



# Mitochondrial genomes from modern and ancient Turano-Mongolian cattle reveal an ancient diversity of taurine maternal lineages in East Asia

Xiao-Ting Xia<sup>1</sup> · Alessandro Achilli<sup>2</sup> · Johannes A. Lenstra<sup>3</sup> · Bin Tong<sup>4</sup> · Yun Ma<sup>5</sup> · Yong-Zhen Huang<sup>1</sup> · Jian-Lin Han<sup>6</sup> · Zhou-Yong Sun<sup>7</sup> · Hong Chen<sup>1</sup> · Chu-Zhao Lei<sup>1</sup> · Song-Mei Hu<sup>7</sup> · Ning-Bo Chen<sup>1</sup>

Received: 1 November 2020 / Revised: 15 March 2021 / Accepted: 15 March 2021 / Published online: 29 March 2021  
© The Author(s), under exclusive licence to The Genetics Society 2021

## Abstract

Turano-Mongolian cattle are a group of taurine cattle from Northern and Eastern Asia with distinct morphological traits, which are known for their ability to tolerate harsh environments, such as the Asian steppe and the Tibetan plateau. Through the analysis of 170 mitogenomes from ten modern breeds, two sub-lineages within T3 (T3<sub>119</sub> and T3<sub>055</sub>) were identified as specific of Turano-Mongolian cattle. These two T3 sub-lineages, together with the previously identified T4, were also present in six Neolithic samples, dated to ~3900 years BP, which might represent the earliest domestic taurine stocks from Southwest Asia. The rare haplogroup Q, found in three Tibetan cattle, testifies for the legacy of ancient migrations from Southwest Asia and suggests that the isolated Tibetan Plateau preserved unique prehistoric genetic resources. These findings confirm the geographic substructure of Turano-Mongolian cattle breeds, which have been shaped by ancient migrations and geographic barriers.

Associate editor: Christine Baes

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41437-021-00428-7>.

✉ Song-Mei Hu  
husongmei@sina.com

✉ Ning-Bo Chen  
ningbochen@nwafu.edu.cn

- <sup>1</sup> Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi, China
- <sup>2</sup> Dipartimento di Biologia e Biotechnologie “L. Spallanzani”, Università di Pavia, Pavia, Italy
- <sup>3</sup> Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands
- <sup>4</sup> The State Key Laboratory of Reproductive Regulation and Breeding of Grassland Livestock, School of Life Sciences, Inner Mongolia University, Hohhot, China
- <sup>5</sup> College of Life Sciences, Xinyang Normal University, Xinyang, Henan, China
- <sup>6</sup> CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Institute of Animal Science, Chinese Academy of Agriculture Sciences (CAAS), Beijing, China
- <sup>7</sup> Shaanxi Provincial Institute of Archaeology, Xi’an, China

## Introduction

The domestication of taurine cattle (*Bos taurus*) is thought to have occurred in the Fertile Crescent about 10,500 years ago (Loftus et al. 1994; Helmer et al. 2005) and was followed by a spread along the Neolithic human migration routes. Turano-Mongolian cattle are a type of taurine cattle that distribute in Northeast Asia, which are morphologically and genetically distinct from the European taurine cattle (Felix 1995; Mannen et al. 2004; Chen et al. 2018) and comprise cattle from Buryat (extinct), northern and central China, Korea, Japan, Kazakhstan, Yakutia of Russia, Mongolia, and Tibet of China. Recent whole-genome-resequencing study has shown that Turano-Mongolian cattle form a separate genetic cluster (Chen et al. 2018). These cattle are not highly productive breeds, but are adapted to harsh climate and extremely cold environments. For example, the breeds lived in Asian steppe and Tibetan Plateau can adapt to the local extreme cold or alpine climatic conditions (Zhang 2011). What’s more, Yakutian cattle in northern Siberia have shown a unique cold adaptation (−60 °C), with their breeding centers close to the Polar circle (Granberg et al. 2009). With the modernization and specialization of animal husbandry, many Turano-Mongolian cattle breeds (e.g., Buryat and Altay cattle) have

been completely replaced and extinct by modern international cattle breeds or through extensive crossbreeding (Kantanen et al. 2009). The existing breeds such as Kazakh Whitehead and Japanese Black cattle are threatened with extinction (Felius 1995).

During the past two decades, studies of mitochondrial DNA (mtDNA) revealed that most domestic cattle carry either *B. taurus* T (T1–T4) or *B. indicus* I (I1 and I2) haplogroups. Haplogroup T3 is widespread, whereas T1 is centered in Africa, T2 in the Near East, and T4 in East Asia (Troy et al. 2001; Mannen et al. 2004; Achilli et al. 2009; Bonfiglio et al. 2002; Lenstra et al. 2014; Xia et al. 2019). Complete mtDNA sequences also identified three rare haplogroups (P, Q, and R) in modern taurine cattle, with Q most likely of Near Eastern origin and P and R possibly introduced by local aurochs introgression (Achilli et al. 2009; Bonfiglio et al. 2010; Olivieri et al. 2015; Noda et al. 2018). Most mitogenome studies were focused on European and African cattle, whereas the maternal origin analysis of East Asian cattle has been based only on the mtDNA D-loop region (Mannen et al. 2004; Kantanen et al. 2009; Lorenzo et al. 2016).

Here, complete mitogenomes of 170 individuals from ten cattle populations were analyzed to investigate the genetic diversity of the maternal lineages of Turano-Mongolian cattle. In order to provide a global context, these data have been compared with 212 available mitochondrial sequences from European, African, American, Southwest Asian cattle, aurochs and yak, giving a total of 382 mitogenomes. To clarify the early domestication events of Northeast Asian cattle, we also included six ancient taurine mitogenomes from the late Neolithic Shimao site (3975–3835 BP) in northern China (Chen et al. 2018). The study conducted a large-scale global survey on northeastern Asian cattle at the highest molecular resolution with the aim to reveal the phylogenetic relationships and early domestication events of Turano-Mongolian cattle.

