Quaternary Science Reviews

Mitochondrial genomes of Late Pleistocene caballine horses from China belong to a separate clade

--Manuscript Draft--

Manuscript Number:	JQSR-D-20-00315R1
Article Type:	Research Paper
Keywords:	Equus dalianensis; Equus przewalskii; Pleistocene caballine horses; ancient DNA; phylogenetic relationship; divergence time
Corresponding Author:	Junxia Yuan CUG China
First Author:	Junxia Yuan
Order of Authors:	Junxia Yuan
	Guilian Sheng
	Michaela Preick
	Boyang Sun
	Xindong Hou
	Shungang Chen
	Ulrike Helene Taron
	Axel Barlow
	Linying Wang
	Jiaming Hu
	Tao Deng
	Xulong Lai
	Michael Hofreiter
Abstract:	There were several species of Equus in northern China during the Late Pleistocene, including Equus przewalskii and Equus dalianensis . A number of morphological studies have been carried out on E. przewalskii and E. dalianensis , but their evolutionary history is still unresolved. In this study, we retrieved near-complete mitochondrial genomes from E. dalianensis and E. przewalskii specimens excavated from Late Pleistocene strata in northeastern China. Phylogenetic analyses revealed that caballoid horses were divided into two subclades: the New World and the Old World caballine horse subclades. The Old World caballine horses comprise of two deep phylogenetic lineages, with modern and ancient Equus caballus and modern E. przewalskii forming lineage I, and the individuals in this study together with one Yakut specimen forming lineage II. Our results indicate that Chinese Late Pleistocene caballoid horses showed a closer relationship to other Eurasian caballine horses than that to Pleistocene horses from North America. In addition, phylogenetic analyses suggested a close relationship between E. dalianensis and the Chinese fossil E. przewalskii , in agreement with previous researches based on morphological analyses. Interestingly, E. dalianensis and the fossil E. przewalskii were intermixed rather than split into distinct lineages, suggesting either that gene flow existed between these two species or that morphology-based species assignment of palaeontological specimens is not always correct. Moreover, Bayesian analysis of the divergence time between the New World and the Old World caballoid horses was dated at 1.02 Ma (95% CI: 0.86 - 1.24 Ma), and the two Old World lineages (I & II) split at 0.88 Ma (95% CI: 0.69 - 1.13 Ma), which indicates that caballoid horses seem to have evolved into different populations in the Old World soon after they migrated from North America via the Bering Land Bridge. Finally, the TMRCA of E. dalianensis was estimated at 0.20 Ma (95% CI: 0.15 - 0.28 Ma), and it showed a relative low genetic diversity compared

	with other Equus species.
Suggested Reviewers:	Ludovic Orlando Toulouse, France, Universite Toulouse III Paul Sabatier Centre de Ressources Documentaire de l'IUT Ponsan lorlando@snm.ku.dk Specialist on paleogenomes and molecular evolution of Equus.
	Ian Barnes London, UK, Natural History Museum London i.barnes@nhm.ac.uk Specialist on mammal evolution in the Quaternary period
Response to Reviewers:	

Dear Editors,

We would like to submit our manuscript **Mitochondrial genomes of Late Pleistocene caballine horses from China belong to a separate clade** to QUATERNARY SCIENCE REVIEWS.

Equus przewalskii and *Equus dalianensis* were two most fascinating caballoid horses in China. Poliakov (1881) erected the species *E. przewalskii* based on a specimen of wild horse from the eastern Junggar Basin, Xinjiang, China. Zhou *et al*. (1985) erected another caballoid species, i.e. *E. dalianensis*, based on a series of dental sets and metapodials excavated from Gulongshan, Dalian, China. Fossil *E. przewalskii* was widely distributed in China during Late Pleistocene and its footprints even reached Europe. *E. dalianensis* was endemic to China, mainly restricted to northeastern China. A number of morphological studies have been carried out on these species, however, their evolutionary history is still unresolved and no genetic study has as yet been performed.

Our study retrieves complete mitochondrial genomes from nine Late Pleistocene horse specimens excavated from northeastern China, including eight *E. dalianensis* and one *E. przewalskii* samples. Our studies indicate that the Old World caballine horses were divided into two lineages, i.e., the individuals in this study together with one Yakut specimen formed a monophyletic lineage, modern and ancient *Equus caballus* and modern *E. przewalskii* formed the other lineage. Fossil *E. przewalskii* showed a closer relationship to *E. dalianensis* than to other caballine horses based on the obtained mitogenomes. In addition, *E. dalianensis* and the fossil *E. przewalskii* were intermixed rather than split into distinct lineages, suggesting either that gene flow existed between these two species or that morphology-based species assignment of palaeontological specimens is not always correct.

Our study also provides significant details of the evolutionary history of the Old World caballoid horses. Bayesian analysis of the divergence time between the New World and the Old World caballoid horses was dated at 1.02 Ma, and the two Old World lineages split at 0.88 Ma, which indicates that caballoid horses seem to have evolved into different populations in the Old World soon after they migrated from North America via the Bering Land Bridge. Moreover, the TMRCA of *E. dalianensis* was estimated at 0.20 Ma, and it showed a relative low genetic diversity compared with other *Equus* species.

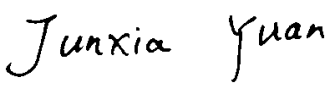
Author Contributions: Junxia Yuan, Xulong Lai and Guilian Sheng conceived the study; Junxia Yuan, Michaela Preick, Xindong Hou and Ulrike Helene Taron performed the experiments; Axel Barlow and Guilian Sheng guided the experiment and bioinformatics

analyses. Junxia Yuan, Axel Barlow, Shungang Chen and Jiaming Hu analyzed the data; Tao Deng and Boyang Sun carried out morphological analyses of the samples; Junxia Yuan, Michael Hofreiter, Guilian Sheng, Boyang Sun, Xindong Hou and Linying Wang wrote the paper. All authors read and gave comments to the final version of the manuscript.

We declare no conflict of interest in the submission of this manuscript, and it is approved by all co-authors for publication. We would like to declare that the work described is original research that has not been published previously, and is not under consideration for publication elsewhere, in whole or in part.

We thank you for your consideration, and look forward to your decision.

Yours sincerely,
Junxia Yuan



Faculty of Materials Science and Chemistry, China University of Geosciences, Wuhan 430074, China

Professor (Associate) Junxia Yuan Email: yuanjx@cug.edu.cn Phone: +0086 027 67883022

Michael Hofreiter

Institute for Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Strasse 24-25, 14476 Potsdam, Germany

Professor Micheal Hofreiter Email: michael.hofreiter@uni-potsdam.de Phone: +0049 (0)331 / 9776321

Respones to Reviewers

Thanks for the reviewers for providing all the helpful suggestions.

Reviewer #1

[1] In order to present a more updated picture of evolutionary relationship in the genus Equus I would suggest to incorporate in the ML and BEAST analysis the recently published *E. hydruntinus* mtDNA sequence (Catalano et al.2020).

We have added the recently published *E. hydruntinus* mtDNA sequence (GenBank no. MK574675) in the ML and BEAST analysis.

[2] I suggest to analyze molecule deamination and fragmentation patterns to estimate the authenticity of ancient *Equus* sequences.

We have analyzed DNA deamination and endogenous fragment length distributions of the obtained mitogenomes in this study, and the results have been provided in supplementary materials (Figure S1 and Figure S2).

[3] Change "Disappeared during the end-Pleistocene extinction" to "Disappeared during the Late Pleistocene extinction"

Yes, we have corrected it.

[4] Add a reference to the period "Both the fossil record and molecular dating suggested that caballoid horses might have migrated from North America to Eurasia during the late Early Pleistocene or early Middle Pleistocene"

We have added a reference (Deng and Xue, 1998) in this sentence.

[5] In Table S1 I suggest to add three more columns with the radiocarbon lab number, radiocarbon age (^{14}C) and the age used for divergence times estimate.

We have added the radiocarbon lab number and radiocarbon age (^{14}C) in Table S1. The ages of our samples used for divergence time estimate are shown in Figure 4, therefore we prefer no to add these information in Table S1.

[6] Substitute: Taxa > Taxon; Tissue > Skeletal element; Geography origin> Location

Yes, we have corrected them.

Reviewer #2

[1] - Section 2.5 of Materials and Methods

The authors produced their mitochondrial genomes by mapping reads to a mitochondrial genome, and I'd like to suggest a couple of alternatives that might improve their final mitochondrial genomes. One thing that is important to be aware of, while mapping reads to mitogenomes, is that they are circular genomes. By mapping the reads to a linearized version of the genome one might lose many reads that could have mapped to the "pseudo-edges" of the genome. One alternative would be simply creating an "extended" version of the linear mitogenome by adding a region from the opposite edge, thus allowing all reads to map, and cutting the final version manually. This would help make the coverage even across the entire mitogenome. Another alternative would be to use MIA (https://github.com/mpieva/mapping-iterative-assembler/). MIA has been used in several publications in the past few years. It is an iterative mapping program that accepts as reference a circular genome. By adding that flag it will automatically take into consideration that the 'edges' of the genome are not real. It will also map the reads iteratively, creating new consensus sequences along the way, which helps mapping reads from a different subspecies to a single reference genome. These two approaches could certainly help the final mitochondrial genomes, but I don't believe the final results of the paper would be affected by the approach currently used by the authors.

We have checked all accession numbers and corrected them.

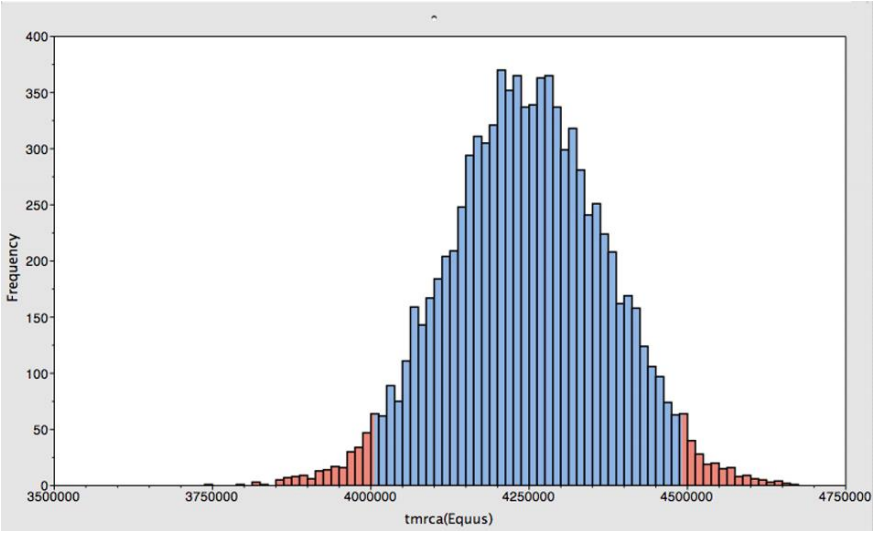
According to the reviewer’s suggestion, the trimmed reads have been mapped to the “extended” reference by adding 20 bp from the opposite edge. New consensus sequences look better than before, especially in 5’ and 3’ end of the sequences.

[2] Section 2.6 of Materials and Methods

The authors describe their methods for divergence time estimates in this section. I would like to have a little more details in their methods section. I'm assuming from the

results section that tip dating was also used to calibrate the tree, since it is mentioned in the results, but in the description of this analysis there's no mention of tip dating. Also, the authors calibrate the tree based on the TMRCA obtained from Orlando et al. 2013. It is a common thing to do, but I found 4.0 - 4.5 Ma to be a very narrow distribution for the calibration node. Was this calibration set with hard or soft bounds? When looking at the Tracer results for this parameter, is there a normal distribution, or does it tend towards one of the edges of the calibration range? This is a crucial parameter that can directly impact all dating findings, and should be further discussed in the methods. Also, the authors mention that they followed a partition scheme based on the PartitionFinder results, but they don't mention in the Methods section what regions they cover and what nucleotide substitution parameters are used in each partition.

In this study, we used root-and-tip-dating calibrations to investigate the divergence times among caballine horse lineages. We considered either the median radiocarbon or the strata age of specimens as tip-dating calibration points (please see 2.6 Phylogenetic analysis section). The detailed age information of all the samples for divergence times estimation was shown in Figure 4. The root calibration of the TMRCA of *Equus* (4.0 - 4.5 Ma) was set with soft bounds. When looking at the Tracer results for this parameter, there is a normal distribution (please see the following screenshot). The posterior age of this calibration parameter is 4.25 Ma and the effective sample size is 431 (ESS > 200). We have added these information in Materials and Methods section. In addition, we have added nucleotide substitution parameters in Section 2.6 of Materials and Methods.



[3] - Results, first paragraph

The authors describe the mitogenomes coverages varying from 14.5 to 137.4 fold. That is a very wide coverage distribution. I would have liked to see a coverage plot for these mitogenomes. Do we see consistently double/triple coverage at a certain region of the genome? Is there just a spike at a small region, how does this variation spread over the genome? This plot could be part of the supplemental material and would strengthen the results, while providing a way of checking these mitogenomes for duplicated regions in the hyper-variable region, for example.

We have provided the coverage plots for the obtained mitogenomes in supplemental materials section (Figure S3). We have not seen consistently double/triple coverage at a certain region of the genomes, however, there is a spike in D-loop region for all the samples.

