into China occurred, thereby suggesting possible private migration routes. The Xinjiang Province may be a transportation center as it demonstrated almost the highest nucleotide diversity for both lineages (Figure 3B, C), consistent with previous studies on genetic diversity among various Chinese breeds (Ge et al., 2007) and written records (Xie, 1987). The genetic diversity of Clade I remained relatively high in the Qinghai and Henan provinces, but declined gradually towards north and south, respectively (Figure 3B), consistent with a migration route from Xinjiang to the Guanzhong Plain through the Ningxia and Gansu provinces, and finally other areas of China inferred previously (Ge et al., 2007). The genetic diversity of Clade II remained relatively high in the Inner Mongolia and Yunnan provinces, but declined substantially in the middle region of China (Figure 3C), consistent with previously inferred migration routes from Xinjiang to Inner Mongolia (north forward) and to Yunnan (south forward) (Ge et al., 2007). These potential migration routes partially match written records, which suggest that the "Taihang donkey" was introduced into Hebei Province through Inner Mongolia; the "Yunnan donkey" was introduced from Xinjiang to Yunnan, as

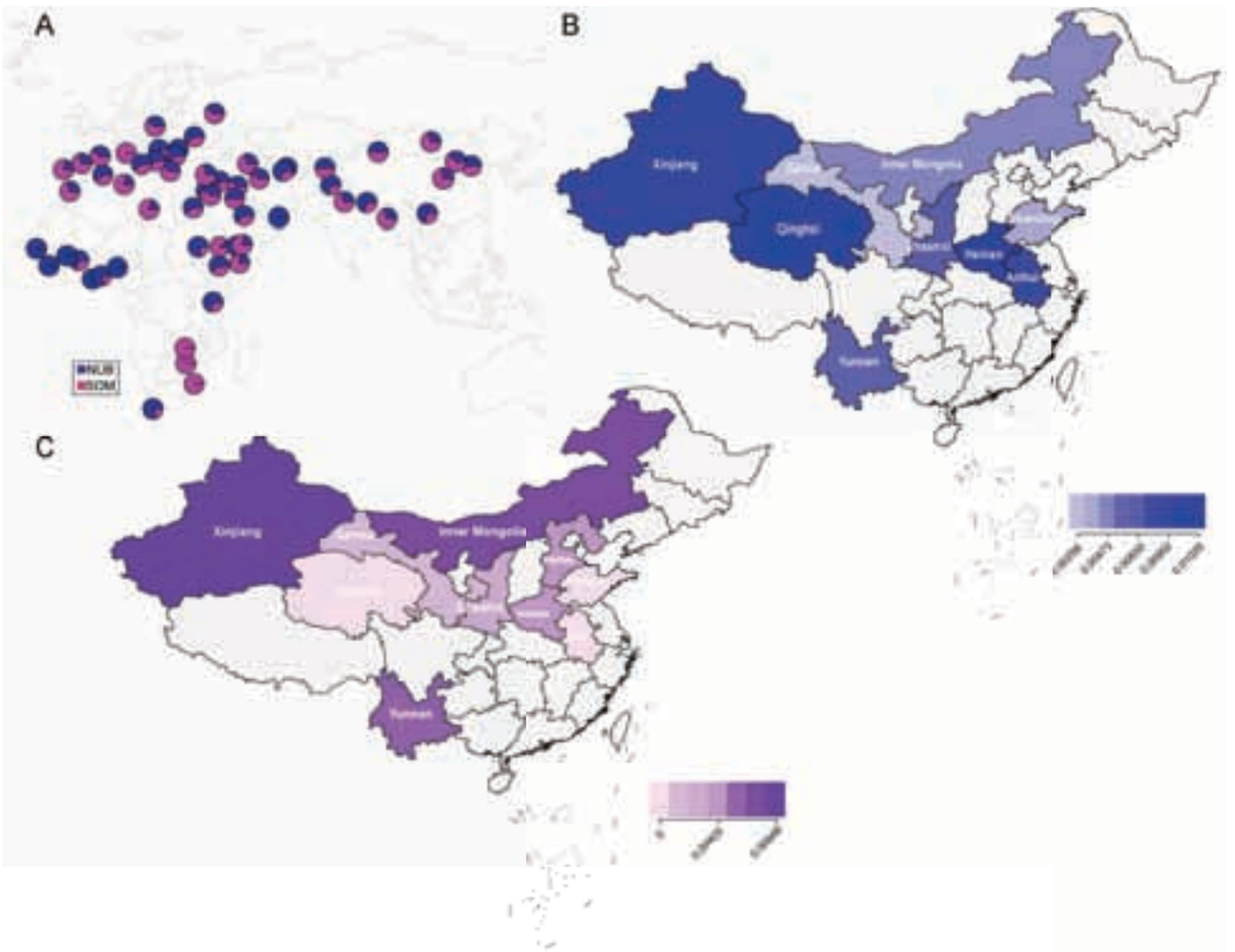

Figure 3 Worldwide geographic distribution of two lineages plotted using "rworldmap" package in R (A) and genetic diversity of Clade I (B) and Clade II (C) lineages across China

A: Blue: Clade I lineage; Magenta: Clade II lineage. B: Dark to light indigo indicates diversity decay. Gray regions indicate no samples available. C: Dark to light magenta indicates diversity decay. Gray regions indicate no samples available.

it shares many morphological traits with the Xinjiang donkey; and many other breeds may have been formed through migration routes from Xinjiang to Gansu and then Guanzhong Plain, Henan, and other areas of China (Xie, 1987). Therefore, it is likely that the two lineages expanded into China through different migration events, although from the same transportation center (Xinjiang).