## Materials and methods

### Animal sampling and ethics statement

A total of 80 cattle (Table S1) from four Turano-Mongolian breeds have been sampled: 17 from Mongolia and 63 from 4 regions of China, including 6 from Inner Mongolia, 46 from Tibet, 5 from Anxi and 6 from Yanbian. To minimize the degree of kinship among individuals, unrelated animals were selected based on pedigree information. The protocols for animal handling have been approved by the Faculty of Animal Policy and Welfare Committee of Northwest A&F University (FAPWC-NWAFU, Protocol number, NWAFAC1008).

### Illumina sequencing, data mining, and reconstruction of mitogenomes

Genomic DNA was extracted from ear tissues by the standard phenol–chloroform method (Sambrook and Russell 2001). Paired-end libraries with an average insert size of 500 bp were constructed for each individual and sequenced using the HiSeq 2000 platform (Illumina). In addition, we retrieved reads for 72 Turano-Mongolian samples from the Short Read Archive (Tsuda et al. 2013; Lee et al. 2014; Chen et al. 2018; Weldenegodguad et al. 2018; Wu et al. 2018), including samples from five additional breeds (Table S1) and eight late Neolithic samples (accessions SRR6942506–SRR6942513 (Chen et al. 2018)), six of which yielded sufficient coverage (15–390×, Table S1). Adapter sequences of ancient data were identified and removed using AdapterRemoval version 2.2.0 (Schubert et al. 2016). All reads were aligned to the *B. taurus* mitochondrial reference genome (V00654.1) using the Burrows-Wheeler Aligner v0.7.15 (Li and Durbin 2009) with the sub-command <aln -t 24 -l 1024 -n 0.01 -o 2>, which were subsequently converted to BAM files using the command samtools view -Sb. To improve alignment to the circularised genome 30 bp of sequence from the end of the mtDNA was attached to the beginning (numbering of accession V00654.1). BAM files were also sorted using SAMtools (v. 0.1.19) (Li et al. 2009) and subsequently filtered for removal of PCR duplicates (rmdup -s). Indel realignment was performed using the Genome Analysis ToolKit (GATK v3.8) (McKenna et al. 2010). Ancient mtDNA coverages were calculated by Qualimap. BAM alignments were transformed to FASTQ files and assembled by using Mapping Iterative Assembler v 1.0 (Briggs et al. 2009) (<https://github.com/mpieva/mapping-iterative-assembler>).

All novel sequences were deposited in GenBank under the accession numbers MT576705–MT576844. Previously published 236 mitogenomes of Asian, European, African, American, West Asian cattle, aurochs, and yak were downloaded from the NCBI database (Table S2).

### Sequence variation and phylogenetic analyses

All unclear sequence positions were verified by Integrative Genomics Viewer (Thorvaldsdóttir et al. 2012). Measures of mitogenome sequence variation, including numbers of haplotypes and variable sites, haplotype diversity ( $Hd$ ), nucleotide diversity ( $Pi$ ), and the average number of nucleotide differences ( $k$ ), were calculated using the program DnaSP v 5.10 (Librado and Rozas 2009). To reduce the sample size bias in assessing genetic diversity, we averaged over three randomly selected sets of ten samples from breeds with more than ten individuals.

Phylogenetic relationships were inferred from the complete mtDNA sequences using the maximum likelihood (ML) approach within IQ-TREE (Nguyen et al. 2015) and the Bayesian inference (BI) approach within MrBayes v3.2.0 (Ronquist and Huelsenbeck 2003). GTR + F + R4 was chosen as the best-fit model for the complete sequences of mitogenomes by using ModelFinder (Kalyaanamoorthy et al. 2017). Bootstrap support values for the ML analysis were generated with 1000 replicates (-bb 1000). The BI analysis was run with Markov chains for 20 million generations with a burn-in value of 2500 (25%). Trees were sampled every 2000 generations. The final tree topology was visualized using iTOL (Letunic and Bork 2019). To identify autochthonous haplogroups of T and Q, a maximum parsimony (MP) phylogeny was constructed using an adapted version of mtPhyl v4.015 (Eltsov and Volodko 2011), as previously described (Achilli et al. 2009; Achilli et al. 2002). A median-joining network was constructed using NETWORK 5.0.1.1 (Bandelt et al. 1999).

### Time estimates

Phylogenetic relationships within the Turano-Mongolian cattle inferred in previous analyses were used to estimate the divergence times between the major haplogroups using BEAST v2.6.0 (Drummond et al. 2002). We constructed a Bayesian tree of Turano-Mongolian cattle on the mtDNA coding region, including 158 modern and 6 ancient sequences. The tree was rooted with the *Bos grunniens*. All positions containing gaps and ambiguous data were eliminated from the dataset (FASTA). The FASTA alignment file was used to generate the BEAST XML input file for BEAST v2.6.0 (Drummond et al. 2002). We performed the analysis with the HKY model (as indicated by ModelFinder), a Relaxed Clock Exponential, and gamma-distributed rates (with four categories). Ancient samples (3975–3835 BP) and the age for PQT of  $75.5 \pm 10.0$  ky (Bonfiglio et al. 2010; Olivieri et al. 2015) were considered as the consistent internal calibration point. The major haplogroups were considered as monophyletic in order of being able to calculate their age estimates. We set the number of generations to 50 million, logging parameters every 1000 steps. Convergence was confirmed by effective sampling size >200 using the Tracer v 1.7 (<http://tree.bio.ed.ac.uk/software/tracer/>). A maximum credibility tree topology was generated using a 10% burn-in using TreeAnnotator in BEAST v2.6.0. We visualized the tree in FigTree v1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>).