[4] - Figure 2

Since the authors chose to make some Figures in color, I think it would help the reader follow more closely the results if different colors were used to identify *E. przewalskii* and *E. dalianensis* and other groups, not only in this figure, but all other figures. Matching the colors from this figure and figures 3 and 4 would be nice and helpful too.

Yes, we have corrected them.

[5] Also, in Figure 4, an indication of time in the x axis would be a nice addition for those interested in the timing of the clades that do not have numbers indicating their age. It would also allow knowing exactly the dates for the 95% HPD bars.

We have added the x axis in Figure 4.

[6] - Figure 5

I believe Figure 5 should be removed from this manuscript. A bar plot figure to indicate numbers is not informative. This information should be presented as a table. A table would have actual numbers that can be used in future studies and allow for seeing the actual differences between species.

We have made Table 1 replacing Figure 5.

Highlights

1. In this study, nine near-complete mitochondrial genomes were retrieved from *Equus dalianensis* and *Equus przewalskii* specimens excavated from Late Pleistocene strata in northeastern China.

2. The Old World caballine horses comprise of two phylogenetic lineages, and Late Pleistocene caballine horses from China belong to a separate clade

3. The divergence time between the New World and the Old World caballoid horses was dated at 1.02 Ma, and the two Old World lineages split at 0.88 Ma, which indicates that caballoid horses seem to have evolved into different populations in the Old World soon after they migrated from North America via the Bering Land Bridge.

4. The TMRCA of *Equus dalianensis* was estimated at 0.20 Ma, and it showed a relative low genetic diversity compared with other *Equus* species

Mitochondrial genomes of Late Pleistocene caballine horses from China belong to a separate clade

Junxia Yuan[a,b,*], Guilian Sheng[b], Michaela Preick[c], Boyang Sun[d], Xindong Hou[b], Shungang Chen[a], Ulrike Helene Taron[c], Axel Barlow[c,e], Linying Wang[a], Jiaming Hu[b], Tao Deng[d], Xulong Lai[b], Michael Hofreiter[c,*]

[a] Faculty of Materials Science and Chemistry, China University of Geosciences, Wuhan 430074, China

[b] State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Wuhan 430074, China

[c] Institute for Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Strasse 24-25, 14476 Potsdam, Germany

[d] Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, Beijing 100044, China

[e] School of Science and Technology, Nottingham Trent University, Clifton Lane, Nottingham, NG11 8NS, UK.

[*] Correspondence: Junxia Yuan, e-mail: yuanjx@cug.edu.cn; Michael Hofreiter, e-mail: michael.hofreiter@uni-potsdam.de

Abstract

There were several species of *Equus* in northern China during the Late Pleistocene, including *Equus przewalskii* and *Equus dalianensis*. A number of morphological studies have been carried out on *E. przewalskii* and *E. dalianensis*, but their evolutionary history is still unresolved. In this study, we retrieved near-complete mitochondrial genomes from *E. dalianensis* and *E. przewalskii* specimens excavated from Late Pleistocene strata in northeastern China. Phylogenetic analyses revealed that caballoid horses were divided into two subclades: the New World and the Old World caballine horse subclades. The Old World caballine horses comprise of two deep phylogenetic lineages, with modern and ancient *E. quus caballus* and modern *E.*

przewalskii forming lineage Ⅰ, and the individuals in this study together with one Yakut specimen forming lineage Ⅱ. Our results indicate that Chinese Late Pleistocene caballoid horses showed a closer relationship to other Eurasian caballine horses than that to Pleistocene horses from North America. In addition, phylogenetic analyses suggested a close relationship between *E. dalianensis* and the Chinese fossil *E. przewalskii*, in agreement with previous researches based on morphological analyses. Interestingly, *E. dalianensis* and the fossil *E. przewalskii* were intermixed rather than split into distinct lineages, suggesting either that gene flow existed between these two species or that morphology-based species assignment of palaeontological specimens is not always correct. Moreover, Bayesian analysis of the divergence time between the New World and the Old World caballoid horses was dated at ~~0.99~~1.02 Ma (95% CI: 0.~~84~~ 86 - 1.~~17~~ 24 Ma), and the two Old World lineages (Ⅰ & Ⅱ) split at 0.~~84~~ 88 Ma (95% CI: 0.~~67~~ 69 - 1.~~04~~ 13 Ma), which indicates that caballoid horses seem to have evolved into different populations in the Old World soon after they migrated from North America via the Bering Land Bridge. Finally, the TMRCA of *E. dalianensis* was estimated at 0.~~16~~ 20 Ma (95% CI: 0.~~12~~ 15 - 0.~~22~~ 28 Ma), and it showed a relative low genetic diversity compared with other *Equus* species.

Keywords

Equus dalianensis, *Equus przewalskii*, Pleistocene caballine horses, ancient DNA, phylogenetic relationship, divergence time

1. Introduction

The genus *Equus* first originated in North America and spread to Eurasia via the Bering Land Bridge during the climatic cold event 2.5 million years ago (Ma) (Azzaroli et al., 1988; Forsten, 1988; Deng and Xue, 1998; Sun and Deng, 2019). Several species, including *Equus przewalskii*, *Equus dalianensis*, *Equus hemionus*, *Equus kiang* and *Equus ovodovi*, have been described from Late Pleistocene northern China (Zhou et al., 1985; Dong et al., 1996; Deng and Xue, 1998; Yuan et al., 2019), with two of these species, the ass species *E. kiang* and *E. hemionus*, having survived

to the present (Deng and Xue, 1998). In contrast, *E. dalianensis* and *E. ovodovi* disappeared during the Late ~~end~~ Pleistocene extinction (Dong et al., 1996). It was believed that *E. przewalskii* also survived and represents the only extant wild horse. However, recent molecular data suggested that modern Przewalski's horses are the feral descendants of horses herded at Botai and not truly wild horses (Gaunitz et al., 2018).

According to the currently available fossil evidence, *E. dalianensis* and *E. przewalskii* both lived in China since the early Late Pleistocene (Deng and Xue, 1998). *E. dalianensis* was endemic to China, mainly restricted to northeastern China (Zhou et al., 1985; Deng and Xue, 1998). In contrast, *E. przewalskii* was more widely distributed in China during the Late Pleistocene, and its habitat ranged from the north of China to the Taiwan Strait (Deng and Xue, 1998; Deng, 1999; Gao, 2000; Nie et al., 2008; Dong et al., 2009). These two species of caballine horses were often found coexisting in Late Pleistocene faunas of northeastern China, such as the Gulongshan, Xiaogushan, Miaohoushan, Yushu, Yanjiagang and Qingshantou faunas as well as others (Xu et al., 1985; Zhou et al., 1985; Dong et al., 1996; Deng, 1999). When the climate became warmer at the beginning of the Holocene, *E. dalianensis* became extinct and the habitat of *E. przewalskii* shrank. Currently, Przewalski's horses only occur as reintroductions. Genomic studies suggest that modern Przewalski's horses are in fact descendants of Botai horses (Gaunitz et al., 2018), raising questions regarding the relationships between Pleistocene and modern Przewalski horses and, by extension, also about the relationship between *E. dalianensis* and Pleistocene *E. przewalskii*.

Morphologically, *E. dalianensis* was a quite large wild horse, described as larger than contemporary *E. przewalskii* and *E. hemionus* (Zhou et al., 1985). Previous palaeontological and morphological studies suggested that *E. dalianensis* was most closely related to Chinese Late Pleistocene *E. przewalskii* (Zhou et al., 1985; Deng and Xue, 1998). Deng and Xue (1998) proposed that *E. dalianensis* and fossil *E.*

przewalskii might be sister tax~~a~~on and that they were the direct descendants of *Equus beijingensis*. Unfortunately, *E. beijingensis* material is comparatively rare, since it has only been found at locality 21 of Zhoukoudian, dating to the late Middle Pleistocene or the early Late Pleistocene (Liu, 1963). Hence, the evolutionary origins of *E. dalianensis* and the Pleistocene *E. przewalskii* remain unclear.

Several studies investigated the occlusal enamel morphology of Late Pleistocene *Equus* cheek tooth to detect taxonomic and phylogenetic signals (Barrón-Ortiz et al., 2017; Cucchi et al., 2017). Ouyang and Xu (1993) also studied the enamel structure of Chinese Late Pleistocene equid cheek tooth, i.e., *E. dalianensis* and *E. przewalskii*, from Gulongshan cave. Surprisingly, their results suggested that it was the Late Pleistocene *E. dalianensis* rather than the Late Pleistocene *E. przewalskii* that was more similar to the modern *E. przewalskii* in terms of the average diameter of enamel rods and the width of buccal lateral enamel.

Despite the abundant materials available for detailed morphological analyses of Chinese *E. dalianensis* and *E. przewalskii* (Zhou et al., 1985; Ouyang and Xu, 1993; Dong et al., 1996; Deng and Xue, 1998; Deng, 1999; Dong et al., 2009), up to now, little is known about the evolutionary histories of these two species, such as their geographical origins, dispersals and relationships with contemporary caballine horses from other regions.

Molecular analysis is an effective tool to explore evolutionary relationship among species. Jansen et al. (2002) and Goto et al. (2011) verified that domestic horses were not descended from *E. przewalskii*, based on both mitochondrial and autosomal sequences. Interestingly, the results by Gaunitz et al. (2018) indicated that it was in fact rather the other way round and *E. przewalskii* are feral descendants of horses herded at Botai. Until now, what we know about *E. dalianensis* and Late Pleistocene *E. przewalskii* from China is based exclusively on information obtained from morphological studies (Zhou et al., 1985; Ouyang and Xu, 1993; Deng and Xue,

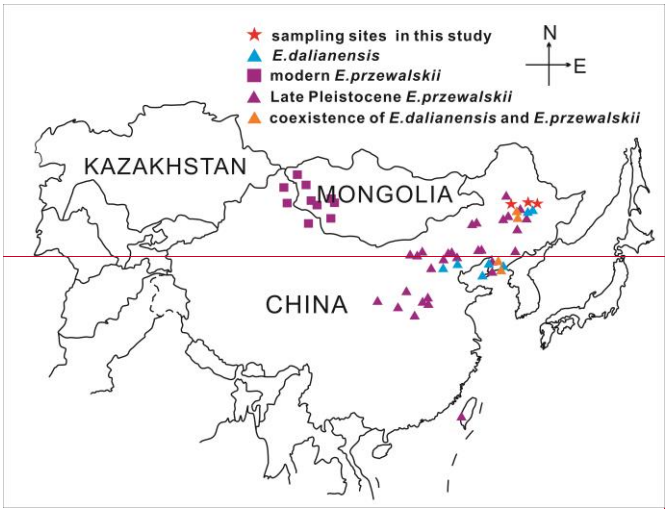
1998), while, to our knowledge, no genetic study has as yet been performed.

In this study, we retrieved nine almost complete mitochondrial genomes of *E. dalianensis* and *E. przewalskii* fossil specimens collected from northeastern China. This dataset allowed us to explore their precise phylogenetic status and revealed the relationship between Chinese Late Pleistocene caballine horses and their ancient and modern counterparts from other regions of the world.

2. Materials and methods

2.1 Samples

Nine Late Pleistocene equid fossil specimens were analyzed, comprising eight *E. dalianensis* and one *E. przewalskii* individuals, which were excavated from three sites, Zhaodong, Tonghe, and Harbin, all in Heilongjiang province, northeastern China (Fig. 1). AMS-^{14}C dating of these specimens was carried out at Peking University & BETA Analytic. Detailed information on the samples is provided in Table S1.



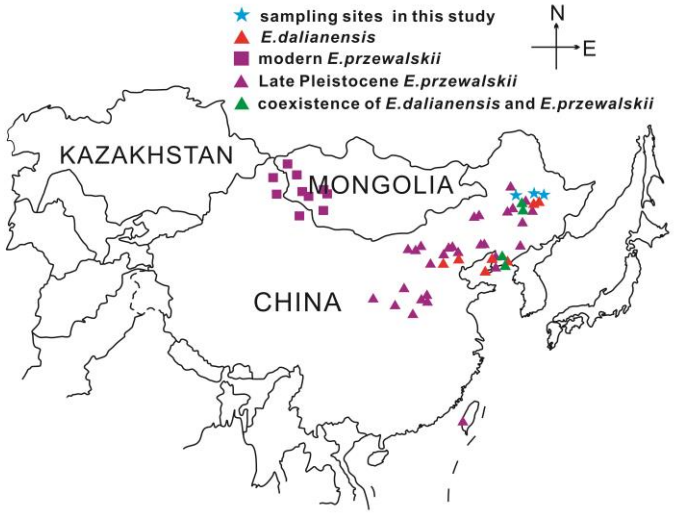


Fig. 1. Geographic sites of *E. dalianensis* and *E. przewalskii* in China. Fossil findings of *E. przewalskii* and *E. dalianensis* and the modern *E. przewalskii* distribution area are shown according to published literatures. Sampling sites in this study are indicated by ~~red~~ blue stars; the modern *E. przewalskii* distribution area is shown by purple squares; fossil findings of *E. dalianensis* are shown by ~~blue~~ red triangles; fossil findings of *E. przewalskii* are shown by purple triangles, and sites of co-existence of fossil *E. przewalskii* and *E. dalianensis* are shown by ~~orange~~ green triangles.