CONCLUSIONS

Overall, our results on the demographic dynamics, haplotype structure, diversity decay pattern, and distribution bias in Africa are in accordance with non-simultaneous domestication events and an independent migration history. Our findings revealed that domestication of the donkey may have been driven by the response of pastoralists to the desertification of the Sahara and by social expansion and trade of ancient humans. Future research on population genomes will be needed to increase our understanding of donkey domestication history.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

Y. P. Z., Y. L., and S. C. O. designed the study. X. Y. M., T. N., and A. C. A. sequenced the samples. X.Y.M. and J.L. performed data analyses. A.E., J. D.L., B.R.A., J.I., A.A.A., S.F.W., H.Q.L., N.T.A., M.E.A., K.B.I., R.A.M.A., S. C.O., O.J.S., M.G.F., X.C., and W.K.Y. helped with sample collection. Y.L., A.C.A., Z.W., and M.S.P. wrote the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We thank Xing Chen, Ya-Jiang Wu, Miao Wu, Mu-Yang Wang, Yao-Ming Li, and Feng Xu for technical assistance.

REFERENCES

Andrews E. 2017. 7 Influential African Empires. From ancient Sudan to medieval Zimbabwe, get the facts on seven African kingdoms that made their mark on history. A&E Television Networks, LLC. URL: https://www.history.com/.amp/news/7-influential-african-empires.

Applied Biosystems. 2009. DNA Sequencing by Capillary Electrophoresis, Applied BiosystemsChemistry Guide. Thermo Fisher Scientific.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**(1): 37−48.

Beja-Pereira A, England PR, Ferrand N, Jordan S, Bakhiet AO, Abdalla MA, Mashkour M, Jordana J, Taberlet P, Luikart G. 2004. African origins of the domestic donkey. *Science*, **304**(5678): 1781.

Blench RM. 2000. A History of Donkeys, Wild Asses and Mules in Africa. London, UK: University College London Press.

Cardoso, J L, Vilstrup JT, Eisenmann V, Orlando L. 2013. First evidence of *Equus asinus* L. in the Chalcolithic disputes the Phoenicians as the first to introduce donkeys into the Iberian Peninsula. *Journal of Archaeological Science*, **40**(12): 4483−4490.

Chen SY, Zhou F, Xiao H, Sha T, Wu SF, Zhang YP. 2006. Mitochondrial DNA diversity and population structure of four Chinese donkey breeds. *Animal. Genetics*, **37**(4): 427−429.

Chen X, Ni G, He K, Ding ZL, Li GM, Adeola AC, Murphy RW, Wang WZ, Zhang YP. 2016. An improved de novo pipeline for enrichment of high diversity mitochondrial genomes from Amphibia to high-throughput sequencing. *BioRxiv*: 080689.

Clutton-Brock J. 1992. Horse Power: A History of the Horse and the Donkey in Human Societies. Cambridge, USA: Harvard University Press.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary. Biology*, **7**: 214.

Endicott P, Ho SYW, Metspalu M, Stringer C. 2009. Evaluating the mitochondrial timescale of human evolution. *Trends in Ecology and Evolution*, **24**(9): 515−521.

Epstein H. 1971. The Origin of the Domestic Animals of Africa. New York, USA: African Publishing Corporation (APC).

Excoffier L, Lischer HEL. 2010. Arlequin suite ver3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**(3): 564−567.

Ge QL, Lei CZ, Jiang YQ, Chen H, Zhang W, Dang RH, Zhen HL, Zhang AL, Li TP. 2007. Genetic diversity and origins of Chinese donkey revealed by mitochondrial D-loop sequencies. *Acta Veterinaria et Zootechnica Sinica*, **38**(7): 641−645.

Han L, Zhu S, Ning C, Cai D, Wang K, Chen Q, Hu S, Yang J, Shao J, Zhu H, Zhou H. 2014. Ancient DNA provides new insight into the maternal lineages and domestication of Chinese donkeys. *BMC Evolutionary Biology*, **14**: 246.

Herrera RJ, Garcia-Bertrand R. 2018. The Bantu Expansion. Ancestral DNA, human origins, and migrations. Academic Press, 395–432.

Hiernaux J. 1968. Bantu expansion: the evidence from physical anthropology confronted with linguistic and archaeological evidence. *The Journal of African History*, **9**(4): 505−515.