We also obtained a Bayesian skyline plot (BSP) from the Turano-Mongolian cattle phylogeny using BEAST v2.6.0 (Drummond et al. 2002). Yak haplotypes were excluded, and only cattle mtDNA haplotypes were considered in this estimation. The effective population size

was estimated assuming a generation time of 6 years (Bollongino et al. 2002).

## Results

### The phylogeny and diversity of Turano-Mongolian mitogenomes

A set of 170 modern Turano-Mongolian mitogenomes was analyzed, among which 8 Tibetan and 4 Mongolian mtDNAs were identified as yak haplotypes and removed from the subsequent analysis (Tables S1 and S2). In the remaining 158 individuals, 588 variable sites were identified after quality control, 143 of which are singletons. A total of 108 haplotypes were defined by 445 sites. Thirty-six haplotypes were shared by at least two individuals (Table S3). Overall, we observed an average of 35.72 nucleotide differences between two sequences. The detailed haplogroup composition and genetic diversity of modern Turano-Mongolian are shown in Table 1.

Geographic distributions of maternal haplogroups are shown in Fig. 1a, c. Phylogenetic analyses confirmed an extremely high haplotype diversity and the prevalence in Turano-Mongolian cattle of haplogroups T3 (~60.8%), T4 (16.5%), and T2 (12.0%) (Table 1). Haplogroup T1, which is dominant in Africa, was also observed in Kazakh (1/8), Yanbian (3/7), and Mishima cattle (1/8). The *B. indicus* I haplogroup has low frequency and was found in Tibetan (3/67), Kazakh (1/8), Mongolian (1/30), and Chaidamu cattle (2/5). Interestingly, haplogroup P was observed in Yanbian (1/7) and Korean cattle (1/28), while another rare haplogroup Q was found exclusively in Tibetan cattle (3/67).

### Relationships of the Turano-Mongolian mitogenomes with those from other areas

To further evaluate the phylogenetic relationships of the Turano-Mongolian cattle, phylogenetic trees were inferred by ML (Figs. 1b and S1a) and BI approach (Fig. S1b) with 370 available mitogenomes (Table S2). The two trees showed similar topologies, except for two Tibetans (RKZ5 and RKZ6), which exhibited unclear phylogenetic clusters within the T branch (Fig. S1). MP tree (Fig. S2) suggested that RKZ5 and RKZ6 could not be ascribed to any of the known T haplogroup (T1/T2/T3/T4) and was separated from T by six mutations (4469-10349-16057-16058-16133-16277A). The median network also supported the result of MP tree (Fig. 1c, H76).

Our analysis confirms a predominance of haplogroup T3 (~43%) among domestic taurine cattle all over the world. The second most frequent haplogroup is T1 (~29%), which is dominant in Africa. Five Turano-Mongolian individuals

**Table 1** Genetic structure and diversity of Turano-Mongolian breeds.

Breed	N	Haplogroup								S	h	k	Hd	Pi
		T1	T2	T3	T4	Q	P	I	Yak					
Yakutian	5		2	3						32	4	15.6	0.900	0.001
Mongolian <sup>a</sup>	30		7	12	6			1	4	205	7	45.7	0.941	0.0028
Kazakh	8	1	3	3				1		273	6	72.8	0.893	0.0045
Chaidamu	5			3					2	251	3	149.6	0.800	0.0092
Anxi	5		3	2						32	3	17.6	0.800	0.0011
Yanbian	7	3			3			1		85	5	29.3	0.905	0.0018
Tibetan <sup>a, b</sup>	67		1	49	3	3		3	8	203	9	44.3	0.970	0.0027
Korean <sup>a</sup>	28		3	20	4			1		109	10	23.5	1.000	0.0014
Mishima	8	1			7					22	3	5.6	0.464	0.0004
Japanese Black	7			4	3					23	7	8.6	1.000	0.0005
Total	170	5	19	96	26	3	2	7	12	588	108	35.7	0.994	0.0022

N = sample size, S = number of variable sites, h = number of haplotypes, k = the average number of differences, Hd = haplotype diversity, Pi = nucleotide diversity.

<sup>a</sup>Indicates that the estimated genetic diversity of the population has been averaged over panels of ten individuals.

<sup>b</sup>Means that we classified RKZ5 and RKZ6 into T3 haplogroup.

(three Yanbian, one Kazakh, and one Mishima) carry T1, but apparently not in the main lineage (Figs. 1b, c, S1, and Table S4). In a previous study, the Asian sub-lineage was denoted as T1f, which lacks the diagnostic mutation at position 16,113 (Bonfiglio et al. 2002). So far, only three European and one Egyptian cattle have been reported to carry T1f (Bonfiglio et al. 2002). Haplogroup T2 has a narrower range with a presence in Europe and Asia, whereas the presence of haplogroup T4 is restricted to Asia and is present in most East Asian breeds. T4 was not detected in our Yakutian cattle, but Kantanen et al. (2009) observed this lineage in the same breed with a frequency of 5/24 (Kantanen et al. 2009). The rare haplogroup Q is currently found in cattle from Italy (16 animals), Egypt (2), and Tibet (3) (Figs. 1b, S1, and S3). Finally, haplogroup P was identified in two Korean and one Yanbian cattle.