2.2 DNA extraction

DNA extractions were performed in a laboratory dedicated to ancient DNA work following the protocol of Dabney et al. (2013), with several modifications previously described in Yuan et al. (2019). Approximately 50 mg of bone or tooth powder was used for each sample, followed by overnight incubation in 1 mL extraction buffer (0.45 M EDTA, 0.25 mg/mL proteinase K) under gentle rotation at 37°C, and blank controls were included in each DNA extraction session. The samples were then centrifuged for 2 minutes at maximum speed to pellet the powder, and the supernatant was purified on a silica-based spin-column. Finally, the DNA was eluted in 25 μL TET buffer.

2.3 Library construction

We built single-stranded DNA libraries using 20 μL DNA extract of each sample following the protocol described by Gansauge and Meyer (2013), with the slight modifications as described in Basler et al. (2017) and Yuan et al. (2019). The optimal number of cycles for the indexing PCR was estimated by qPCR. Indexing PCR was performed using 20 μL template library in a total reaction volume of 80 μL. After amplification, PCR products were purified using silica spin columns (Qiagen MinElute) following the manufacturer's instructions, and DNA was eluted twice by adding 10 μL EB buffer each time. Libraries were then quantified using Qubit 2.0 and 2200 TapeStation (Agilent Technologies) to measure the final library concentration and fragment size distribution. Additionally, blank controls were also included in the library preparation to monitor potential contamination.

2.4 Hybridization capture

Hybridization capture of the complete mitochondrial genome was carried out following previously published procedures (González-Fortes and Paijmans, 2019). Total DNA was extracted from a modern horse sample, and the baits were prepared by PCR amplifying the mitochondrial genome using of four overlapping long range PCR primer pairs (Vilstrup et al., 2013; Yuan et al., 2019). The amplified modern horse mitochondrial DNA fragments were then sheared, blunt-end repaired and ligated to biotinylated adapters for use as hybridization capture baits. We carried out two rounds of capture to improve the enrichment rate as detailed in Yuan et al. (2019). Finally, the enriched libraries were purified using MinElute columns (Qiagen), pooled and sequenced on the Illumina NextSeq 500 sequencing platform, using 75 bp single-end sequencing and custom sequencing primers for the single-stranded libraries, following the procedures described in Paijmans et al. (2017).

2.5 Data analysis

Raw reads were trimmed to remove adapter sequences using Cutadapt v1.18 (Martin, 2011), requiring a minimum adapter overlap of 1 bp and discarding fragments shorter

than 30 bp. All other Cutadapt parameters were left as default. The trimmed reads were then mapped to an ~~*Equus* *caballus*~~ *E. przewalskii* mitochondrial reference genome (GenBank No. ~~X79547~~AP013095) using the "aln" and "sampe" algorithms in Burrows-Wheeler aligner (BWA 0.7.15-r1140) (Li and Durbin, 2010) with default parameters. It should be noted that the reference sequence was an "extended" version of the linear mitogenome by adding 20 bp from the opposite edge. Next, reads with mapping quality less than 30 were excluded using samtools v1.9 "view" and potential PCR duplicates removed with samtools "rmdup" (Li et al., 2009). Finally, consensus sequences were determined by using ~~ANGSD~~Geneious v10.~~916~~ 1.3 (https://www.geneious.com/~~Korneliussen et al., 2014~~), with parameters a minimum read depth 2 and 75% majority rule for consensus calling.

2.6 Phylogenetic analysis

To investigate the phylogenetic relationships of *E. dalianensis* and *E. przewalski* mitochondrial haplotypes, we carried out ML phylogenetic analysis with RAxML-HPC v8.2.12 (Stamatakis, 2014). The eight newly obtained *E. dalianensis* and one *E. przewalski* near-complete mitochondrial genomes were aligned with ~~576~~ 577 *Equus* and 16 *Haringtonhippus* sequences downloaded from GenBank (Table S2) using the MAFFT (Katoh et al., 2002) algorithm in the Cipres Science Gateway v 3.3 (Miller et al., 2010). In addition, 6 *Hippidion* sequences were chosen as out-group to root the tree (Table S2). Regions that were difficult to align were deleted resulting in a final alignment of 16,~~521~~ 557 bp. Selection of the most appropriate partitioning scheme and substitution models were performed in PartitionFinder v2.1.1 (Lanfear et al., 2016), resulting in the GTR+G nucleotide substitution model for eight partitions (File S1). The reliability of the branches was assessed using 500 bootstrap replicates.

To better visualize the relationship of *E. dalianensis* and *E. przewalskii*, a median-joining network was also reconstructed with Network v~~5.0.0.3~~10100 (fluxus-engineering.com/sharenet.htm) using the near-complete mitochondrial DNA

sequences newly retrieved in this study and modern *E. przewalskii* mitochondrial sequences from GenBank (Table S2), default settings were applied.

To investigate the divergence times among caballine horse lineages, we performed a molecular dating analysis based on the whole mitogenome data using BEAST 1.8.2 (Drummond et al., 2012). Vilstrup et al. (2013) estimated the time to the most recent common ancestor (TMRCA) of all equids around 4.3 Ma (95% CI: 4.0 - 4.7 Ma) according to complete mitochondrial genomes. Orlando et al. (2013) inferred a minimal date of 4.07 Ma for the TMRCA of *Equus*, and they proposed 4.0 - 4.5 Ma for the TMRCA of all living *Equus*. In this study, we assumed TMRCA of all equids of 4.0 - 4.5 Ma (soft bounds)~~(Orlando et al., 2013),~~ and considered the median radiocarbon or strata age of specimens as calibration points (root-and-tip-dating calibrations). A total of ~~95~~ 96 mitochondrial genomes were used in this analysis, including seven specimens from this study, together with 52 caballine and ~~36~~ 37 non-caballine horse sequences (Table S2) retrieved from GenBank. The estimation analysis was conducted under a relaxed uncorrelated lognormal molecular clock, coalescent Bayesian Skyline tree model, ~~and~~ a partitioning scheme with eight partitions indicated by PartitionFinder v2.1.1 (Lanfear et al., 2016), and the GTR substitution model was considered. Markov Chain Monte Carlo (MCMC) runs were carried out with 50,000,000 iterations each, sampling every 5,000 steps. Results were checked using the program Tracer v1.7 (Rambaut et al., 2018). The posterior age of the TMRCA of all living *Equus* was estimated around 4.25 Ma and there was a normal distribution for this calibration parameters (ESS > 200), which is also similar to the estimate by Vilstrup et al. (2013). The first ~~12~~30,~~500~~000,000 iterations were discarded as an appropriate burn-in, ~~verified using the program Tracer v1.7 (Rambaut et al., 2018). The~~ the maximum clade credibility tree was annotated with relevant statistics using TreeAnnotator v1.5.4 (Drummond et al., 2012) and viewed in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree).

Nucleotide diversity (π) of *E. dalianensis* haplotypes compared with other caballine,

i.e., modern *E. przewalskii* (Table S2) and *Equus caballus* (Table S3), and non-caballine equids (Table S2), i.e., *E. ovodovi*, *Equus burchellii*, *Equus grevyi*, *Equus zebra*, *E. kiang*, *E. hemionus* and *Equus asinus*, was calculated using MEGA 7 (Kumar et al., 2016) , excluding all positions containing gaps or missing data.

3. Results

In this study, we retrieved the mitochondrial genomes from eight *E. dalianensis* individuals and one Late Pleistocene *E. przewalskii* sample. We successfully obtained 16,~~274~~ 277 - 16,~~553~~ 556 bp mitochondrial DNA from all of these specimens with the sequencing depth varying between 14.5 ~~—~~and ~~137~~138.~~4~~ 3 fold (Table S1). DNA damage rates, endogenous fragment length distributions and coverage plots for the analyzed samples in this study were provided in supplementary materials (see Fig. S1 - S3).

The maximum-likelihood phylogenetic tree (Fig. 2) reveals similar relationships to previous molecular studies (Orlando et al., 2009; Vilstrup et al., 2013; Der Sarkissian et al., 2015a; Druzhkova et al., 2017; Yuan et al., 2019). The mitochondrial genomes were divided into two main clades: caballine and non-caballine horses. Non-caballine horses include *E. zebra*, *E. grevyi*, *E. burchellii*, *E. asinus*, *E. hemionus*, *E. kiang* and *E. ovodovi*. The caballine cluster includes the species *E. caballus*, *E. przewalskii*, *E. dalianensis*, *Equus* cf. *scotti* and *Equus* cf. *lambei*, and this clade is further divided into three main lineages, all supported with 100% bootstrap value (Fig. 2). *E. caballus* and modern *E. przewalskii* form lineage Ⅰ. The specimens in this study (Late Pleistocene *E. przewalskii* and *E. dalianensis*) form lineage Ⅱ. One Late Pleistocene individual (GenBank No. KT757749) collected from Yakutia (Russian Federation) also clusters in lineage Ⅱ in a basal position. Finally, Pleistocene individuals from Yukon (*E.*cf. *scotti*, *E.* cf. *lambei* and *E. caballus*) form the third monophyletic group (lineage Ⅲ). This North American lineage was already identified by Vilà et al. (2001). Moreover, our results potentially indicate that the two Old World mitochondrial lineages (lineage Ⅰ and lineage Ⅱ) are sister groups, although this relationship is only

Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic

supported with ~~75~~81% bootstrap value. Notably, the Chinese Late Pleistocene *E. przewalskii* sample TH3 falls within lineage Ⅱ in our phylogenetic analyses, while present-day *E. przewalskii* are scattered across lineage Ⅰ (Fig. 2).

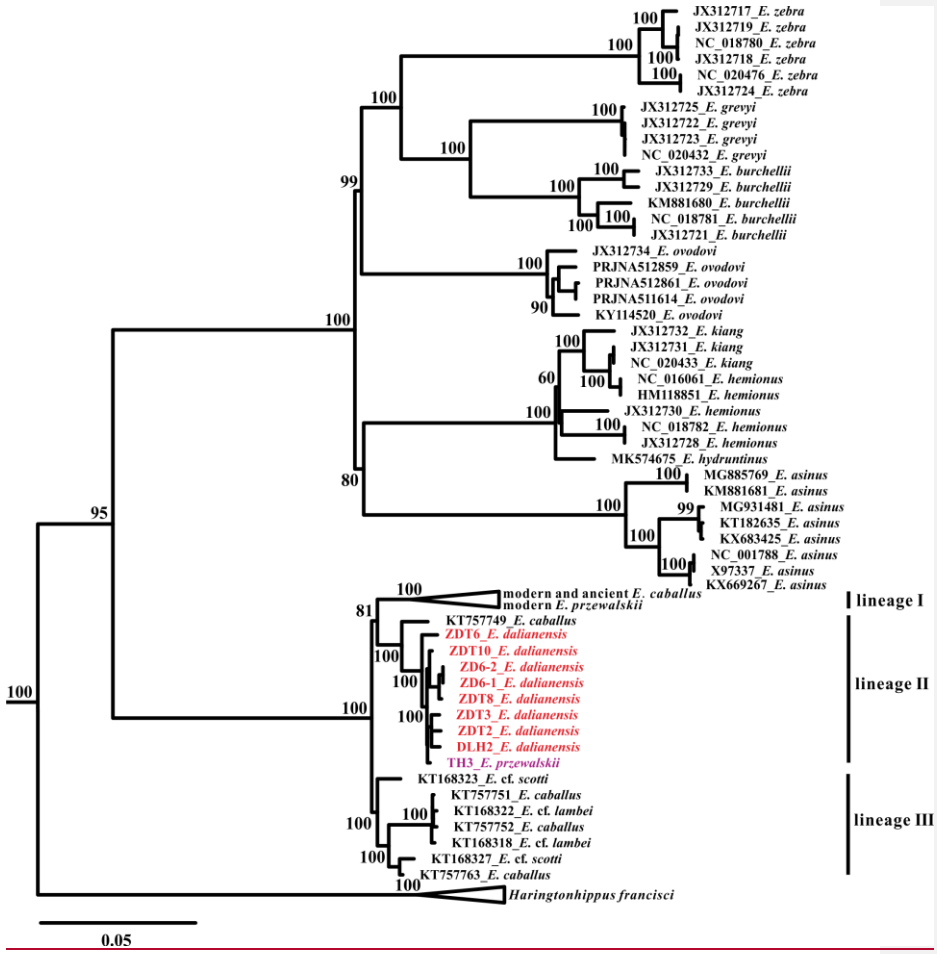


Fig. 2. Maximum-likelihood phylogenetic tree of complete mitochondrial genomes of equids using the *Hippidion* clade as outgroup. Branch labels show bootstrap values derived from 500 replications.

The median-joining network analysis of *E. przewalskii* and *E. dalianensis* haplotypes further supported the results obtained from the RAxML tree. The network shows two distinct haplogroups separated by 176 substitutions, consisting of our ancient specimens and modern *E. przewalskii*, respectively (Fig. 3). In the specimen haplogroup, seven haplotypes were identified from eight *E. dalianensis* individuals, with the Late Pleistocene *E. przewalskii* haplotype falling within the diversity of the *E.*

dalianensis haplotypes, while it was highly divergent from the modern *E. przewalskii* haplotypes with 208 - ~~243~~ 240 bp differences.