Jordana J, Ferrando A, Miró J, Goyache F, Loarca A, Martinez López OR, Canelón JL, Stemmer A, Aguirre L, Lara MAC, Álvarez LA, Llambí S, Gómez N, Gama LT, Nóvoa MF, Martínez RD, Pérez E, Sierra A, Contreras MA, Guastella AM, Marletta D, Arsenos G, Curik I, Landi V, Martinez A, Delgado JV. 2016. Genetic relationships among American donkey populations: insights into the process of colonization. *Journal of Animal Breeding and Genetics*, **133**(2): 155−164.

Kefena E, Dessie T, Tegegne A, Beja-Pereira A, Yusuf Kurtu M, Rosenbom S, Han JL. 2014. Genetic diversity and matrilineal genetic signature of native Ethiopian donkeys (*Equus asinus*) inferred from mitochondrial DNA sequence polymorphism. *Livestock Science*, **167**: 73−79.

Kimura B, Marshall FB, Chen S, Rosenbom S, Moehlman PD, Tuross N, Sabin RC, Peters J, Barich B, Yohannes H, Kebede F, Teclai R, Beja-Pereira A, Mulligan CJ. 2011. Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication. *Proceedings of the Royal Society B: Biological Sciences*, **278**(1702): 50−57.

Lindsay EH, Opdyke ND, Johnson NM. 1980. Pliocene dispersal of the horse Equus and late Cenozoic mammalian dispersal events. *Nature*, **287**: 135−138.

Ma XY. 2018. Global Perspective of Phylogeography and Demongraphic History of *Equus asinus* Based on Mitochondrial Analysis. Master thesis, Yunnan University, China.

Mark JJ. 2011. Punt. Ancient History Encyclopedia. https://www.ancient.eu/punt/.

Marshall F. 2000. The Origins and Development of African Livestock. London: University College London Press.

Marshall F. 2007. Rethinking Agriculture: Archeological and Ethnoarcheological Perspectives. Walnut Creek, CA, USA: Left Coast Press.

Meutchieye F, Kwalar NN, Nyock RAF. 2017. Donkey Husbandry and Production Systems. Current Donkey Production and Functionality · Relationship with Humans. F. J. N. González, UCO Press.

Mukhtār MJa-D . 1981. Ancient Civilizations of Africa. London: Heinemann Educational Books.

Pérez-Pardal L, Grizelj J, Traoré A, Cubric-Curik V, Arsenos G, Dovenski T, Marković B, Fernández I, Cuervo M, Álvarez I, Beja-Pereira A, Curik I, Goyache F. 2014. Lack of mitochondrial DNA structure in Balkan donkey is consistent with a quick spread of the species after domestication. *Animal Genetics*, **45**(1): 144−147.

R-Core-Team. 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/.

Rogers AR, Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**(3): 552−569.

Rosenbom S, Costa V, Al-Araimi N, Kefena E, Abdel-Moneim AS, Abdalla MA, Bakhiet A, Beja-Pereira A. 2015. Genetic diversity of donkey populations from the putative centers of domestication. *Animal Genetics*, **46**(1): 30−36.

Rossel S, Marshall F, Peters J, Pilgram T, Adams MD, O'Connor D. 2008. Domestication of the donkey: timing, processes, and indicators. *Proceedings of the National Academy of Sciences of the United States of America*, **105**(10): 3715−3720.

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, **34**(12): 3299−3302.

Shi NN, Fan L, Yao YG, Peng MS, Zhang YP. 2014. Mitochondrial genomes of domestic animals need scrutiny. *Molecular Ecology*, **23**(22): 5393−5397.

Smith DG, Pearson RA. 2005. A review of the factors affecting the survival of donkeys in semi-arid regions of Sub-Saharan Africa. *Tropical Animal Health and Production*, **37**(S1): 1−19.

South A. 2011. Rworldmap: a new R package for mapping global data. *The R Journal*, **3**(1): 35−43.

Starkey P. 2000. The History of Working Animals in Africa. London, UK: University College London Press.

Sun Y, Jiang Q, Yang C, Wang X, Tian F, Wang Y, Ma Y, Ju Z, Huang J, Zho X u, Zhong J, Wang C. 2016. Characterization of complete mitochondrial genome of Dezhou donkey (*Equus asinus*) and evolutionary analysis. *Current Genetics*, **62**(2): 383−390.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version6.0. *Molecular Biology and Evology*, **30**(12): 2725−2729.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings Bioinformatics*, **14**(2): 178−192.

Xia X, Yu J, Zhao X, Yao Y, Zeng L, Ahmed Z, Shen S, Dang R, Lei C. 2019. Genetic diversity and maternal origin of Northeast African and South American donkey populations. *Animal Genetics*, **50**(3): 266−270.

Xie CX. 1987. The *Equus* Breeds in China. Shanghai: Shanghai Science & Technical Publishers. (in Chinese)

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