In addition to the haplogroup T4, we found two Turano-Mongolian-specific sub-lineages within T3 (Figs. 1b, 1c, S1, and S2), here denoted as T3<sub>119</sub> and T3<sub>055</sub>. T3<sub>119</sub> was initially termed T3<sub>119C</sub> (Cai et al. 2014) and has one diagnostic mutations at position 16,119 (with respect to the reference sequence V00654.1) and comprises 42 sequences mainly from Asia, including Tibetan (n = 26), Korean (n = 9), Japanese Black (n = 2), Mongolian (n = 2), and Yakut (n = 1) cattle. Only two individuals from Europe (EU177828 and AY676857) belong to T3<sub>119</sub>. Noting that except for five Korean cattle, most of the Turano-Mongolian cattle of the T3<sub>119</sub> lineage also share a mutation on site 2232 (Figs. 1b, c and S2). The second Turano-Mongolian-specific lineage T3<sub>055</sub> is defined by two mutations at positions 12,908 and 16,055 and was found in 12 samples, one from southwest Asia (EU177837) and 11

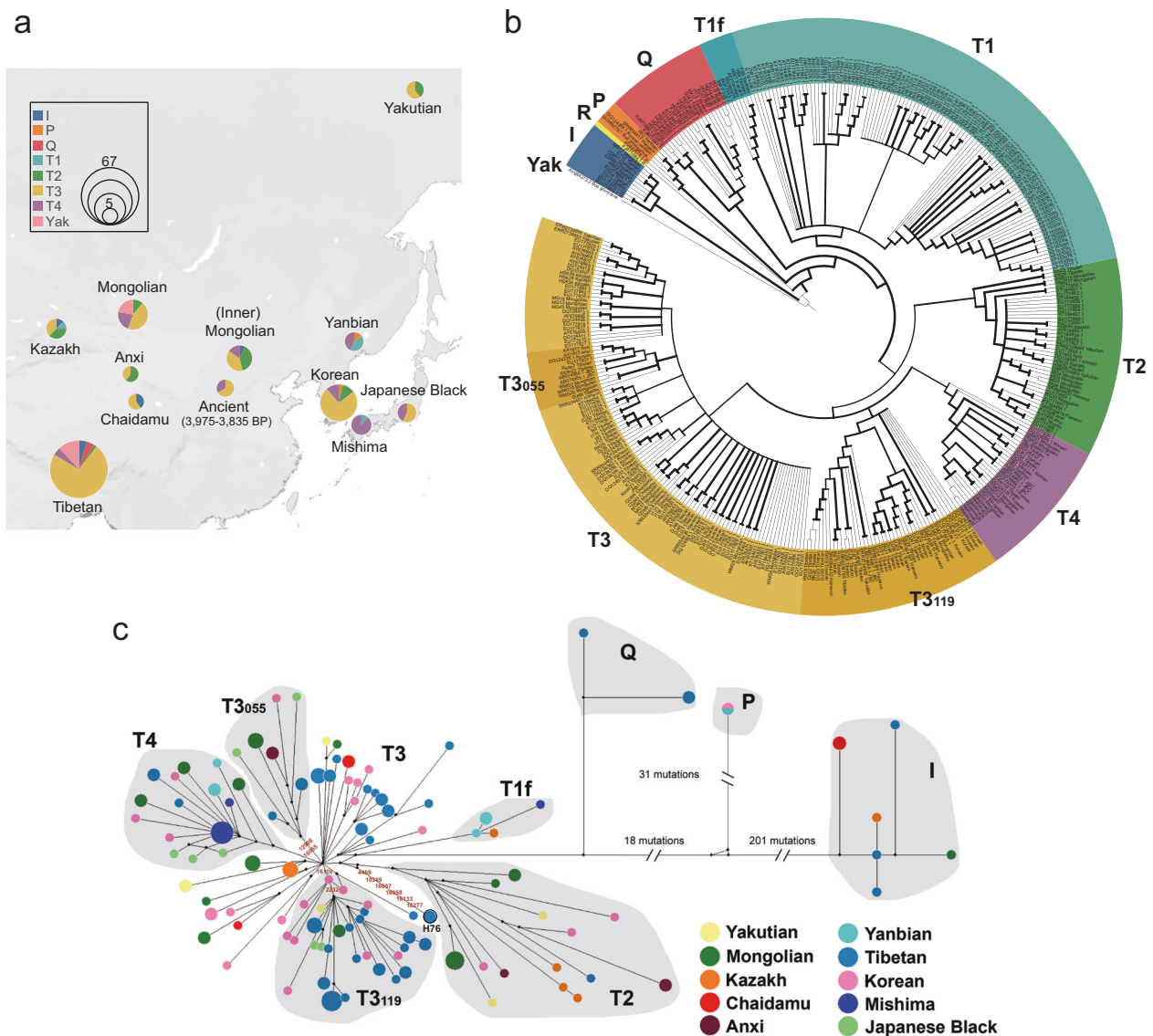
from East Asia (three from Tibet, two from Anxi, three from Mongolia, one from Korea, and two from Japan).

### Ancient mtDNA variation

To investigate the relationship between modern and ancient East Asian cattle, we aligned our sequences to six ancient sequences, which were collected in the late Neolithic Shima site (3975–3835-yr BP) in northern China (Chen et al. 2018). All six samples were classified into Turano-Mongolian-specific sub-lineage: three T3<sub>119</sub>, one T3<sub>055</sub>, and two T4 (Table 2 and Fig. S2). It is worth noting that these sub-lineages were previously mentioned (T4 and T3<sub>119</sub>) or labeled in table (T3<sub>055</sub>) by Cai et al. (2014) in ancient domestic cattle based on mtDNA partial regions, with T3<sub>119</sub> being the dominant lineage. Thus, our results indicated that T4, T3<sub>119</sub>, and T3<sub>055</sub> might represent the earliest arrival of domestic taurine cattle that entered north China at least ~3900-yr BP and have been retained until the present day.