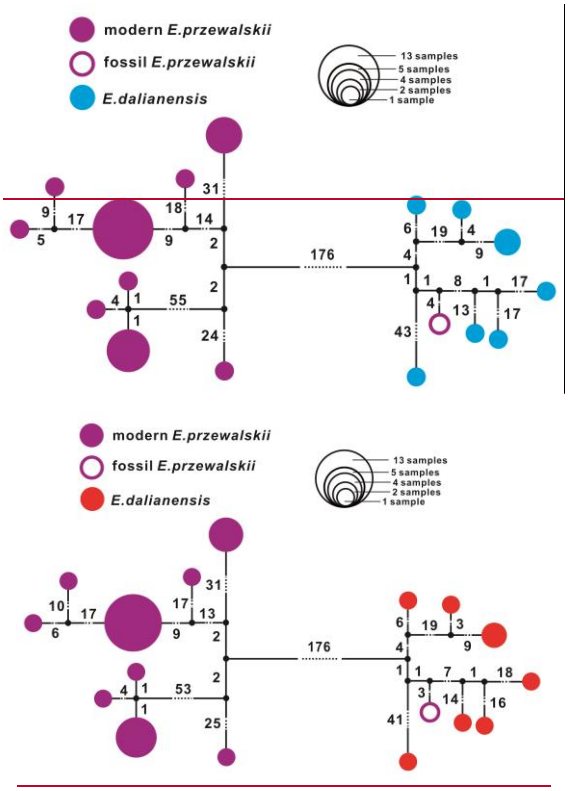
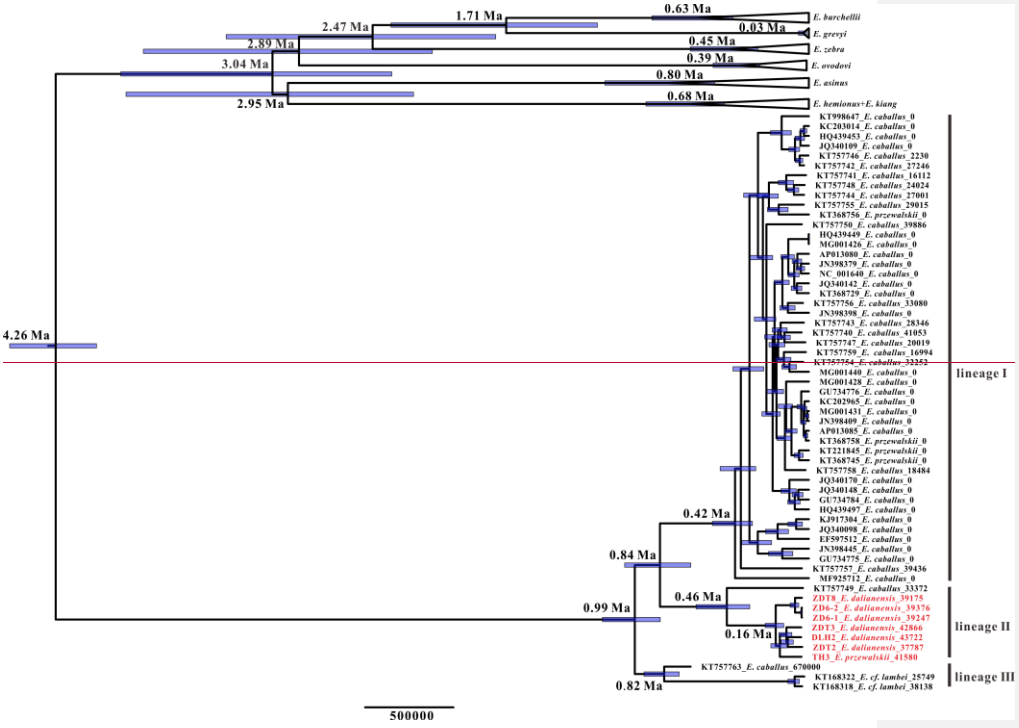


Fig. 3. Median-joining Network generated with mitochondrial genomes of the specimens in this study and modern *E. przewalskii* from GenBank. The size of circles corresponds to the number of individuals. Branch lengths are not proportional to the mutational steps, the numbers of mutations between two haplotypes are shown on the dash lines.

The molecular dating analysis in BEAST (Fig. 4) yielded the same topology as the maximum-likelihood phylogenetic analysis in RAxML (Fig. 2). The mean substitution rate across the whole mitochondrial tree was estimated at 1.~~30~~50×10^{-8} substitutions/ site/ year~~, with a 95% credibility interval 1.07 - 1.54×10^{-8} substitutions/ site/ year~~. Our result is similar to the lower estimate obtained by Vilstrup et al. (2013), who used different model parameters and calibration dates and obtained mean substitution rates

for *Equus* ranging from 1.17×10^{-8} to 3.96×10^{-8} substitutions/ site/ year. Moreover, we obtained a divergence time of mt-lineage Ⅰ/Ⅱ and lineages Ⅲ, i.e., the TMRCA of caballine horses was dated to ~~0.99~~1.02 Ma (95% CI: 0.~~84~~ 86 - 1.~~17~~ 24 Ma). The divergence time for the Old World caballine lineage combining clades Ⅰ and Ⅱ was estimated to around 0.~~84~~ 88 Ma (95% CI: 0.~~67~~ 69 - 1.~~04~~ 13 Ma). Our estimate of the time to the MRCA of lineage Ⅱ was 0.~~46~~ 49 Ma (95% CI: 0.33 - 0.~~64~~ 68 Ma), and the estimated coalescence of the available Chinese individuals was about 0.~~16~~ 20 Ma (95% CI: 0.~~12~~ 15 - 0.~~22~~ 28 Ma) (Table S4).



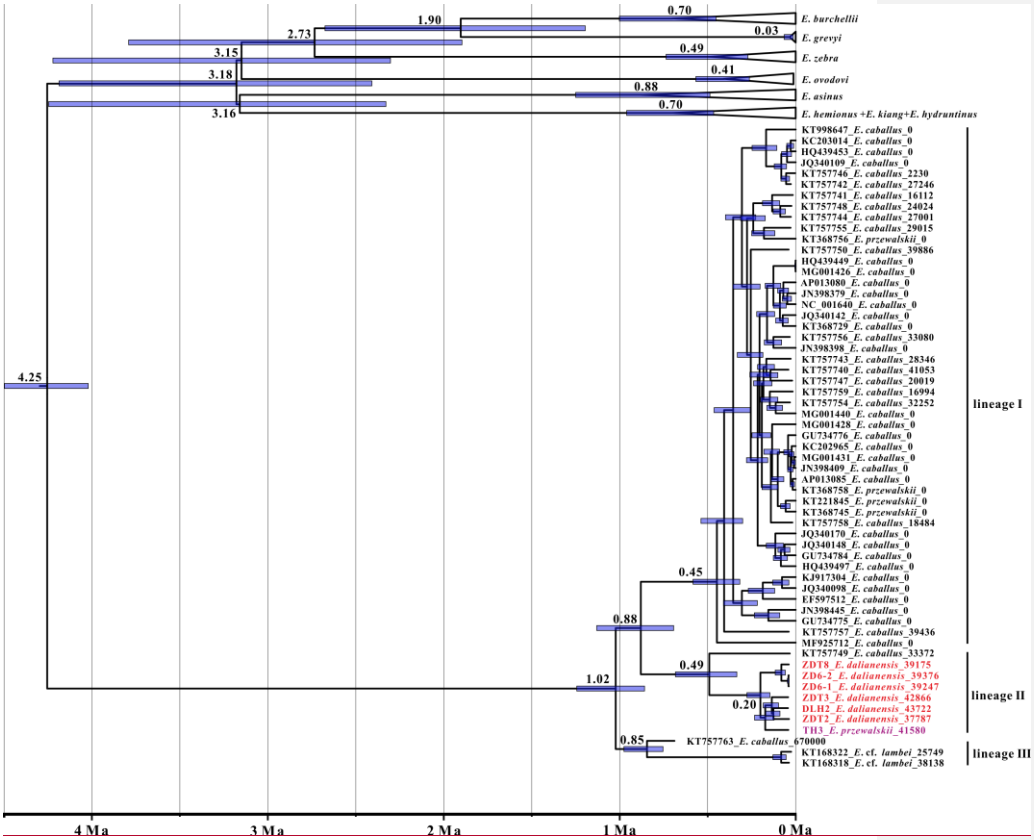
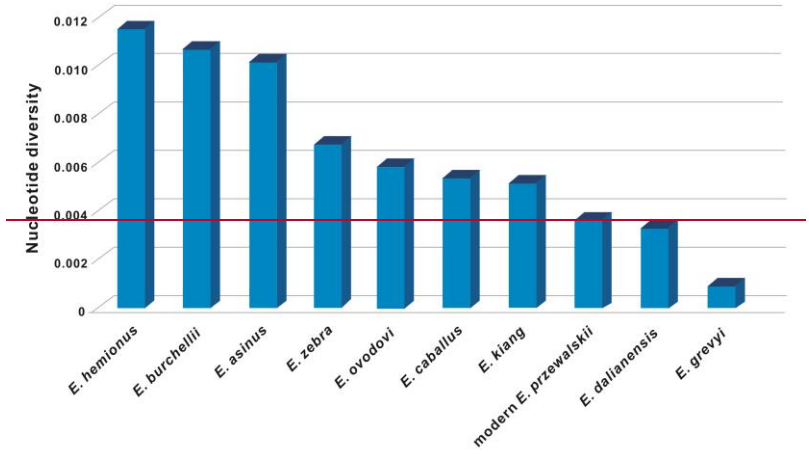


Fig. 4. Maximum clade credibility tree of the genus *Equus* in BEAST based on complete mitogenomes. Nodes heights are centered on the median posterior age estimates (x-axis; Ma). Blue node bars show 95% credibility intervals of the divergence times. Tip dates were used to calibrate the tree as followings: Ages of caballine specimens are indicated following the sample/accession number. The non-caballine equids, samples of *E. burchellii*, *E. grevyi*, *E. zebra*, *E. hemionus*, *E. kiang* and *E. asinus* are all from modern individuals; the ages of *E. ovodovi* individuals used were the same as Yuan *et al*. (2019) and the age of *E. hydruntinus* specimen (GenBank no. MK574675) was set at 22,000 BP (Catalano *et al*., 2020).

Finally, we calculated nucleotide diversity within *Equus* species (~~Fig. 5~~Table 1). Compared with other caballine and non-caballine *Equus* species, our analyses suggested that *E. dalianensis* (based on all *E. dalianensis* sequences in this study)

possessed relatively low mitochondrial diversity, which is similar to modern *E. przewalskii* and only a little higher than that of *E. grevyi*, two species listed by IUCN as critically endangered.



~~Fig.~~Table ~~5~~ 1

Nucleotide diversity within equid species based on mitogenomes.

Taxon	Nucleotide diversity	Taxon	Nucleotide diversity
E. hemionus	0.011095	*E. caballus*	0.004930
E. burchellii	0.010068	*E. kiang*	0.004736
E. asinus	0.009745	modern *E. przewalskii*	0.002942
E. zebra	0.006183	*E. dalianensis*	0.002825
E. ovodovi	0.005283	*E.grevyi*	0.000355

4. Discussion

4.1 Evolutionary relationship among Chinese Late Pleistocene caballine horses and their Eurasian and North American counterparts

Pleistocene caballine horses have been assigned to a number of species, including, among others, *E.* cf. *scotti*, *E.* cf. *lambei*, *E. caballus*, *E. dalianensis* and *E. przewalskii*. During the Late Pleistocene period, *E. dalianensis* and *E. przewalskii* were widely distributed in northeastern China (Zhou et al., 1985; Deng and Xue,

Formatted: Font: Not Italic

Formatted: Font: Not Italic

1998). According to morphological characteristics, the most commonly accepted hypothesis suggests that *E. dalianensis* and *E. przewalskii* descended from *E. beijingensis* (Deng and Xue, 1998). The ancestor of *E. beijingensis* is assumed to have migrated to China either from North America or Europe. Deng and Xue (1998) suggest that *E. beijingensis* were most likely the descendants of *Equus mosbachensis*, a Europe caballine horse species. The evolutionary relationship between Chinese caballine horses and other populations from the world is unclear.

Our phylogenetic analyses revealed that the genus *Equus* was divided into two clades, i.e., caballine horse clade and non-caballine clade, as suggested in previous publications (Heitzmann et al., 2017; Yuan et al., 2019). The caballine horse clade was divided into two deep subclades, which are the New World caballine horse subclade and the Old World caballine horse subclade (Fig. 2 & Fig. 4), which is similar to the finding by Heitzmann et al. (2017). Most noteworthy, our phylogenetic trees further showed that the Old World caballine horses were grouped into two lineages (Ⅰ & Ⅱ). All sampled individuals in this study together with a Russian ancient horse formed lineage Ⅱ, while lineage Ⅰ included other Eurasian ancient and modern caballine horses. Thus, according to the mitochondrial genomes obtained so far (Fig. 2), the Late Pleistocene lineage Ⅱ might represent extinct populations. Gaunitz et al. (2018) also found, based on nuclear sequences that modern Przewalski's horses are not truly wild horses, but rather the descendants of Botai horses.