### Age estimates and population expansion of Turano-Mongolian cattle

To assess population expansions that might have involved the Turano-Mongolian cattle, BSP was obtained (Fig. 2a). The overall BSP points to a steep increase of the female effective population size about ten to eight thousand years ago (kya), which corresponds to the early Neolithic period. Concerning the differentiation of the major branches in the Bayesian phylogenetic tree, we found that the Turano-Mongolian-specific T3 sub-lineages possibly emerged



**Fig. 1** Spatial frequency distribution and phylogenetic analysis of Turano-Mongolian cattle mtDNA haplogroups. **a** The geographical distribution of mtDNA haplogroups in the Turano-Mongolian cattle breeds. **b** A maximum likelihood (ML) phylogenetic tree of 370 mitogenomes, including 134 assembled in this study, 235 obtained from NCBI and one *Bos grunniens* (accession no. AY684273) as outgroup. The thickness of the line indicates the height of the

bootstrap value. The GenBank accession numbers for the haplogroup references are shown in labels. **c** Median-joining network constructed from 158 modern Turano-Mongolian sequences, including 134 assembled in this study and 24 downloaded from NCBI. The areas of the circles are proportional to haplotype frequencies. Black points are intermediate and unsampled sequences. The haplotype of RKZ5 and RKZ6 is labeled H76.

around ~12.49 (T3<sub>119</sub>, 4.98–34.86 kya) to 11.00 kya (T3<sub>055</sub>, 4.01–34.68 kya) (Table S5 and Fig. 2b). The most recent common ancestor (TMRCA) of the Q haplogroup detected in Tibetan cattle corresponded to 11.04 kya, which was lower than that of the entire Q1 (15.5 kya ± 3.5 ky) (Bonfiglio et al. 2010) but higher than the sister lineage Q1a reported in Italian cattle (5.8 ± 9.0 ky) (Olivieri et al. 2015) (Fig. S3). The TMRCA of T1f found in Northeast Asian cattle dates to around 9.55 kya, which was similar with that reported in Europe and African cattle (11.4 ± 2.4 ky) (Bonfiglio et al. 2002). Given the possibility that purifying

selection might slightly inflated the estimated age of haplogroup nodes, the real divergence time might be later (Soares et al. 2009). The details of the tree with the mean ages (nodes) and the 95% credibility intervals are shown in Table S5.

## Discussion

Phylogenetic analyses of 158 taurine and indicine mitogenomes confirmed the dominance of *B. taurus* haplogroups

**Table 2** Variable positions in mitogenome sequences of ancient samples.

Sample	Variable positions <sup>a</sup>	Haplogroup
	1 1	
	2 2 2 3 5 5 6 9 9 9 1 2 2 2 2 2 2 3 4 4 4 4 5 6 6 6 6 6 6 6 6	
	1 3 3 2 5 9 1 8 9 3 3 6 7 1 1 4 5 8 8 9 3 6 6 6 9 5 0 0 0 0 1 1 3 3	
	6 6 7 3 3 5 1 8 0 2 3 8 2 7 5 0 8 8 9 0 1 8 8 8 9 1 2 4 5 9 1 4 0 0	
	9 3 4 2 6 2 9 7 5 9 6 2 9 4 8 4 9 2 2 8 0 4 7 9 0 0 2 2 5 3 9 1 1 2	
V00654.1	A C C G C T T C C C C G G C T C C C C C A C C C T C G T T G T T C G	
Ancient05	G G . . A C . . . . C . . C . . . . C . . . . T . C . A . . . A	T4
Ancient01	G G T . A . . - - - C . T C T . T T . C - - - . - . C . A . C . A	T4
Ancient04	G . . . A . . . . . C . . . . . T C . . . . . C . . . . .	T3 <sub>055</sub>
Ancient07	G . . A A . . . . . C A . . . . . C . . . . . A . . . C . T .	T3 <sub>119</sub>
Ancient08	G . . A A . . . . . C A . . . T . . . C . . . C . . . . C . . .	T3 <sub>119</sub>
Ancient02	G . . A A . C . . . T C - . . . . . C . . . . . C . . . . C . . .	T3 <sub>119</sub>

Sequence codes are given in the first column, and only variable sites are shown. Dots denote identity with the reference sequence. Missing sequence is denoted by short lines.

<sup>a</sup>The variable positions were aligned to the taurine reference haplotype sequence (accession number V00654.1).

T3, T4, and T2 in Turano-Mongolian breeds (Mannen et al. 2004; Lenstra et al. 2014; Gao et al. 2017; Cai et al. 2018; Chen et al. 2018). We identified two Turano-Mongolian-specific sub-lineages in T3 (T3<sub>119</sub> and T3<sub>055</sub>), which were also present frequently in Neolithic Shimao samples (3975–3835-yr BP). T3<sub>119</sub> and T3<sub>055</sub>, together with T4, have a wide distribution in Asian modern and ancient cattle, which probably spread eastward after domestication with human activities and reached northern China at least ~3900-yr BP. Japanese Black (100.0%) showed the highest frequency of Turano-Mongolian-specific sub-lineages (T3<sub>119</sub>, T3<sub>055</sub>, and T4), followed by Mishima (87.5%), Mongolian (52.9%), Korean (50.0%), Tibetan (47.8%), Yanbian (42.8%), Anxi (40.0%), Yakutian (20.0%), and Inner Mongolian cattle (15.4%) (Fig. S4). Among them, most of the cattle breeds, previously classified as East Asian taurine cattle on the autosomal genome, are reported genetically similar to the late Neolithic cattle from Shimao site (Chen et al. 2018). In terms of maternal pattern, our results also support a close relationship between East Asian taurine and ancient Shimao cattle. Even if a samples bias cannot be completely ruled out, the lack of other taurine lineages suggests that T2 or other haplotypes belonging to T3 in the present-day cattle were likely introduced later, gradually replacing the original haplogroups. For example, for Kazakh and Chaidamu, the frequency of Turano-Mongolian-specific sub-lineages is zero. However, there are no Chinese ancient mtDNA samples available that are younger than 3975 years ago, so we do not know when the shift of the (sub-)haplogroup composition occurred. In addition, Tibetan cattle share more T3<sub>119</sub> haplotypes (38.8%) than other Turano-Mongolian breeds (Fig. S4), which is also common in ancient Chinese cattle (3/6).

This indicates that female cattle carrying the T3<sub>119</sub> sub-lineage successfully spread onto Tibet Plateau after 3900-yr BP. Due to the natural geographical barrier of the Tibetan Plateau, these original haplotypes have been retained until present days.

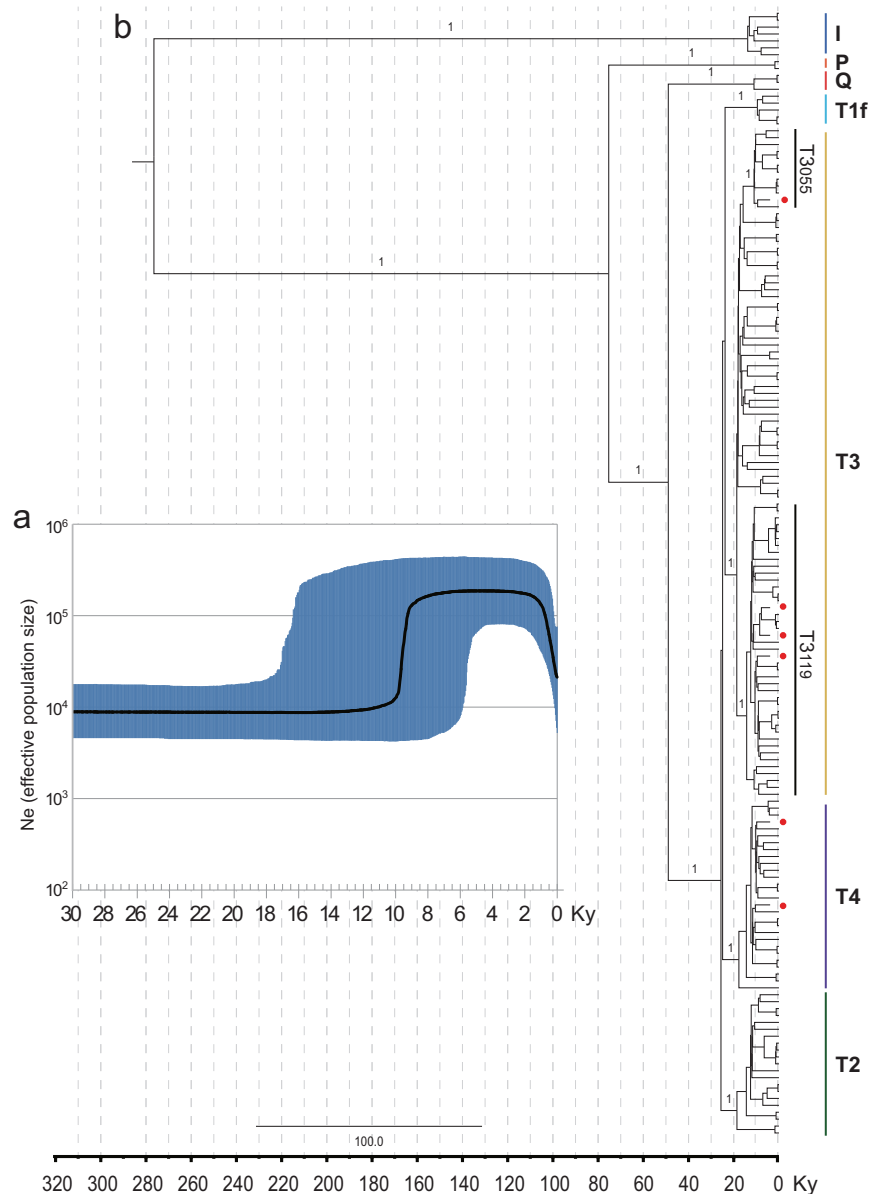
Four Mongolian and eight Tibetan cattle carry yak mtDNAs, which is consistent with previous studies (Medugorac et al. 2017; Wu et al. 2018) and reflects the coexistence of domestic cattle and yaks in Mongolia and on the Tibetan Plateau. Indicine mtDNAs were detected in Mongolian, Chaidamu, Kazakh, and Tibetan cattle at a low frequency. Previous studies have indicated that the secondary introgression of *B. indicus* in Mongolian and Kazakh breeds may have occurred during the period of Silk Road in ancient China (2nd to 7th centuries AD), which connected northwest China with India (Mannen et al. 2004; Cai et al. 2007; Yue et al. 2014).

We also found that one Yanbian and one Korean cattle carry haplogroup P, which was the predominant haplotype of the European aurochs. It has so far been detected only in modern cattle from Japan, Korea, and China, with a single substitution at position 16,247 (V00654.1) relative to the aurochs sequence (Noda et al. 2018). In this study, haplogroup P showed a more recent coalescence time (~1.79 kya), but its estimate might be imprecise, due to the small size (only two individuals with the same haplotype, Fig. 1c). Our phylogenetic analysis also showed that P haplotypes from modern Asian cattle and aurochs are separated with 100% high bootstrap values (Figs. 1b and S1), which may indicate that Asian aurochs has contributed to the gene pool of domestic cattle in Northeast Asia. Most recently, Mannen et al. (2020) proposed that the migration route of haplogroup P in northeast Asian cattle might be

**Fig. 2 Estimation of effective population size and divergence time of Turano-Mongolian cattle.**

**a** Bayesian skyline plot showing the Turano-Mongolian population size trend. The Y-axis indicates the effective number of females. A generation time of 6 years was considered (Bollongino et al. 2002). The black solid line is the median estimate and the blue shading shows the 95% limits of the highest posterior density.