The trees suggested that the New World caballine horses (lineage Ⅲ) diverged from caballoid horse clade first (Fig. 2), which means that the Chinese Late Pleistocene caballine horses, *E. dalianensis* and *E. przewalskii*, were closer to other Eurasian caballine horses than to the Late Pleistocene North America counterparts. Current molecular evidence thus supports the hypothesis that the direct ancestors of *E. dalianensis* and *E. przewalskii* immigrated to China from other areas of Eurasia rather than from North America, similar to the result suggested by Deng and Xue (1998) based on morphological data. However, it should be noted that the sister group

relationship between lineage Ⅰ and lineage Ⅱ was not supported by high bootstrap value (81~~75~~%; Fig. 2). Therefore, nuclear DNA sequences will be essential for further resolution for the phylogenetic relationship between these two lineages.

4.2 Phylogenetic relationship of *E. dalianensis* and Late Pleistocene *E. przewalskii*

In 1880, Russian explorer Przhevalsky obtained a specimen of wild horse from the eastern Junggar Basin, Xinjiang, China. Poliakov (1881) erected the species *E. przewalskii* based on this specimen, and its fossil representative is the most abundant Pleistocene caballoid horse in northern China. Zhou et al. (1985) erected another caballoid species, i.e. *E. dalianensis*, based on a series of dental sets and metapodials. The new species was based on measurements of the total lengths of Mc III and Mt III, which were 231 - 249.5 mm and 265 - 293 mm, respectively, in *E. dalianensis* and thus longer than those of *E. przewalskii* (216 - 235 mm and 255 - 275.5 mm, respectively). However, the data of *E. przewalskii* and *E. dalianensis* show partial overlap, and if combined become a constant distribution, which is difficult to divide into two areas with a clear boundary. Based on morphological characteristics of the upper and lower molars, *E. dalianensis* is similar to fossil *E. przewalskii*, and it was proposed that *E. dalianensis* and fossil *E. przewalskii* represent two closely related species (Zhou et al., 1985; Deng and Xue, 1998; Deng and Xue, 1999). Zhou et al. (1985) argued that the body size of *E. dalianensis* was larger than that of *E. przewalskii*, and researchers have tended to identify relatively large molar with more elongated protocone as that of *E. dalianensis*. Overall, after the establishment of *E. dalianensis* as a new species, the morphological distinction of Late Pleistocene equids in northern China became blurred.

In the present study, we successfully retrieved mitochondrial genomes from *E. dalianensis* and Late Pleistocene *E. przewalskii* specimens collected from northeastern China. The sequences enable us to obtain a better understanding of the phylogenetic status of *E. dalianensis* and *E. przewalskii* compared with other ancient and modern horses (Fig. 2 & Fig. 4). The results of our phylogenetic analyses all

suggest that the Late Pleistocene *E. przewalskii* and *E. dalianensis* individuals in this study, together with a single lineage from Late Pleistocene Russia formed a distinct branch (lineage Ⅱ) within the caballine equids. Moreover, our DNA analyses suggest that Late Pleistocene *E. przewalskii* may fall within the phylogenetic diversity of *E. dalianensis* (Fig. 2 & Fig. 4), as we are unable to separate them from each other at least for mitochondrial DNA. This intermixing phenomenon was also observed in other members of family Equidae despite extensive chromosomal plasticity (Jónssen et al., 2014), e.g. for *E. kiang* and *E. hemionus* (Vilstrup et al., 2013). The validity of the species designations of *E. kiang* and *E. hemionus* was questioned on the basis of these results, but Rosenbom et al. (2015) revealed that mitochondrial introgression between *E. kiang* and *E. hemionus* had occurred in the Late Pleistocene or early Holocene at the latest, although they are currently distributed in different geographical areas, and quite deeply divergent according to nuclear data (Jónsson et al., 2014). Therefore, more Pleistocene *E. przewalski* samples as well as the analysis of nuclear data will be necessary to clarify the evolutionary relationship of Late Pleistocene *E. przewalskii* and *E. dalianensis.*

Until now, the relationship among modern Przewalski's horses, Late Pleistocene Przewalski's horses and domestic horses still remains contentious. Der Sarkissian et al. (2015b) suggested that domestic horses and Przewalski's horses split about 45,000 years ago and they had remained connected by gene-flow thereafter. Surprisingly, the recent study by Gaunitz et al. (2018) based on ancient and modern horse genomes revealed that modern Przewalski's horses were the feral descendants of horses herded at Botai. However, modern Przewalski's horses and domestic horses exhibit different karyotypes, as Przewalski's horses possess 2n=66 chromosomes while domestic horses have only 2n=64 chromosomes (Kefena et al., 2012). In this study, mitochondrial phylogenetic trees (Fig. 2 & Fig. 4) indicated that all modern *E. przewalskii* specimens were scattered in lineage Ⅰ, and did not group together with the Late Pleistocene *E. przewalskii* specimen in lineage Ⅱ. Our results suggest that Late Pleistocene Przewalski's horses were probably different from modern

Przewalski's horses. The fossil Przewalski's horse clustered within the now extinct *E. dalianensis* lineage, suggesting that fossil and modern Przewalski's horses may represent different evolutionary lineages. Again, more samples and nuclear DNA sequences will be required to resolve the status of fossil Przewalski's horses including their relationship to their modern namesake.

4.3 Divergence times of the Old World caballoid horse lineages

Caballoid horses are thought to have appeared first in North America. To date, the oldest record of caballoid horse is believed to originate from the early Irvingtonian (1.9 - 1.3 Ma) Red Cloud Formation of Nebraska (Eisenmann, 1992). A previous study using the nuclear genome of a 0.56 - 0.78 Ma Alaskan horse estimated the TMRCA of all equids at around 4.0 - 4.5 Ma (Orlando et al., 2013). According to this node age and tip-calibration, we inferred the divergence time between the New World caballine horses and the Old World populations at about ~~0.99~~1.02 Ma (95% CI: 0.~~84~~ 86 - 1.~~17~~ 24 Ma) based on the mitochondrial DNA sequences obtained so far (Fig. 4), which is slightly older than the estimate by Heintzmann et al. (2017). *E.* cf. *scotti* was the earliest known representative of caballoid horse (Eisenmann, 1992), but it seems that there might have been other caballoid populations before *E.* cf. *scotti* appeared. Up to now, the oldest Eurasian remains of caballoid horse were found in Siberia, dating to around 0.7 Ma (Sher, 1986). Both the fossil record (Deng and Xue, 1998) and molecular dating (Fig. 4) suggested that caballoid horses might have migrated from North America to Eurasia during the late Early Pleistocene or early Middle Pleistocene. In addition, our divergence estimate suggests a split of Eurasian caballoid horses between lineage Ⅰ and lineage Ⅱ dating back to about 0.~~84~~ 88 Ma (Fig. 4). Thus, soon after they migrated from North America via the Bering Land Bridge, caballoid horses seem to have diverged into different populations in the Old World. Based on current analysis, lineage Ⅱ was mainly distributed in Northeast Asia, while the specimens of lineage Ⅰ scattered in Eurasia. It would be interesting to further investigate the geographical distribution of lineage Ⅱ in future studies.

Formatted: Font: Not Italic

Formatted: Font: Not Italic

In comparison to other members of horse family, *E. dalianensis* exhibits a relatively low level of nucleotide diversity, similar to modern *E. przewalskii* and only slightly higher than that of *E. grevyi* (~~Fig. 5~~Table 1), which are currently limited to small geographic ranges. Both of them have experienced severe bottlenecks and suffered losses of genetic diversity (Cordingley et al., 2009; Goto et al., 2011). The low nucleotide diversity may reflect that *E. dalianensis* experienced one or several bottlenecks during its evolutionary history, or alternatively, that its population size was always restricted. Although our exploration from a limited number of individuals cannot provide a conclusive answer, the current analyses established yet another mitochondrial clade within the Pleistocene representatives of the genus *Equus*. Fully understanding the evolutionary history of this important group of species will clearly require substantial palaeogenomic data.

Acknowledgments: This research was financially supported by the National Natural Science Foundation of China (Nos. 41472014; 41672017) and the "PPP" project jointly funded by CSC and DAAD (No. 2016-2041). We are thankful to Mr. Guoqing Peng for his help in collecting the samples. We thank Federica Alberti, Stefanie Hartmann and Johanna Paijmans (University of Potsdam, Germany) for their help in the process of this research. We also appreciate two anonymous reviewers for their valuable comments that improved the manuscript.

Author Contributions: Junxia Yuan, Xulong Lai and Guilian Sheng conceived the study; Junxia Yuan, Michaela Preick, Xindong Hou and Ulrike Helene Taron performed the experiments; Axel Barlow and Guilian Sheng guided the experiment and bioinformatics analyses. Junxia Yuan, Axel Barlow, ~~and~~ Shungang Chen and Jiaming Hu analyzed the data; Tao Deng and Boyang Sun carried out morphological analyses of the samples; Junxia Yuan, Michael Hofreiter, Guilian Sheng, Boyang Sun, Xindong Hou and Linying Wang wrote the paper. All authors read and gave comments to the final version of the manuscript.

References

Azzaroli, A., De Giuli, C., Ficcarelli, G., Torre, D., 1988. Late Pliocene to early Mid-Pleistocene mammals in Eurasia: faunal succession and dispersal events. Palaeogeography, Palaeoclimatology, Palaeoecology 66, 77-100. https://doi.org/10.1016/0031-0182(88)90082-X.

Barrón-Ortiz, C.I., Rodrigues, A.T., Theodor, J.M., Kooyman, B.P., Yang, D.Y., Speller, C.F., 2017. Cheek tooth morphology and ancient mitochondrial DNA of Late Pleistocene horses from the western interior of North America: implications for the taxonomy of North American Late Pleistocene *Equus*. PLoS ONE 12, e0183045. https://doi.org/10.1371/journal.pone.0183045.

Basler, N., Xenikoudakis, G., Westbury, M.V., Song, L., Sheng, G., Barlow, A., 2017. Reduction of the contaminant fraction of DNA obtained from an ancient giant panda bone. BMC Res. Notes. 10, 754. https://doi.org/10.1186/s13104-017-3061-3.

Catalano, G., Modi, A., Mangano, G., Sineo, L., Lari, M., Bonfiglio, L., 2020. A mitogenome sequence of an *Equus hydruntinus* specimen from Late Quaternary site of San Teodoro Cave (Sicily, Italy). Quat. Sci. Rev. 236, 106280. https://doi.org/10.1016/j.quascirev.2020.106280.

Cordingley, J.E., Sundaresan, S.R., Fischhoff, I.R., Shapiro, B., Ruskey, J., Rubenstein, D.I., 2009. Is the endangered Grevy's zebra threatened by hybridization? Anim. Conserv. 12, 505-513. https://doi.org/10.1111/j.1469-1795.2009.00294.x.

Cucchi, T., Mohaseb, A., Peigné, S., Debue, K., Orlando, L., Mashkour, M., 2017. Detecting taxonomic and phylogenetic signals in equid cheek teeth: towards new palaeontological and archaeological proxies. R. Soc. Open Sci. 4, 160997. https://doi.org/10.1098/rsos.160997.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J., Meyer, M., 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. U.S.A. 110, 15758-15763.

https://doi.org/10.1073/pnas.1314445110.

Deng, T., Xue, X., 1998. The systematic evolution of the Chinese genus *Equus* fossils. Sci. China 28, 505- 510.

Deng, T., 1999. Ages of some Late Pleistocene faunas based on the presence of *Equus przewalskii* (Perissodactyla, Equidae). Vertebrata Palasiatica 23, 51-56.

Deng, T., Xue, X., 1999. Chinese fossil horses of *Equus* and their environment. Beijing: China Ocean Press.

Der Sarkissian, C., Vilstrup, J.T., Schubert, M., Seguin-Orlando, A., Eme, D., Weinstock, J., Alberdi, M.T., Martin, F., Lopez, P.M., Prado, J.L., Prieto, A., Douady, C.J., Stafford, T.W., Willerslev, E., Orlando, L., 2015a. Mitochondrial genomes reveal the extinct *Hippidion* as an outgroup to all living equids. Biology Letter 11, 1-5. https://doi.org/10.1098/rsbl.2014.1058.

Der Sarkissian, C., Ermini, L., Schubert, M., Yang, M.A., Librado, P., Fumagalli, M., Jónsson, H., Bar-Gal, G.K., Albrechtsen, A., Vieira, F.G., Petersen, B., Ginolhac, A., Seguin-Orlando, A., Magnussen, K., Fages, A., Gamba, C., Lorente-Galdos, B., Polani, S., Steiner, C., Neuditschko, M., Jagannathan, V., Feh, C., Greenblatt, C.L., Ludwig, A., Abramson, N.I., Zimmermann, W., Schafberg, R., Tikhonov, A., Sicheritz-Ponten, T., Willerslev, E., Marques-Bonet, T., Ryder, O.A., McCue, M., Rieder, S., Leeb, T., Slatkin, M., Orlando, L., 2015b. Evolutionary genomics and conservation of the endangered Przewalski's horse. Curr. Biol. 25, 2577-2583. https://doi.org/10.1016/j.cub.2015.08.032.

Dong, W., Xu, Q., Jin, C., Li, Y., 1996. Succession of the large herbivorous fauna of Northeast China in the Quaternary and its relationship with the paleoclimate. Vertebrata Palasiatica 34, 58-70. https://doi.org/10.19615/j.cnki.1000-3118.1996.01.009.