**b** Schematic phylogeny of mtDNA coding region from 158 Turano-Mongolian modern and 6 ancient cattle. This tree was rooted by a published yak sequence (AY684273). The topology was inferred by ML approach based on the complete sequences, while the BI time divergence scale (Ky) based on coding region sequences is shown on the bottom. Posterior probabilities are marked on the branches. The red dots indicate ancient DNA samples.



from Mongolia/Russia eastward and eventually reached Japan. The Yanbian and Korean cattle carrying haplotype P found in our study were located on this migration route, probably further confirming this view.

Interestingly, three Yanbian, one Kazakh, and one Mishima cattle belong to sub-haplogroup T1f within haplogroup T1, which is common in African cattle. Outside Asia, T1f has only been found in three European and one Egyptian cattle (Bonfiglio et al. 2002), indicating that T1 independently spread to Europe/Africa and Asia and that in Asia only T1f survived.

Previously, Q has been reported for Italian and Egyptian breeds (Achilli et al. 2008; Achilli et al. 2009; Bonfiglio et al. 2010; Olivieri et al. 2015), which support the domestication of Q in southwest Asia together with haplogroup T

(Bonfiglio et al. 2010; Olivieri et al. 2015), but at a lower frequency. In our study, three out of sixty-seven Tibetan mtDNA belong to the rare haplogroup Q. Their coalescence age (~11.04 ky) was close to the Near Eastern domestication about 10 kya, further supporting the southwest Asia origin of the Q haplogroup. Our observations indicate that female cattle carried haplogroup Q together with other three T sub-haplogroups (T3<sub>119</sub>, T3<sub>055</sub>, and T4) spread eastward and reached high frequencies on Tibetan plateau due to the geographical isolation. Alternatively, Tibetan cattle might have been influenced by the local Asian aurochs.

Based on the above results, we suggest that these specific lineages are the legacy of the introduction of domestic cattle in Asia and still contribute significantly to the gene pool of Asian modern cattle.

## Conclusions

To date, mtDNA has been a very useful genetic marker for tracing the maternal origin of population. We present a comprehensive mitogenome analysis in modern and late Neolithic Turano-Mongolian cattle. Haplotype distributions support the southwest Asian origins of Turano-Mongolian cattle. We identified two specific lineages, T3<sub>119</sub> and T3<sub>055</sub>, in Turano-Mongolian cattle, which are also shared with Neolithic Shimaó samples (~3900 BP), indicating that they were present among the earliest domesticated cattle arrived in East Asia. Our results demonstrate the uniqueness of East Asian taurine cattle from their maternal origin and provide a theoretical basis for the conservation of genetic resources of these breeds.

## Data availability

All novel sequences were deposited in GenBank under the accession numbers MT576705–MT576844. Previously published 236 mitogenomes of Asian, European, African, American, West Asian cattle, aurochs, and yak were downloaded from the NCBI database (Table S2).

**Acknowledgements** This work was supported by National Natural Science Foundation of China (31872317), the Program of National Beef Cattle and Yak Industrial Technology System (CARS-37), the Central Plains Technological Innovation Leading Talents Project of Henan Province (194200510022), and the Major Projects of the National Social Science Foundation of China (17ZDA221 and 18ZDA218). Finally, we thank High-Performance Computing (HPC) of Northwest A&F University (NWFU) for providing computing resources.

## Compliance with ethical standards

**Conflict of interest** The authors declare no competing interests.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