Dong, W., Fu, R., Feng, X., Zhang, S., 2009. Late Pleistocene mammalian fauna from the Mashandong, Chaoyang, Liaoning province. Acta Anthropologica Sinica 28, 95-109. https://doi.org/10.16359/j.cnki.cn11-1963/q.2009.01.005.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969-1973.

https://doi.org/10.1093/molbev/mss075.

Druzhkova, A.S., Makunin, A.I., Vorobieva, N.V., Vasiliev, S.K., Ovodov, N.D., Shunkov, M.V., Trifonov, V.A., Graphodatsky, A.S., 2017. Complete mitochondrial genome of an extinct *Equus* (*Sussemionus*) *ovodovi* specimen from Denisova cave (Altai, Russia). Mitochondrial DNA Part B 2, 79-81. https://doi.org/10.1080/23802359.2017.1285209.

Eisenmann, V., 1992. Origins, dispersals, and migrations of *Equus* (Mammalia, Perissodactyla). Courier Forschungs Institut Senckenberg 153, 161-170.

Forsten, A., 1988. Middle Pleistocene replacement of stenonid horses by caballoid horses-ecological implications. Palaeogeography, Palaeoclimatology, Palaeoecology 65, 23-33. https://doi.org/10.1016/0031-0182(88)90109-5.

Gansauge, M., Meyer, M., 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nat. Protoc. 8, 737-748. https://doi.org/10.1038/nprot.2013.038.

Gao, C., 2000. Pleistocene mammal in Dalian aera. Liaoning Geology 17, 34-38.

Gaunitz, C., Fages, A., Hanghøj, K., Albrechtsen, A., Khan, N., Schubert, M., Seguin-Orlando, A., Owens, I.J., Felke, S., Bignon-Lau, O., Damgaard, P., Mittnik, A., Mohaseb, A.F., Davoudi, H., Alquraishi, S., Alfarhan, A.H., Al-Rasheid, K.A.S., Crubézy, E., Benecke, N., Olsen, S., Brown, D., Anthony, D., Massy, K., Pitulko, V., Kasparov, A., Brem, G., Hofreiter, M., Mukhtarova, G., Baimukhanov, N., Lõugas, L., Onar, V., Stockhammer, P.W., Krause, J., Boldgiv, B., Undrakhbold, S., Erdenebaatar, D., Lepetz, S., Mashkour, M., Ludwig, A., Wallner, B., Merz, V., Merz, I., Zaibert, V., Willerslev, E., Librado, P., Outram, A.K., Orlando, L., 2018. Ancient genomes revisit the ancestry of domestic and Przewalski's horses. Science 360, 111-114. https://doi.org/10.1126/science.aao3297.

González-Fortes, G., Paijmans, J.L.A., 2019. Whole-genome capture of ancient DNA using home-made baits. In Ancient DNA: methods and protocols. Second Edition. New York: Humana Press. https://doi.org/10.1007/978-1-4939-9176-1_11.

Goto, H., Ryder, O.A., Fisher, A.R., Schultz, B., Pond, S.L.K., Nekrutenko, A., Makova, K.D., 2011. A massively parallel sequencing approach uncovers ancient origins and high genetic variability of endangered Przewalski's horses. Genome Biol. Evol. 3, 1096-1106. https://doi.org/10.1093/gbe/evr067.

Heitzmann, P.D., Zazula, G.D., MacPhee, R.D.E., Scott, E., Cahill, J.A., McHorse, B.K., Kapp, J.D., Stiller, M., Wooller, M.J., Orlando, L., Southon, J., Froese, D.G., Shapiro, B., 2017. A new genus of horse from Pleistocene North America. eLife 6, e29944. https://doi.org/10.7554/eLife.29944.001.

Jansen, T., Forster, P., Levine, M.A., Oelke, H., Hurles, M., Renfrew, C., Weber, J., Olek, K., 2002. Mitochondrial DNA and the origins of the domestic horse. Proc. Natl. Acad. Sci. U.S.A. 99, 10905-10910. https://doi.org/10.1073/pnas.152330099.

Jónsson, H., Schubert, M., Seguin-Orlando, A., Ginolhac, A., Petersen, L., Fumagalli, M., Albrechtsen, A., Petersen, B., Korneliussen, T.S., Vilstrup, J.T., Lear, T., Myka, J.L., Lundquist, J., Miller, D.C., Alfarhan, A.H., Alquraishi, S.A., Al-Rasheid, K.A.S., Stagegaard, J., Strauss, G., Bertelsen, M.F., Sicheritz-Ponten, T., Antczak, D.F., Bailey, E., Nielsen, R., Willerslev, E., Orlando, L., 2014. Speciation with gene flow in equids despite extensive chromosomal plasticity. Proc. Natl. Acad. Sci. U.S.A. 111, 18655-18660. https://doi.org/10.1073/pnas.1412627111.

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059-3066. https://doi.org/10.1093/nar/gkf436.

Kefena, E., Mekasha, Y., Han, J.L., Rosenbom, S., Haile, A., Beja-Pereira, A., 2012. Discordances between morphological systematics and molecular taxonomy in the stem line of equids: a review of the case of taxonomy of genus *Equus*. Livest. Sci. 143, 105-115. https://doi.org/10.1016/j.livsci.2011.09.017.

~~Korneliussen, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: analysis of next generation sequencing data. BMC Bioinformatics 15, 1-13. https://doi.org/10.1186/s12859-014-0356-4.~~

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870-1874. https://doi.org/10.1093/molbev/msw054.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Boil. Evol. 34(3): 772-773. https://doi.org/10.1093/molbev/msw260.

Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589-595. https://doi.org/10.1093/bioinformatics/btp698.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and samtools. Bioinformatics 25, 2078-2079. https://doi.org/10.1093/bioinformatics/btp352.

Liu, H., 1963. A new species of *Equus* from locality 21 of Zhoukoudian. Vertebrata Palasiatica 7, 318-322. https://doi.org/10.19615/j.cnki.1000-3118.1963.04.003.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17, 10-12. https://doi.org/10.14806/ej.17.1.200.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA pp 1-8.

Nie, Z., Li, H., Ma, B., 2008. Fossil fauna in the Late Pleistocene in the Hetao basin, Inner Mongolia. Quaternary Sciences 28, 14-25.

Orlando, L., Metcalf, J.L., Alberdi, M.T., Telles-Antunes, M., Bonjean, D., Otte, M., Martin, F., Eisenmann, V., Mashkour, M., Morello, F., Prado, J.L., Salas-Gismondi, R., Shockey, B.J., Wrinn, P.J., Vasil'ev, S.K., Ovodov, N.D., Cherry, M.I., Hopwood, B., Male, D., Austin, J.J., Hänni, C., Cooper, A., 2009. Revising the recent evolutionary history of equids using ancient DNA. Proc. Natl. Acad. Sci. U.S.A. 106, 21754-21759. https://doi.org/10.1073/pnas.0903672106.

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., Johnson, P.L.F., Fumagalli, M., Vilstrup, J.T., Raghavan, M., Korneliussen, T., Malaspinas, A., Vogt, J., Szklarczyk, D., Kelstrup, C.D., Vinther, J., Dolocan, A., Stenderup, J., Velazquez, A.M.V., Cahill, J., Rasmussen, M., Wang, X., Min, J., Zazula, G.D., Seguin-Orlando, A., Mortensen, C., Magnussen, K., Thompson, J.F., Weinstock, J., Gregersen, K., Røed, K.H., Eisenmann, V., Rubin, C.J., Miller, D.C., Antczak, D.F., Bertelsen, M.F., Brunak, S., Al-Rasheid, K.A.S., Ryder, O., Andersson, L., Mundy, J., Krogh, A., Gilbert, M.T.P., Kjær, K., Sicheritz-Ponten, T., Jensen, L.J., Olsen, J.V., Hofreiter, M., Nielsen, R., Shapiro, B., Wang, J., Willerslev, E., 2013. Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. Nature 499, 74-78. https://doi.org/10.1038/nature12323.

Ouyang, L., Xu, Q., 1993. Enamel structure of the equid cheek teeth from the Late Pleistocene of Dalian, Liaoning. Vertebrata Palasiatica 31, 208-216. https://doi.org/10.19615/j.cnki.1000-3118.1993.03.005.

Paijmans, J.L.A., Baleka, S., Henneberger, K., Taron, U.H., Trinks, A., Westbury, M.V., Barlow, A., 2017. Sequencing single-stranded libraries on the Illumina NextSeq 500 platform. arXiv: Other Quantitative Biology. https://arxiv.org/abs/1711.11004.

Poliakov, I.S., 1881. Przewalski's horse (*Equus przewalski* n. sp.). Isvestia Russki Geographiceski obsch-va, St Petersburg 17, 1-20.

Rambaut, A., Suchard, M.A., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol*. 67, 901-904. https://doi.org/10.1093/sysbio/syy032.

Rosenbom, S., Costa, V., Chen, S., Khalatbari, L., Yusefi, G.H., Abdukadir, A., Yangzom, C., Kebede, F., Teclai, R., Yohannes, H., Hagos, F., Moehlman, P.D., Beja-Pereira, A., 2015. Reassessing the evolutionary history of ass-like equids: insights from patterns of genetic variation in contemporary extant populations. Mol. Phylogenet. Evol. 85, 88-96. https://doi.org/10.1016/j.ympev.2015.01.005.

Sher, A.V., 1986. Olyorian land mammal age of northeastern Siberia.

Palaeontographica Italica 74, 97-112.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312-1313. https://doi.org/10.1093/bioinformatics/btu033.

Sun, B., Deng, T., 2019. The *Equus* Datum and the early radiation of *Equus* in China. Front. Ecol. Evol. 7, 429. https://doi.org/10.3389/fevo.2019.00429.

Vilà, C., Leonard, J.A., Götherström, A., Marklund, S., Sandberg, K., Lidén, K., Wayne, R.K., Ellegren, H., 2001. Widespread origins of domestic horse lineages. Science 291, 474-477. https://doi.org/10.1126/science.291.5503.474.

Vilstrup, J.T., Seguin-Orlando, A., Stiller, M., Ginolhac, A., Raghavan, M., Nielsen, S., Weinstock, J., Froese, D., Vasiliev, S.K., Ovodov, N.D., Clary, J., Helgen, K.M., Fleischer, R.C., Cooper, A., Shapiro, B., Orlando, L., 2013. Mitochondrial phylogenomics of modern and ancient equids. PLoS ONE 8, e55950. https://doi.org/10.1371/journal.pone.0055950.

Xu, Q., Jin, C., Li, C., 1985. Climate change and mammal extinction ten thousand years ago in Northeast China. Jilin Geology 1, 39-42.

Yuan, J., Hou, X., Barlow, A., Preick, M., Taron, U.H., Alberti, F., Basler, N., Deng, T., Lai, X., Hofreiter, M., Sheng, G., 2019. Molecular identification of Late and terminal Pleistocene *Equus ovodovi* from northeastern China. PLoS ONE 14, e0216883. https://doi.org/10.1371/journal.pone.0216883.

Zhou, X., Sun, Y., Xu, Q., Li, Y., 1985. Note on a new Late Pleistocene *Equus* from Dalian. Vertebrata Palasiatica 23, 69-76. https://doi.org/10.19615/j.cnki.1000-3118.1985.01.009.

Figure

Supplemental Figures S1-S3

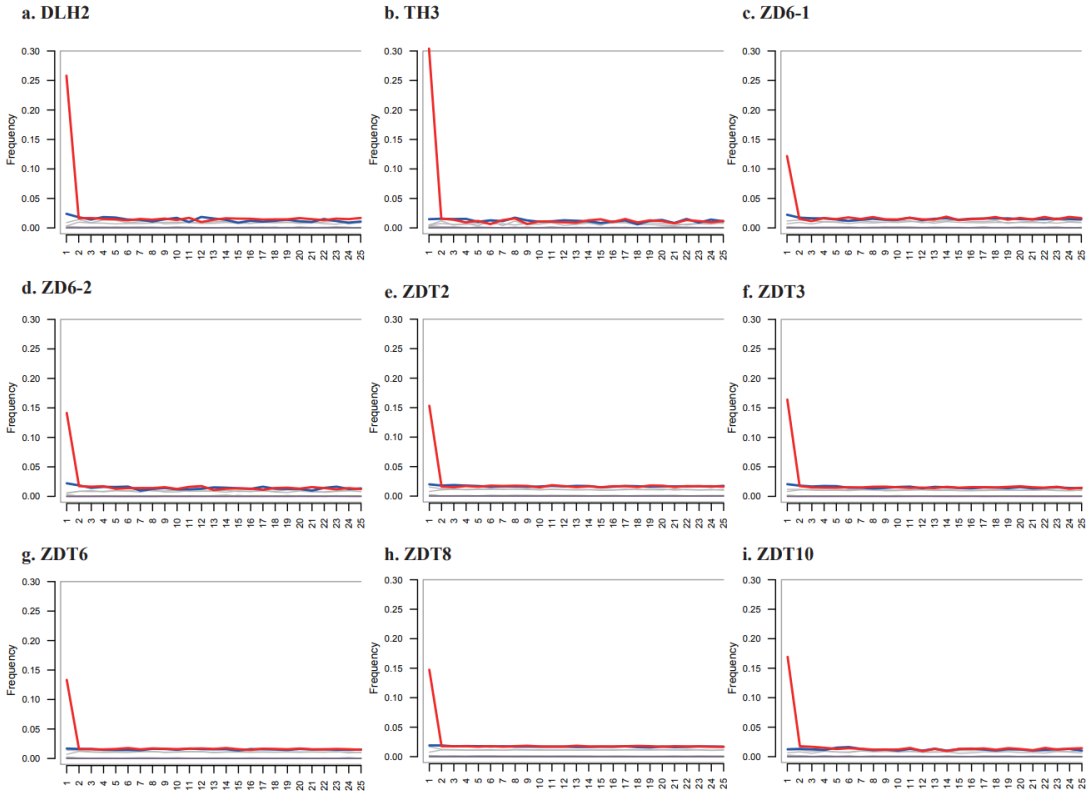


Fig. S1. Cytosine deamination frequency inferred from Late Pleostocene *Equus przewalskii* and *Equus dalianensis* samples analyzed in this study. Red lines show the rates of C to T substitutions for the first 25 bases of the 5' end of the endogenous fragments.