- Achilli A, Bonfiglio S, Olivieri A, Malusà A, Pala M, Hooshiar Kashani B et al. (2009) The multifaceted origin of taurine cattle reflected by the mitochondrial genome. *Plos ONE* 4:e5753
- Achilli A, Olivieri A, Pellecchia M, Ubaldi C, Colli L, Al-Zahery N et al. (2008) Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Curr Biol* 18:R157–158
- Achilli A, Olivieri A, Soares P, Lancioni H, Hooshiar Kashani B, Perego UA et al. (2002) Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proc Natl Acad Sci U S A* 109:2449–2454
- Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48
- Bollongino R, Burger J, Powell A, Mashkour M, Vigne JD, Thomas MG (2002) Modern taurine cattle descended from small number of Near-eastern founders. *Mol Biol Evol* 29:2101–2104
- Bonfiglio S, Achilli A, Olivieri A, Negrini R, Colli L, Liotta L et al. (2010) The enigmatic origin of bovine mtDNA haplogroup R: sporadic interbreeding or an independent event of *Bos primigenius* domestication in Italy? *Plos ONE* 5:e15760
- Bonfiglio S, Ginja C, Gaetano AD, Achilli A, Olivieri A, Colli L et al. (2002) Origin and spread of *Bos taurus*: new clues from mitochondrial genomes belonging to haplogroup T1. *Plos ONE* 7:e38601
- Briggs AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U et al. (2009) Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* 325:318
- Cai D, Sun Y, Tang Z, Hu S, Li W, Zhao X et al. (2014) The origins of Chinese domestic cattle as revealed by ancient DNA analysis. *J Archaeol Sci* 41:423–434
- Cai X, Chen H, Lei C, Wang S, Xue K, Zhang B (2007) mtDNA Diversity and genetic lineages of eighteen cattle breeds from *Bos taurus* and *Bos indicus* in China. *Genetica* 131:175–183
- Cai Y, Jiao T, Lei Z, Liu L, Zhao S (2018) Maternal genetic and phylogenetic characteristics of domesticated cattle in north-western China. *Plos ONE* 13:e0209645
- Chen N, Cai Y, Chen Q, Li R, Wang K, Huang Y et al. (2018) Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nature Communications* 9:2337
- Drummond AJ, Suchard MA, Xie D, Rambaut A (2002) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29:1969–1973
- Eltsov NP, Volodko, NV (2011) MtPhyl: software tool for human mtDNA analysis and phylogeny reconstruction. <http://eltsov.org>
- Felius M (1995) Cattle breeds: an encyclopedia. Misset Uitgeverij, Doetinchem
- Gao Y, Gautier M, Ding X, Zhang H, Wang Y et al. (2017) Species composition and environmental adaptation of indigenous Chinese cattle. *Sci Rep* 7:16196
- Granberg L, Soini K, Kantanen J (2009) Sakha Ynaga-Cattle of the Yakuts. *Annales Academiae Scientiarum Fennicae. Humaniora* no. 355. Gummerrus, Vaajakoski, Finland
- Helmer D, Gourichon L, Monchot H, Peters J, Seguí MS (2005) Identifying early domestic cattle from pre-pottery neolithic sites on the Middle Euphrates using sexual dimorphism. In: Vigne J-D, Peters J, Helmer D (eds). *The first steps of animal domestication*. Oxbow Books, Oxford, UK, p 86–95
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589
- Kantanen J, Edwards CJ, Bradley DG, Viinalass H, Thessler S, Ivanova Z et al. (2009) Maternal and paternal genealogy of Eurasian taurine cattle (*Bos taurus*). *Heredity* 103:404–415
- Lee HJ, Kim J, Lee T, Son JK, Yoon HB, Baek KS et al. (2014) Deciphering the genetic blueprint behind Holstein milk proteins and production. *Genome Biol Evol* 6:1366–1374
- Lenstra JA, Ajmone-Marsan P, Beja-Pereira A, Bollongino R, Bradley DG, Colli L et al. (2014) Meta-analysis of mitochondrial DNA reveals several population bottlenecks during worldwide migrations of cattle. *Diversity* 6:178–187
- Letunic I, Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 47:W256–W259
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452
- Loftus RT, Machugh DE, Bradley DG, Sharp PM, Cunningham P (1994) Evidence for two independent domestications of cattle. *Proc Natl Acad Sci USA* 91:2757–2761



- Lorenzo PD, Lancioni H, Ceccobelli S, Curcio L, Panella F, Lasagna E (2016) Uniparental genetic systems: a male and a female perspective in the domestic cattle origin and evolution. *Electron J Biotechnol* 23:69–78
- Mannen H, Kohno M, Nagata Y, Tsuji S, Bradley DG, Yeo JS et al. (2004) Independent mitochondrial origin and historical genetic differentiation in North Eastern Asian cattle. *Mol Phylogenet Evol* 32:539–544
- Mannen H, Yonezawa T, Murata K, Noda A, Kawaguchi F, Sasazaki S et al. (2020) Cattle mitogenome variation reveals a post-glacial expansion of haplogroup P and an early incorporation into northeast Asian domestic herds. *Sci Rep* 10(1):20842
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303
- Medugorac I, Graf A, Grohs C, Rothammer S, Zagdsuren Y, Gladyr E et al. (2017) Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. *Nat Genet* 49:470–475
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274
- Noda A, Yonesaka R, Sasazaki S, Mannen H (2018) The mtDNA haplogroup P of modern Asian cattle: a genetic legacy of Asian aurochs? *Plos One* 13:e0190937
- Olivieri A, Gandini F, Achilli A, Fichera A, Rizzi E, Bonfiglio S et al. (2015) Mitogenomes from Egyptian cattle breeds: new clues on the origin of haplogroup Q and the early spread of *Bos taurus* from the Near East. *Plos ONE* 10:e0141170
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574
- Sambrook J, Russel D (2001) *Molecular cloning: a laboratory manual*, 3rd edn
- Schubert M, Lindgreen S, Orlando L (2016) AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes* 9:88–88
- Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A et al. (2009) Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 84:740–759
- Troy CS, Machugh DE, Bailey JF, Magee DA, Loftus RT, Cunningham P et al. (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature* 410:1088–1091
- Tsuda K, Kawahara-Miki R, Sano S, Imai M, Noguchi T, Inayoshi Y et al. (2013) Abundant sequence divergence in the native Japanese cattle Mishima-Ushi (*Bos taurus*) detected using whole-genome sequencing. *Genomics* 102:372–378
- Thorvaldsdóttir H, Robinson JT, Mesirov JP (2012) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14(2):178–192
- Weldenegodguad M, Popov R, Pokharel K, Ammosov I, Ming Y, Ivanova Z et al. (2018) Whole-genome sequencing of three native cattle breeds originating from the northernmost cattle farming regions. *Front Genet* 9:728
- Wu DD, Ding XD, Wang S, Wójcik JM, Zhang Y, Tokarska M et al. (2018) Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nat. Ecol Evol* 2:1139–1145
- Xia X, Qu K, Zhang G, Jia Y, Ma Z, Zhao X et al. (2019) Comprehensive analysis of the mitochondrial DNA diversity in Chinese cattle. *Anim Genet* 50:70–73
- Yue X, Li R, Liu L, Zhang Y, Huang J, Chang Z et al. (2014) When and how did *Bos indicus* introgress into Mongolian cattle? *Gene* 537:214–219
- Zhang Y (2011) *Animal genetic resources in China-Bovines* (in Chinese). China Agriculture Press, Beijing, China