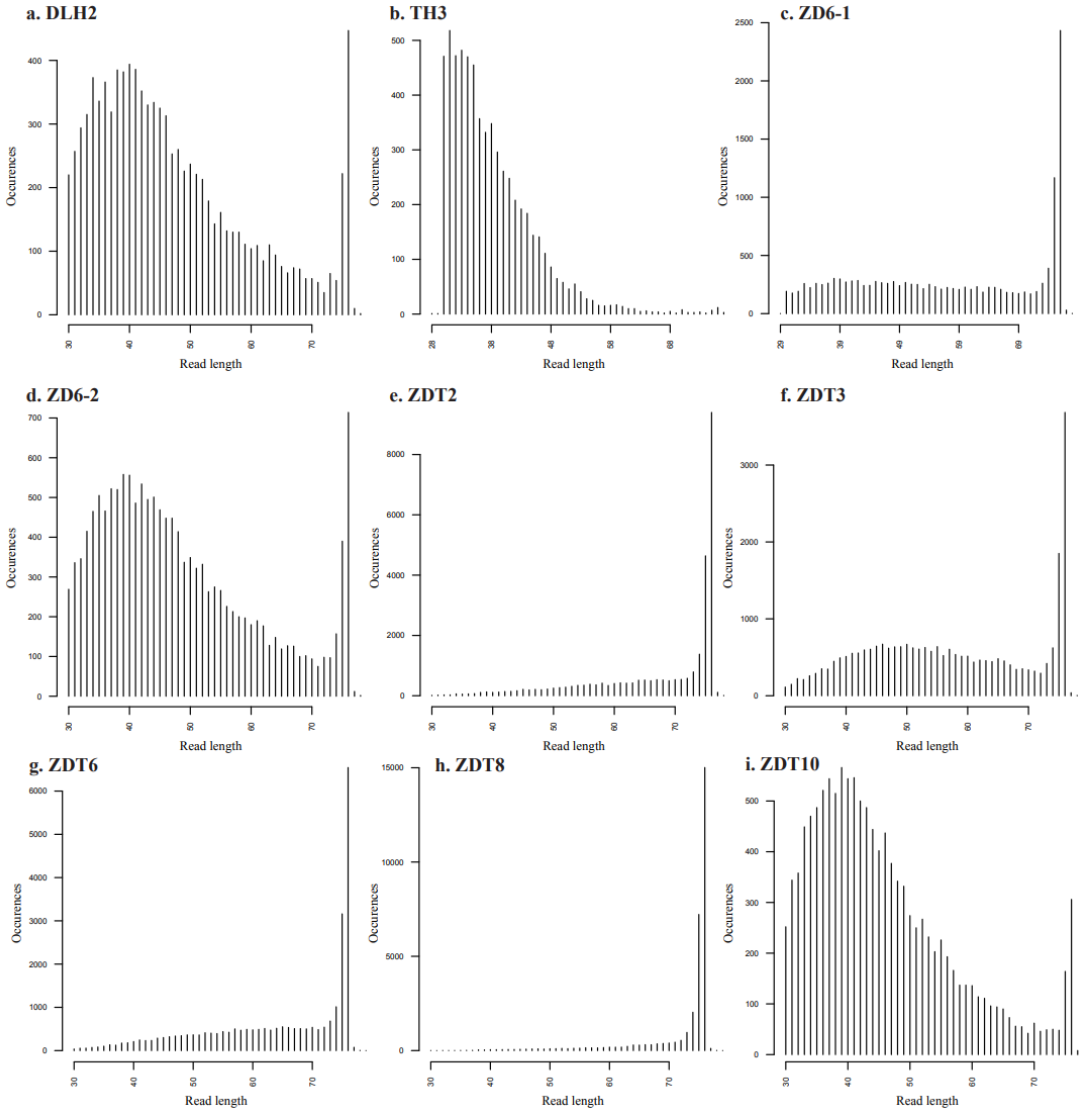
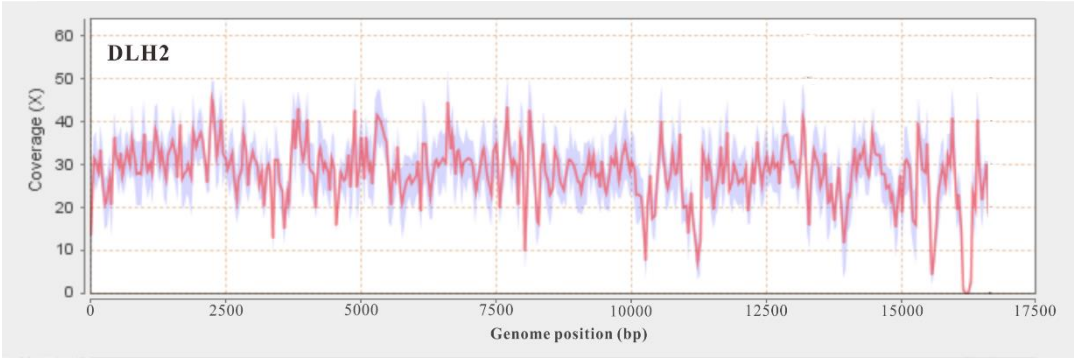
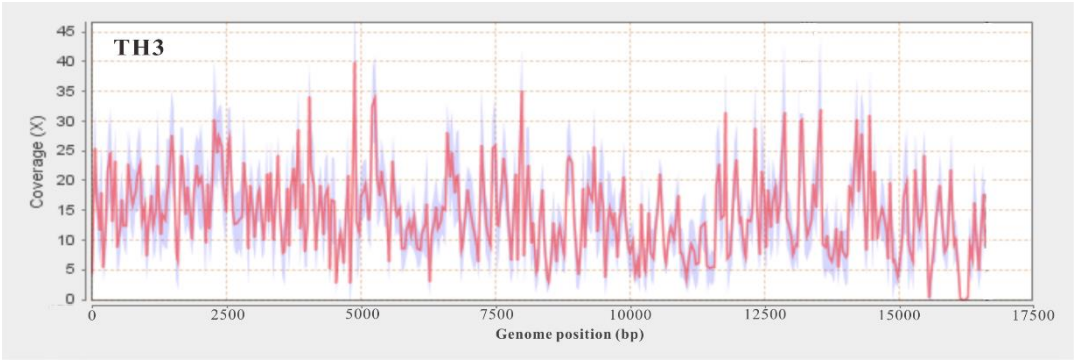
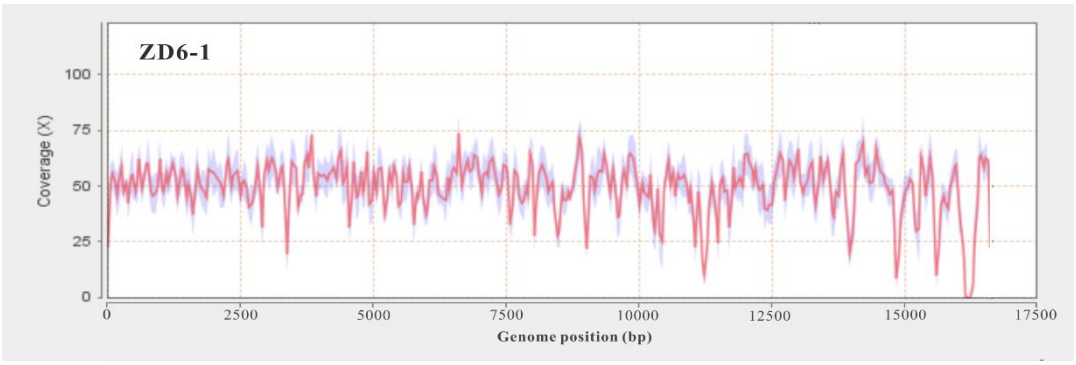
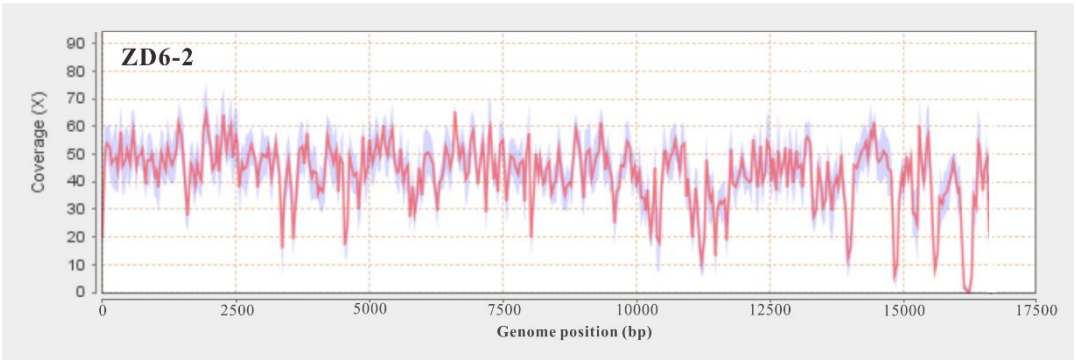


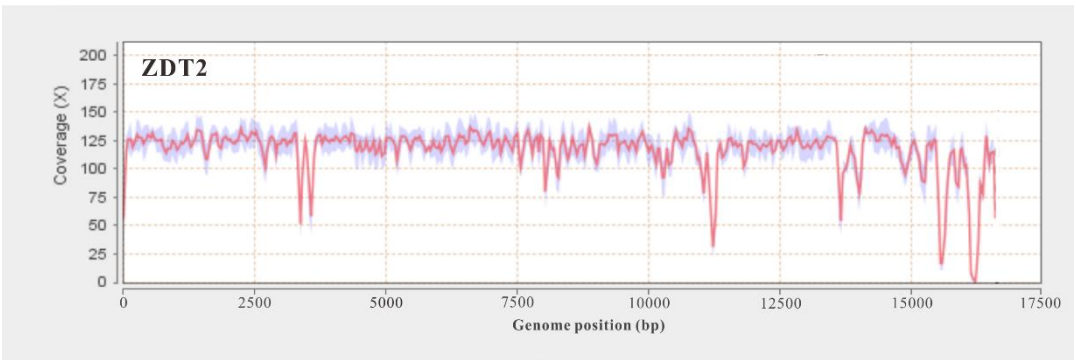
Fig. S2. Estimated endogenous fragment length distributions for Late Pleistocene *Equus przewalskii* and *Equus dalianensis* samples analyzed in this study.


DLH2
Coverage (X)
0
10
20
30
40
50
60
0
2500
5000
7500
10000
12500
15000
17500
Genome position (bp)


TH3
Coverage (X)
0
5
10
15
20
25
30
35
40
45
0
2500
5000
7500
10000
12500
15000
17500
Genome position (bp)


ZD6-1
Coverage (X)
0
25
50
75
100
0
2500
5000
7500
10000
12500
15000
17500
Genome position (bp)


ZD6-2
Coverage (X)
0
10
20
30
40
50
60
70
80
90
0
2500
5000
7500
10000
12500
15000
17500
Genome position (bp)


ZDT2
Coverage (X)
0
25
50
75
100
125
150
175
200
0
2500
5000
7500
10000
12500
15000
17500
Genome position (bp)

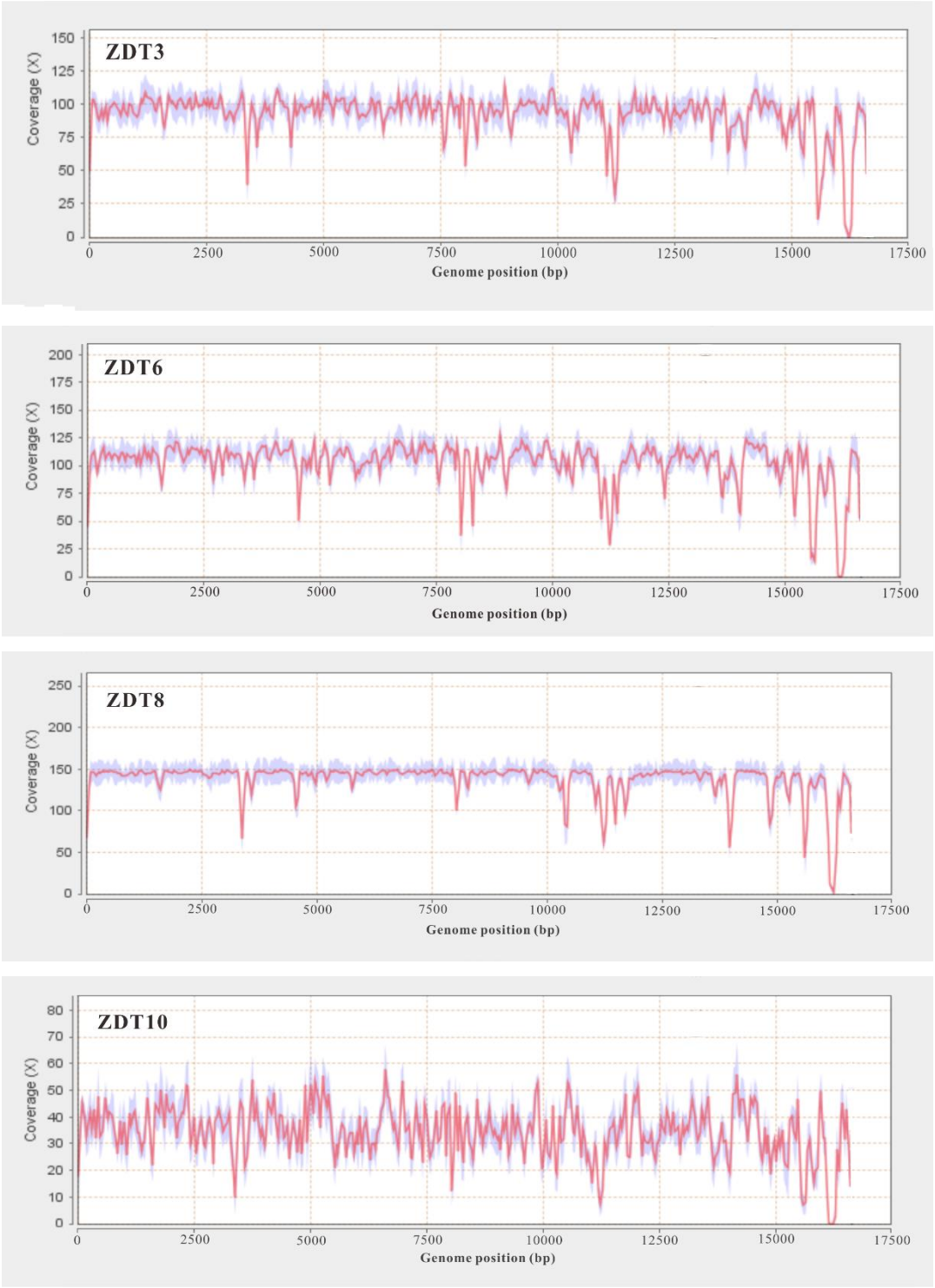


Fig. S3. Coverage plots of the mitochondrial genomes obtained in this study.

Figure

```
Settings used

alignment         : ./infile.phy
branchlengths     : linked
models            : GTR+G
model_selection   : bic
search            : greedy


Best partitioning scheme

Scheme Name       : step_53
Scheme lnL        : -81768.91015625
Scheme BIC        : 176086.459652
Number of params  : 1292
Number of sites   : 16524
Number of subsets : 8

Subset | Best Model | # sites    | subset id                         |
Partition names
1      | GTR+G      | 5455       | baa65ad27dabee39a7abea8e5c5d4633 |
ND6_CP1, ATP8_ATP6_CP1, ND2_CP1, ND5_CP1, ND4_CP3, Arg, Asp, Val, Lys,
Ala, Pro, Trp, Gly, Ile, Ser2, His, Leu2, 12s, Glu, Tyr, Phe, 16s
2      | GTR+G      | 1894       | 0e314fb30d40c4c69c154c47a335a541 |
COX1_CP2, Met, Leu1, Asn, Ser1, ND1_CP3, Gln, COX3_CP3, ND4L_CP2,
COX2_CP1, ND3_CP1
3      | GTR+G      | 2606       | 2558f62ee9afbf4a73a78dd31765fabf |
ND2_CP2, ND5_CP2, ATP8_ATP6_CP2, CYTB_CP3, ND4L_CP3, ND4_CP1, ND1_CP1,
ND3_CP2
4      | GTR+G      | 2381       | 34fad4b6b875ac4787b07b6fffed7029 |
ND5_CP3, ND6_CP2, CYTB_CP1, ND4L_CP1, ND1_CP2, ND4_CP2, ND2_CP3
5      | GTR+G      | 722        | 31fd2326a9d4be99fc4e99d04b0a0850 |
CYTB_CP2, Cys, ND6_CP3, Thr
6      | GTR+G      | 1401       | 867eb2e885d3f53af5fa2d35ba3b7fc8 |
COX3_CP2, ND3_CP3, COX2_CP3, ATP8_ATP6_CP3, COX1_CP1
7      | GTR+G      | 1004       | afa7d98b689c498f25622d2b99c2d41d |
COX1_CP3, COX3_CP1, COX2_CP2
8      | GTR+G      | 1061       | fd7ba48fc1765b1572559b168ddf7d4f | D-
loop


Scheme Description in PartitionFinder format
Scheme_step_53 = (ND6_CP1, ATP8_ATP6_CP1, ND2_CP1, ND5_CP1, ND4_CP3, Arg,
Asp, Val, Lys, Ala, Pro, Trp, Gly, Ile, Ser2, His, Leu2, 12s, Glu, Tyr,
Phe, 16s) (COX1_CP2, Met, Leu1, Asn, Ser1, ND1_CP3, Gln, COX3_CP3,
ND4L_CP2, COX2_CP1, ND3_CP1) (ND2_CP2, ND5_CP2, ATP8_ATP6_CP2, CYTB_CP3,
ND4L_CP3, ND4_CP1, ND1_CP1, ND3_CP2) (ND5_CP3, ND6_CP2, CYTB_CP1,
ND4L_CP1, ND1_CP2, ND4_CP2, ND2_CP3) (CYTB_CP2, Cys, ND6_CP3, Thr)
(COX3_CP2, ND3_CP3, COX2_CP3, ATP8_ATP6_CP3, COX1_CP1) (COX1_CP3,
COX3_CP1, COX2_CP2) (D-loop);

Nexus formatted character sets
begin sets;
```

```
      charset Subset1 = 13602-14110\3 7799-8641\3 3935-4973\3 11779-
13599\3 10210-11578\3 9840-9909 6969-7043 1047-1113 7728-7798 5044-5117
15397-15463 4974-5043 9425-9493 3725-3793 11648-11707 11579-11647 11708-
11778 71-1046 14111-14179 5289-5356 1-70 1114-2691;
      charset Subset2 = 5358-6902\3 3865-3934 2692-2766 5118-5191 6903-
6968 2769-3724\3 3794-3864 8644-9424\3 9911-10207\3 7044-7727\3 9494-
9839\3;
      charset Subset3 = 3936-4973\3 11780-13599\3 7800-8641\3 14182-
15323\3 9912-10207\3 10208-11578\3 2767-3724\3 9495-9839\3;
      charset Subset4 = 11781-13599\3 13601-14110\3 14180-15323\3 9910-
10207\3 2768-3724\3 10209-11578\3 3937-4973\3;
      charset Subset5 = 14181-15323\3 5192-5288 13600-14110\3 15324-
15396;
      charset Subset6 = 8643-9424\3 9496-9839\3 7046-7727\3 7801-8641\3
5357-6902\3;
      charset Subset7 = 5359-6902\3 8642-9424\3 7045-7727\3;
      charset Subset8 = 15464-16524;
      charpartition PartitionFinder = Group1:Subset1, Group2:Subset2,
Group3:Subset3, Group4:Subset4, Group5:Subset5, Group6:Subset6,
Group7:Subset7, Group8:Subset8;
end;


Nexus formatted character sets for IQtree
Warning: the models written in the charpartition are just the best model
found in this analysis. Not all models are available in IQtree, so you
may need to set up specific model lists for your analysis

#nexus
begin sets;
      charset Subset1 = 13602-14110\3 7799-8641\3 3935-4973\3 11779-
13599\3 10210-11578\3 9840-9909 6969-7043 1047-1113 7728-7798 5044-5117
15397-15463 4974-5043 9425-9493 3725-3793 11648-11707 11579-11647 11708-
11778 71-1046 14111-14179 5289-5356 1-70 1114-2691;
      charset Subset2 = 5358-6902\3 3865-3934 2692-2766 5118-5191 6903-
6968 2769-3724\3 3794-3864 8644-9424\3 9911-10207\3 7044-7727\3 9494-
9839\3;
      charset Subset3 = 3936-4973\3 11780-13599\3 7800-8641\3 14182-
15323\3 9912-10207\3 10208-11578\3 2767-3724\3 9495-9839\3;
      charset Subset4 = 11781-13599\3 13601-14110\3 14180-15323\3 9910-
10207\3 2768-3724\3 10209-11578\3 3937-4973\3;
      charset Subset5 = 14181-15323\3 5192-5288 13600-14110\3 15324-
15396;
      charset Subset6 = 8643-9424\3 9496-9839\3 7046-7727\3 7801-8641\3
5357-6902\3;
      charset Subset7 = 5359-6902\3 8642-9424\3 7045-7727\3;
      charset Subset8 = 15464-16524;
      charpartition PartitionFinder = GTR+G:Subset1, GTR+G:Subset2,
GTR+G:Subset3, GTR+G:Subset4, GTR+G:Subset5, GTR+G:Subset6,
GTR+G:Subset7, GTR+G:Subset8;
end;


RaxML-style partition definitions
```

```
DNA, Subset1 = 13602-14110\3, 7799-8641\3, 3935-4973\3, 11779-13599\3,
10210-11578\3, 9840-9909, 6969-7043, 1047-1113, 7728-7798, 5044-5117,
15397-15463, 4974-5043, 9425-9493, 3725-3793, 11648-11707, 11579-11647,
11708-11778, 71-1046, 14111-14179, 5289-5356, 1-70, 1114-2691
DNA, Subset2 = 5358-6902\3, 3865-3934, 2692-2766, 5118-5191, 6903-6968,
2769-3724\3, 3794-3864, 8644-9424\3, 9911-10207\3, 7044-7727\3, 9494-
9839\3
DNA, Subset3 = 3936-4973\3, 11780-13599\3, 7800-8641\3, 14182-15323\3,
9912-10207\3, 10208-11578\3, 2767-3724\3, 9495-9839\3
DNA, Subset4 = 11781-13599\3, 13601-14110\3, 14180-15323\3, 9910-10207\3,
2768-3724\3, 10209-11578\3, 3937-4973\3
DNA, Subset5 = 14181-15323\3, 5192-5288, 13600-14110\3, 15324-15396
DNA, Subset6 = 8643-9424\3, 9496-9839\3, 7046-7727\3, 7801-8641\3, 5357-
6902\3
DNA, Subset7 = 5359-6902\3, 8642-9424\3, 7045-7727\3
DNA, Subset8 = 15464-16524


MrBayes block for partition definitions

begin mrbayes;

      charset Subset1 = 13602-14110\3 7799-8641\3 3935-4973\3 11779-
13599\3 10210-11578\3 9840-9909 6969-7043 1047-1113 7728-7798 5044-5117
15397-15463 4974-5043 9425-9493 3725-3793 11648-11707 11579-11647 11708-
11778 71-1046 14111-14179 5289-5356 1-70 1114-2691;
      charset Subset2 = 5358-6902\3 3865-3934 2692-2766 5118-5191 6903-
6968 2769-3724\3 3794-3864 8644-9424\3 9911-10207\3 7044-7727\3 9494-
9839\3;
      charset Subset3 = 3936-4973\3 11780-13599\3 7800-8641\3 14182-
15323\3 9912-10207\3 10208-11578\3 2767-3724\3 9495-9839\3;
      charset Subset4 = 11781-13599\3 13601-14110\3 14180-15323\3 9910-
10207\3 2768-3724\3 10209-11578\3 3937-4973\3;
      charset Subset5 = 14181-15323\3 5192-5288 13600-14110\3 15324-
15396;
      charset Subset6 = 8643-9424\3 9496-9839\3 7046-7727\3 7801-8641\3
5357-6902\3;
      charset Subset7 = 5359-6902\3 8642-9424\3 7045-7727\3;
      charset Subset8 = 15464-16524;

      partition PartitionFinder = 8:Subset1, Subset2, Subset3, Subset4,
Subset5, Subset6, Subset7, Subset8;
      set partition=PartitionFinder;

      lset applyto=(1) nst=6 rates=gamma;
      lset applyto=(2) nst=6 rates=gamma;
      lset applyto=(3) nst=6 rates=gamma;
      lset applyto=(4) nst=6 rates=gamma;
      lset applyto=(5) nst=6 rates=gamma;
      lset applyto=(6) nst=6 rates=gamma;
      lset applyto=(7) nst=6 rates=gamma;
      lset applyto=(8) nst=6 rates=gamma;
```

```
      prset applyto=(all) ratepr=variable;
      unlink statefreq=(all) revmat=(all) shape=(all) pinvar=(all)
tratio=(all);

end;
```

Table

Conflict of Interest

All co-authors declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Statement

All co-authors would like to declare that this work described is original research that has not been published previously, and is not under consideration for publication elsewhere, in whole or in part. If accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder.