



ORIGINAL  
ARTICLE



# Phylogeographical analyses of domestic and wild yaks based on mitochondrial DNA: new data and reappraisal

Zhaofeng Wang<sup>1†</sup>, Xin Shen<sup>2†</sup>, Bin Liu<sup>2†</sup>, Jianping Su<sup>3†</sup>,  
Takahiro Yonezawa<sup>4</sup>, Yun Yu<sup>2</sup>, Songchang Guo<sup>3</sup>, Simon Y. W. Ho<sup>5</sup>,  
Carles Vilà<sup>6</sup>, Masami Hasegawa<sup>4</sup> and Jianquan Liu<sup>1\*</sup>

<sup>1</sup>Molecular Ecology Group, Key Laboratory of Arid and Grassland Ecology, School of Life Science, Lanzhou University, Lanzhou 730000, Gansu, China, <sup>2</sup>Beijing Genomics Institute, Chinese Academy of Sciences, Beijing 101300, China, <sup>3</sup>Key Laboratory of Evolution and Adaptation of Plateau Biota, Northwest Plateau Institute of Biology, Chinese Academy of Sciences, Xining 810001, Qinghai, China, <sup>4</sup>School of Life Sciences, Fudan University, Shanghai 200433, China, <sup>5</sup>Centre for Macroevolution and Macroecology, Research School of Biology, Australian National University, Canberra, ACT 0200, Australia, <sup>6</sup>Department of Evolutionary Biology, Uppsala University, 75236 Uppsala, Sweden; Estación Biológica de Doñana-CSIC, Avd. María Luisa s/n, 41013 Sevilla, Spain

## ABSTRACT

**Aim** We aimed to examine the phylogeographical structure and demographic history of domestic and wild yaks (*Bos grunniens*) based on a wide range of samples and complete mitochondrial genomic sequences.

**Location** The Qinghai-Tibetan Plateau (QTP) of western China.

**Methods** All available D-loop sequences for 405 domesticated yaks and 47 wild yaks were examined, including new sequences from 96 domestic and 34 wild yaks. We further sequenced the complete mitochondrial genomes of 48 domesticated and 21 wild yaks. Phylogeographical analyses were performed using the mitochondrial D-loop and the total genome datasets.

**Results** We recovered a total of 123 haplotypes based on the D-loop sequences in wild and domestic yaks. Phylogenetic analyses of this dataset and the mitochondrial genome data suggested three well-supported and divergent lineages. Two lineages with six D-loop haplogroups were recovered for all morphological breeds of domestic yaks across their distributions in the QTP, while one more lineage and more endemic haplogroups or haplotypes were found for wild yaks. Based on the mitochondrial genome data, the divergences of the three lineages were estimated to have occurred around 420,000 and 580,000 years ago, consistent with the geological records of two large glaciation events experienced in the QTP.

**Main conclusions** There are distinct phylogeographical differences between wild and domestic yaks. However, there is no apparent geographical correlation between identified haplogroups and distributions of domestic yaks. Three differentiated lineages of yaks probably evolved allopatrically in different regions during the Pleistocene glaciation events, then reunited into a single gene pool during post-glacial population expansion and migrations before the start of the domestication of yaks in the Holocene.

## Keywords

*Bos grunniens*, China, demographic history, D-loop, domestication, mitochondrial genome, phylogenomics, phylogeographical structure, Qinghai-Tibetan Plateau.

\*Correspondence: Jianquan Liu, Key Laboratory of Arid and Grassland Ecology, Lanzhou University, Lanzhou, Gansu 730000, China.  
E-mail: liujq@nwiwb.ac.cn  
†These authors contributed equally to this work.

## INTRODUCTION

The domestication of animals made available stable sources of protein and enabled the development of human civilization (Mason, 1984; Diamond, 2002). Molecular and archaeological

studies indicate complex phylogeographical histories of the animals used for meat (e.g. Lau *et al.*, 1998; Luikart *et al.*, 2001; Troy *et al.*, 2001; Jansen *et al.*, 2002; Beja-Pereira *et al.*, 2004; Larson *et al.*, 2005; Pedrosa *et al.*, 2005; Meadows *et al.*, 2007; Naderi *et al.*, 2008). For example, distinct phylogeographical

patterns have been recovered for wild boars and domestic pigs, suggesting multiple centres of pig domestication (Larson *et al.*, 2005). In goats, the distributions of highly divergent lineages and haplogroups show weak phylogeographical structure because of extensive transportations across the world (Luikart *et al.*, 2001). It is usually suggested that such divergent intraspecific lineages have derived from multiple, independent domestication events associated with genetically discrete wild populations or species with a history of ancient divergence (e.g. Luikart *et al.*, 2001; Larson *et al.*, 2005; Pedrosa *et al.*, 2005).

Domesticated yaks (*Bos grunniens*) are distributed across the Qinghai-Tibetan Plateau (QTP) and adjacent highlands (Wiener *et al.*, 2003). The QTP is the largest continuous high-elevation ecosystem in the world, occupying nearly 2.5 million km<sup>2</sup> of the Asian continent and reaching an average elevation of more than 4000 m a.s.l. (Li *et al.*, 1995). More than 14 million yaks are currently herded there and they provide many of life's necessities (food, hides, dung fuel and transport power) for the nomadic Tibetan pastoralists living in this extremely harsh region (Wiener *et al.*, 2003). It has been suggested that these yaks were the product of local domestication of wild individuals (Qian, 1979); indeed, around 15,000 wild yaks currently inhabit northern Tibet and west Qinghai on the QTP (Schaller & Liu, 1996; Leslie & Schaller, 2009).

Genetic analyses based on mitochondrial DNA (mtDNA) D-loop sequences have suggested that two highly divergent lineages are present in both domesticated and wild yaks (Tu *et al.*, 2002; Qi *et al.*, 2005; Guo *et al.*, 2006; Lai *et al.*, 2007; Ho *et al.*, 2008). However, there are distinct geographical and morphological differences between wild and domestic yaks, the former being of larger size, having long hair, and occurring mainly in the northwestern QTP. Accordingly, they have been treated as separate species (e.g. Leslie & Schaller, 2009); here we treat them as a single species because of the absence of a reproductive barrier between them. Studies based on D-loop sequences, as well as those based on nuclear microsatellite alleles (Zhang *et al.*, 2008) and blood protein electrophoresis (Tu *et al.*, 1997), also suggest that identified genetic groups show weak geographical or morphological correlations within domestic yaks. Because of limited sampling in previous studies, especially for wild yaks (Qi *et al.*, 2005; Guo *et al.*, 2006; Lai *et al.*, 2007), it remains unknown whether additional differentiated maternal lineages or haplogroups occur between and within wild and domestic yaks. In addition, Guo *et al.* (2006) estimated that the divergence of the two major lineages occurred around 130,000–100,000 years ago, while Ho *et al.* (2008) suggested that this divergence occurred more recently, around 75,000 years ago, based on D-loop sequences. However, intraspecific divergences recorded in a few biota occurring on the QTP have been linked to extensive glaciations during the Pleistocene (e.g. Jin *et al.*, 2008; Wang *et al.*, 2009), for example, around 600,000 years ago, much earlier than the estimates for the divergence of the two lineages within the yaks.

In this study we used mtDNA sequence data to study the historical domestication, intraspecific diversification and

geographical distribution of lineages of yaks. MtDNA sequences have been widely used when studying the population genetics and phylogeographical histories of animals; their value lies in the occurrence of multiple copies per cell, rare recombination (Olivo *et al.*, 1983; Ingman *et al.*, 2000), maternal inheritance (Giles *et al.*, 1980; Gyllensten *et al.*, 1985), and high rates of change (Brown *et al.*, 1979). Most previous studies of the phylogeographical histories and domestication patterns of animals have been based on the mtDNA D-loop region (Luikart *et al.*, 2001; Troy *et al.*, 2001; Jansen *et al.*, 2002; Beja-Pereira *et al.*, 2004; Larson *et al.*, 2005; Guo *et al.*, 2006; Lai *et al.*, 2007). This region is highly variable and informative in determining intraspecific diversification, but it is also subject to frequent parallel mutations (Tamura & Nei, 1993; Ingman & Gyllensten, 2001) – it is a mutation 'hotspot' region (Wakeley, 1993). Complete mitochondrial genomes, containing 18 times the number of sites in the D-loop region, may allow us to recover finer details and reach more confident conclusions regarding the geologically recent evolution of animals, especially for molecular estimates of intraspecific, divergent lineages (e.g. Ingman & Gyllensten, 2001). Recently, a number of studies have adopted this phylogenomic approach to examine intraspecific diversification or to clarify the domestication history of animals (Wu *et al.*, 2007; Achilli *et al.*, 2008; Gilbert *et al.*, 2008).

The major goals of this study were: (1) to examine intraspecific genetic diversification and geographical distributions of genetic lineages of both wild and domesticated yaks based on sequences of part of the D-loop region and a larger sample of animals than previously available (Guo *et al.*, 2006); and (2) to clarify when and where major lineages diverged based on the complete mitochondrial sequences of the selected samples. This fine-scale phylogenomic analysis of matrilineal components of wild and domesticated yaks provides new insights into the intraspecific diversification and phylogeographical history of wild and domestic yaks.

## MATERIALS AND METHODS

### Sample collection

About 15,000 wild yaks occur only in the Kokohili region in the northern and northwestern QTP, where the arrival of herdsmen is rare (Leslie & Schaller, 2009). However, domestic yaks are widely distributed in other regions of the QTP (Wiener *et al.*, 2003). We collected genetic material from 96 domesticated yaks, including samples from two new breeds, and samples from a few locations where we had not previously collected (see previous study by Guo *et al.*, 2006). Therefore our domestic yaks represent all the morphological breeds and distributional range over the QTP.

We also collected another 34 wild samples in the Kokohili, including 15 wild individuals in one location where more than 100 wild yaks died of hunger in the snowy spring. The other wild samples were sparsely collected from corpses of wild yaks in the central Kokohili. (These were considered to be wild yaks

based on taxonomic characters: long hair and large skeleton.) We placed the domestic samples collected within a diameter of 200 km as a separate population. Similarly, the wild yaks were also treated as three separate populations because these samples were collected at a spacing of more than 200 km.

We sequenced the partial D-loop region of these samples. Together with our previous samples (Guo *et al.*, 2006) and other sequences from GenBank, we collated data on the D-loop sequences of a total of 405 domesticated and 47 wild yaks. Of the samples from which we had sequenced the D-loop region, 48 domestic yaks and 21 wild yaks were randomly selected for complete mitochondrial sequencing (see Appendix S1 in Supporting Information). In addition, three complete mtDNA sequences from domesticated yaks were downloaded from GenBank, resulting in a dataset comprising 51 domesticated yaks; this dataset represented most morphological breeds as well as the range of species across the QTP (Appendix S2). One cattle (*Bos indicus*) sample was also collected from Yunnan province. Muscle tissue and/or skin samples were collected and stored in 100% ethanol.

### DNA extraction, amplification and sequencing

Genomic DNA was extracted using the standard phenol-chloroform method (Sambrook & Russell, 2001). The partial D-loop region that has a length of 637 base pairs (bp) was amplified using the primers and methods described by Guo *et al.* (2006). For complete mitochondrial sequencing, mitochondrial DNA fragments were amplified using a long and accurate-polymerase chain reaction (LA-PCR) kit (Takara, Dalian, China). Six pairs of PCR primers were designed to amplify the complete sequence of the mitochondrial genome based on the conserved sequences of three species (cattle, sheep and pig) (Appendix S3). PCR amplifications were carried out in 50- $\mu$ L reaction mixtures containing 50 ng DNA, each primer (0.5  $\mu$ L of a 5  $\mu$ M solution), dNTPs (4  $\mu$ L of a 2.5 mM solution), 5  $\mu$ L of 10  $\times$  buffer and 0.2  $\mu$ L of 5 U  $\mu$ L<sup>-1</sup> LA-Taq DNA polymerase. The PCR programme, run in a Tpersonal Thermocycler (Biometra, Germany), consisted of an initial denaturing step at 95 °C for 3 min, 35 amplification cycles (95 °C for 20 s, 52 °C for 40 s and 72 °C for 4 min) and a final extension at 72 °C for 10 min. PCR products were purified using a Tiangen PCR purification kit following the manufacturer's instructions (Tiangen, Beijing, China). The purified PCR products were sequenced on an ABI 3130XL Sequencer (Applied Biosystems, Foster, CA, USA) using a BigDye Terminator Cycle Sequencing Kit (Applied Biosystems). Seventy-two primers were used for sequencing (Appendix S4) and overlapping contigs were compiled using the programs PHRED (Ewing & Green, 1998), PHRAP (Green, 2010) and CONSED (Gordon *et al.*, 1998), in order to assemble a continuous sequence.

### Phylogenetic analyses

The newly determined D-loop and complete mitochondrial sequences from yaks were lodged in GenBank (GQ464116–

GQ464314). Three complete yak mitochondrial sequences (EF494177–EF494179) and 322 D-loop sequences were downloaded from GenBank (DQ138998–DQ139260, AY521137–AY521161, DQ007210–DQ007224, DQ856594–DQ856600, DQ856602–DQ856604, EF494177–EF494179, AY722118, AY749414, AY684273, AY374125, AF485066 and AF083355). The complete mitochondrial sequence of bison (*Bison bison*) (EU177871) was also added to the final data matrix. All sequences were aligned using the CLUSTALX program (Thompson *et al.*, 1997). Two datasets were collated for phylogenetic analyses: yak D-loop haplotypes and yak–bison complete mitochondrial sequences. We used the sister species of yak, *Bison bison* (Kierstein *et al.*, 2004; Gu *et al.*, 2005) (U12936) as an outgroup to root all yak D-loop haplotypes. The cattle (GU256940) was further selected as an outgroup to analyse complete mitochondrial sequences (including both yak and bison) (Kierstein *et al.*, 2004; Gu *et al.*, 2005). For the complete mitochondrial dataset, a total of 12 protein genes were used; the ND6 gene was excluded because it was associated with significant bias in codon usage. The initiation and termination codons and overlapping regions between ATP6 and ATP8, ND4 and ND4L, and ND5 and ND6 were also excluded.

Each dataset was subjected to maximum likelihood (ML) analyses in PAUP\* 4.0b10 (Swofford, 2002), as well as Bayesian analyses in MRBAYES 3.1 (Huelsenbeck & Ronquist, 2001). MODELTEST (Posada & Crandall, 1998) was used to select parameters and assumptions for the ML analyses. The heuristic search parameters used for the PAUP\* analyses were the simple addition sequence of taxa with tree bisection–reconnection (TBR) branch swapping, MULTREES and COLLAPSE. We further used GARLI 0.96 (Zwickl, 2010) to obtain the support values for the nodes of ML tree. GARLI (Genetic Algorithm for Rapid Likelihood Inference) is a program that performs phylogenetic searches on aligned sequence datasets using the maximum-likelihood criterion. It uses a stochastic genetic algorithm-like approach to simultaneously find the topology, branch lengths and substitution model parameters that maximize the log-likelihood. The GTR + I + G model was used, and the bootstrap analyses were carried out with 1000 replicates. Posterior probabilities were calculated for nodes in the tree. For the Bayesian analyses, four simultaneous Markov chain Monte Carlo (MCMC) analyses were run for 10,000,000 steps, with one tree saved every 1000 steps and with the first 25% discarded as 'burn-in'. The standard deviation of split frequencies was below 0.01 after 10,000,000 steps, indicating the convergence of the four chains to the stationary distribution.

### Analysis of molecular diversity

The haplotype diversity ( $H_d$ ) and nucleotide diversity ( $\pi$ ) of all domesticated yaks and wild yaks were estimated separately for the D-loop region, the complete mtDNA sequences and the protein coding regions using ARLEQUIN 3.1 (Excoffier *et al.*, 2005). Fu's  $F_s$  test (Fu, 1997) was used to determine whether

or not each haplogroup and all samples exhibited neutrality. A population that has experienced expansion may result in a rejection of the null neutrality hypothesis with significant negative values. The genetic structure of domestic yaks and wild yaks was examined across the distribution regions by analysis of molecular variance (AMOVA, Excoffier *et al.*, 1992) as implemented in ARLEQUIN 3.1. NETWORK 4.1 (available from <http://www.fluxus-engineering.com/sharenet.htm>) was used to determine D-loop haplotype relationships among the wild and domesticated yaks (Bandelt *et al.*, 1999).

### Molecular calibration and divergence estimation

A likelihood-ratio (LR) test (Huelsenbeck & Rannala, 1997) in PAUP\* 4.0b10 (Swofford, 2002) was used to test the hypothesis of a molecular clock for the protein-coding genes of bison and yaks. This test was performed by comparing the log likelihood ( $\ln L$ ) of the ML trees with and without assuming a molecular clock. The hypothesis of a molecular clock could not be rejected because constrained and unconstrained analyses were not significantly different (GTR + G,  $2\Delta\ln L = 33.27$ , d.f. = 49,  $P = 0.96$ ).

Divergence times of three divergent lineages of yaks were estimated using the Bayesian method implemented in the program BEAST v. 1.5.3 (Drummond & Rambaut, 2007). To address the problem of time-dependent rates, whereby rates appear to be elevated over short time-scales owing to incomplete purifying selection (Ho *et al.*, 2005), our dating analysis was limited to the third codon sites in the alignment. This approach follows recent dating studies, which have suggested that rates at synonymous coding sites, for which third codon sites are taken as a proxy here, are subject to a much smaller degree of time-dependence than non-synonymous sites (Kivisild *et al.*, 2006; Endicott & Ho, 2008; Subramanian *et al.*, 2009). We compared five population models (constant size, exponential growth, logistic growth, expansion growth, and Bayesian skyline plot) using Bayes factors (Nylander *et al.*, 2004) (indicating the strength of the evidence in favour of the best of compared models), then estimated divergence times under the chosen model (constant size). Following a burn-in of 1,000,000 steps, all parameters

were sampled once every 1000 steps from 10,000,000 MCMC steps. We checked convergence of the chains to the stationary distribution by visual inspection of plotted posterior estimates using the program TRACER (Rambaut & Drummond, 2010), and the effective sample size for each parameter sampled from the MCMC analysis was found to exceed 100, usually by an order of magnitude.

To calibrate the divergence-time analysis, the divergence between bison and yak was set to be around 2.5 million years ago (Ma). This was done by specifying a lognormal prior on the age of the bison–yak divergence, with a minimum bound of 2 Ma, a mean of 2.5 Ma, and with 95% of the prior probability between 2 and 3 Ma (Ho & Phillips, 2009). We chose a mean value of 2.5 Ma rather than the value of 1.8 Ma adopted by Guo *et al.* (2006), for the following reasons. First, fossils of yaks were well established before 1.8 Ma in the QTP (Flerow, 1980). These fossil records suggested that origins of yaks should have been earlier than this time (Yonezawa *et al.*, 2007). In addition, the earliest bison-like fossil (which might be a yak fossil) was dated at around 2.5 Ma during the late Pliocene (Olsen, 1990; Tedford *et al.*, 1991). Finally, previous molecular estimates have placed the bison–yak divergence earlier than 2.5 Ma, at about 3.0 Ma (e.g. Kierstein *et al.*, 2004; Gu *et al.*, 2005).

## RESULTS

### MtDNA D-loop sequences

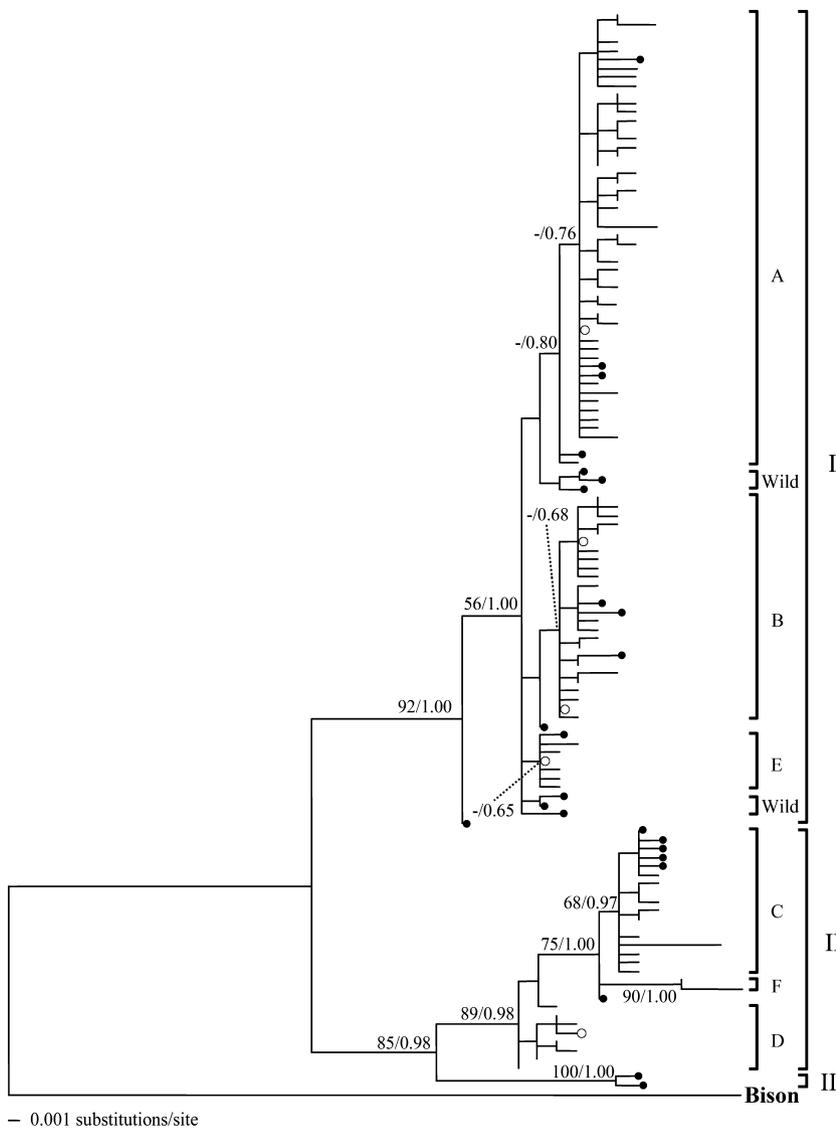
The final analyses used data from 405 domestic yaks (see Appendix S2 for locations) and 47 wild yaks from western China. A total of 87 variable sites and 123 haplotypes were identified from these (452) sequences (Table 1). We identified a total of 29 haplotypes from the wild samples, with five haplotypes shared by domesticated and wild yaks. The average haplotype diversity was clearly higher in wild than in domesticated yaks (Table 1).

Phylogenetic analyses of the 123 haplotypes revealed two highly supported divergent lineages (I and II) (Fig. 1). However, two haplotypes from the wild yaks formed a more divergent lineage, and its clustering with lineage II was poorly

**Table 1** Summary of statistical parameters for the D-loop region and mitochondrial genomes of yaks (*Bos grunniens*) in western China.

	Dataset	Length	V	N	h	Hd ± SD	π ± SD
D-loop	All yaks	636–637	87 (62)	452	123	0.927 ± 0.008	0.0165 ± 0.0084
	Domesticated yaks	636–637	79 (52)	405	99	0.916 ± 0.009	0.0156 ± 0.0079
	Wild yaks	636–637	49 (41)	47	29	0.967 ± 0.013	0.0221 ± 0.0112
Complete mtDNA	All yaks	16,321–16,325	306 (153)	72	64	0.995 ± 0.004	0.00315 ± 0.00153
	Domesticated yaks	16,321–163,25	187 (132)	51	45	0.992 ± 0.007	0.00294 ± 0.00144
	Wild yaks	16,322–163,23	235 (127)	21	19	0.991 ± 0.018	0.00365 ± 0.00183
Protein-coding regions	All yaks	10,710	193 (82)	72	50	0.944 ± 0.022	0.00262 ± 0.00129
	Domesticated yaks	10,710	105 (71)	51	34	0.904 ± 0.039	0.00238 ± 0.00118
	Wild yaks	10,710	148 (67)	21	17	0.981 ± 0.020	0.00323 ± 0.00164

V, number of variable sites; numbers in parentheses indicate number of parsimony-informative sites; N, number of sequences; h, number of haplotypes; Hd, haplotype diversity; SD, standard deviation; π, nucleotide diversity.



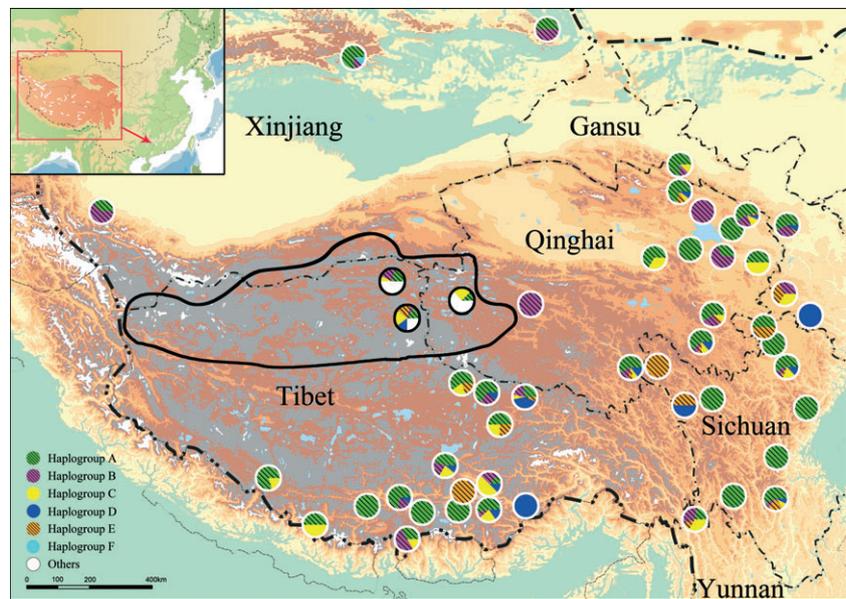
**Figure 1** Phylogenetic tree of all domesticated and wild yak (*Bos grunniens*) D-loop haplotypes in western China constructed by maximum likelihood analysis, rooted with one sequence from *Bison bison*. The length of the D-loop sequences is 637 bp. Filled circles indicate haplotypes found in wild yaks; open circles, haplotypes shared by domestic and wild yaks. Branches with no circle are those found in domestic yaks. Support is indicated at the nodes as percentage bootstrap support for 1000 maximum likelihood replicates and Bayesian posterior probabilities.

supported. All domesticated yak haplotypes nested within lineages I and II, within which six major haplogroups (A–F) were identified (Fig. 1). The wild haplotypes were evenly distributed throughout the entire tree, and about 60% of wild individuals had haplotypes belonging to the six domestic haplogroups (Fig. 2). However, there were more haplogroups or haplotypes specific to the wild yaks.

Geographical structure was weak among domestic yak populations, as revealed by phylogeographical analyses of the 405 D-loop sequences. There were also no correlations between phylogenetic lineages or haplogroups and the geographical distribution of domesticated yaks (Fig. 2). The two main lineages both consist of domestic yak haplotypes from Qinghai, Tibet, Gansu, Sichuan, Yunnan and Xinjiang. A total of 55.5% of domestic yaks had haplotypes belonging to haplogroup A, and 83.9% had types belonging to haplogroups A, B and C (Table 2). Haplogroups A and B were represented in all geographical regions, and haplogroup C in all regions except Xinjiang. Thus these three haplogroups constitute a common

source for a very large proportion of the mtDNA genetic variation in all domestic yak populations. Haplogroup D was found in all regions except Xinjiang and Yunnan, and haplogroup E was represented in all geographical regions except Xinjiang, while haplogroup F was found only regionally in Xinjiang. Furthermore, the frequencies of haplogroups A–E were similar in all regions except Yunnan due to the small sample size from that province (Table 2). There is also no clear correlation between the phylogenetic haplogroups and the 12 morphological breeds of domestic yak (Table 3). Most breeds contain haplotypes from the five major haplogroups (A–E).

The similarity of genetic variation among regions can also be seen from the haplotype network of most of the major haplogroups (Fig. 3). Each haplogroup harboured a few widely distributed major haplotypes, found in most geographical regions and at high frequencies, surrounded by less frequent types that were generally unique to a specific region. This indicated demographic expansions following domestication, further supported for four of the major haplogroups (A, B, C



**Figure 2** Geographical distribution of the yak (*Bos grunniens*) D-loop haplogroups in western China. These haplogroups are shown in Fig. 1. Populations of domesticated yaks are indicated by white circles; wild populations by black circles. The current range of wild yaks is indicated by a continuous line.

**Table 2** Numbers and proportions of individuals of yaks (*Bos grunniens*), and numbers of haplotypes and unique haplotypes representing haplogroups A–F in each geographical region in western China.

Region	N	h (U)	Hd ± SD	π ± SD	A	B	C	D	E	F
					N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Qinghai	180	52 (33)	0.892 ± 0.017	0.0136 ± 0.0012	107 (59.4)	29 (16.1)	18 (10.0)	14 (7.8)	12 (6.7)	0 (0)
Tibet	94	33 (19)	0.921 ± 0.017	0.0188 ± 0.0013	47 (50.0)	15 (16.0)	14 (14.9)	14 (14.9)	4 (4.2)	0 (0)
Gansu	36	17 (5)	0.914 ± 0.031	0.0124 ± 0.0025	20 (55.6)	8 (22.2)	2 (5.6)	3 (8.3)	3 (8.3)	0 (0)
Sichuan	45	22 (9)	0.939 ± 0.018	0.0156 ± 0.0023	25 (55.5)	3 (6.7)	5 (11.1)	5 (11.1)	7 (15.6)	0 (0)
Xinjiang	35	10 (6)	0.842 ± 0.034	0.0097 ± 0.0025	21 (60.0)	12 (34.3)	0 (0)	0 (0)	0 (0)	2 (5.7)
Yunnan	9	7 (2)	0.917 ± 0.092	0.0221 ± 0.0039	2 (22.2)	3 (33.3)	3 (33.3)	0 (0)	1 (11.2)	0 (0)
Domestic yaks	405	99	0.916 ± 0.009	0.0156 ± 0.0079	225 (55.5)	72 (17.8)	43 (10.6)	36 (8.9)	27 (6.7)	2 (0.5)

N, number of sequences; h, number of haplotypes; U, unique haplotypes; Hd, haplotype diversity; SD, standard deviation; π, nucleotide diversity.

**Table 3** The distribution of sampled individuals of 12 morphological breeds of yak (*Bos grunniens*) representing the six identified haplogroups (A–F) in western China.

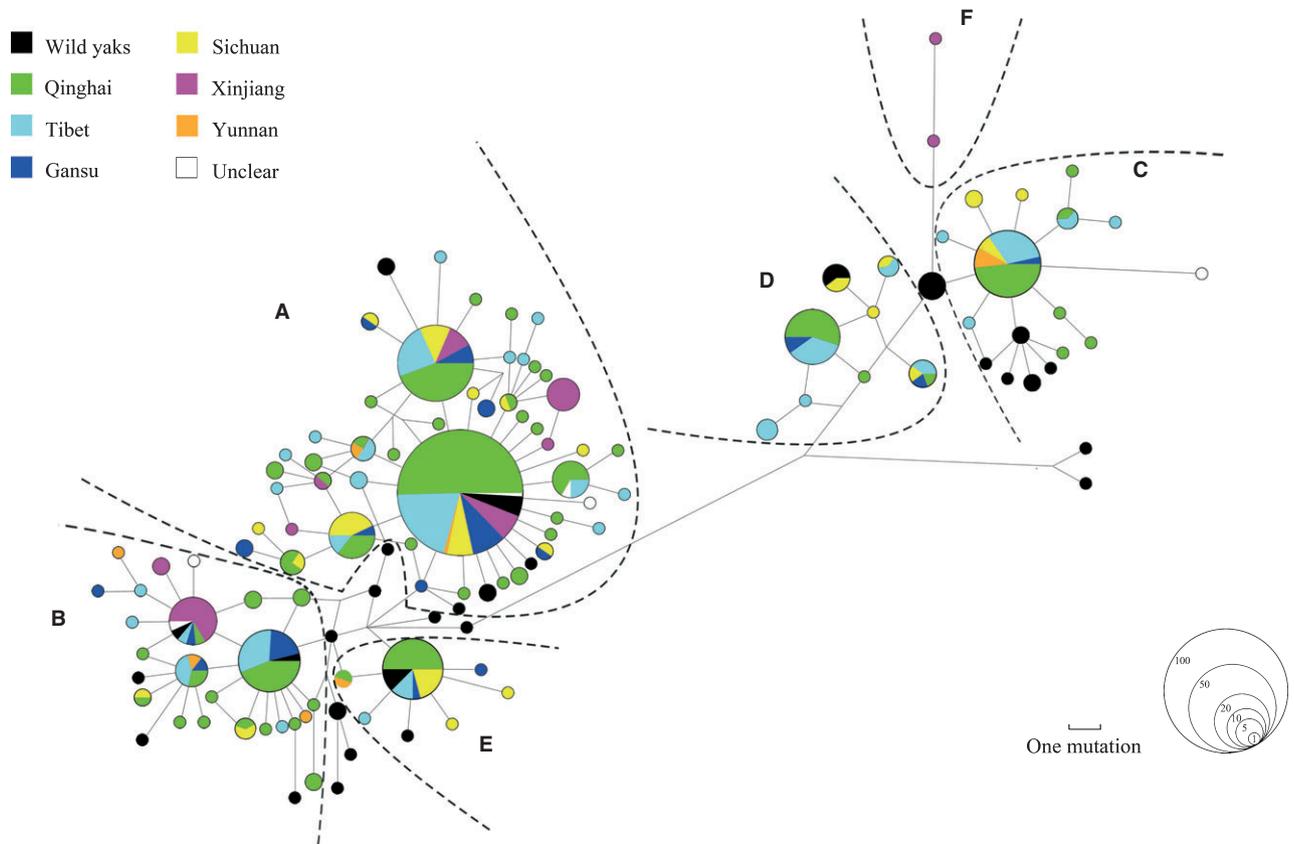
Breed	A	B	C	D	E	F
Tianzhu	12	6		2	1	
Gannan	2	1	1	1	2	
Datong	17	7	4	2	1	
Huanhu	38	7	5	5	6	
Plateau	52	15	9	7	5	
Jiali	19	7	4	9	3	
Sibu	5	2	2	3		
Pali	23	6	8	2	1	
Maiwa	14	2	4	3	1	
Jiulong	10	1	1	1	3	
Bazhou	21	12				2
Zhongdian	2	3	3		1	

and E) by Fu's  $F_S$  test, which is particularly sensitive to population growth (Fu, 1997). The test exhibited significant ( $P < 0.01$ ) departure from neutrality (Table 4).

Analysis of molecular variance (AMOVA) of the D-loop haplotypes suggested that 1.92% of the total variance was distributed between wild and domestic yaks (Table 5). Within domestic yaks, most of the total variances (93.91%) were attributed within populations and 5.46% among populations, with a very small percentage (0.64%) among regions (Table 5).

### Complete mitochondrial sequences

The 72 complete mitochondrial sequences from yaks varied in length from 16,321 to 16,325 bp. A total of 311 segregating sites were identified, of which 157 showed the same polymorphism in at least two individuals. When ND6 was excluded, the mitochondrial protein-coding genes comprised 10,710 bp (3570 codons). Based on this sequence length, we identified 34 haplotypes from the domesticated yaks and 17 from the wild individuals. Only one haplotype was shared by domesticated and wild yaks and therefore a total of 50 haplotypes were recovered. Genetic diversity was higher in wild than domesticated yaks.



**Figure 3** Parsimonious network of D-loop sequences of domesticated and wild yaks (*Bos grunniens*) in western China. Each haplotype is represented by a circle, with the area of the circle proportional to its frequency. Samples from different regions are indicated by different colours. The length of each branch is proportional to the number of mutations on the respective branch.

**Table 4** Genetic diversity and neutrality test for the major haplogroups of yak (*Bos grunniens*) based on the D-loop regions of all sampled yaks in western China.

Haplogroup	Including wild yaks				Excluding wild yaks			
	<i>N</i>	<i>Hd</i> ± SD	$\pi$ ± SD	<i>F<sub>S</sub></i> test	<i>N</i>	<i>Hd</i> ± SD	$\pi$ ± SD	<i>F<sub>S</sub></i> test
A	235	0.777 ± 0.025	0.00241 ± 0.00160	-27.900**	225	0.777 ± 0.026	0.00237 ± 0.00159	-27.894**
B	77	0.852 ± 0.031	0.00269 ± 0.00176	-24.839**	72	0.845 ± 0.032	0.00252 ± 0.00167	-20.652**
C	50	0.664 ± 0.077	0.00203 ± 0.00144	-14.785**	43	0.546 ± 0.092	0.00157 ± 0.00120	-8.952**
D	39	0.709 ± 0.069	0.00264 ± 0.00176	-1.428	36	0.671 ± 0.079	0.00251 ± 0.00169	-1.752
E	31	0.404 ± 0.111	0.00081 ± 0.00078	-5.398**	27	0.399 ± 0.117	0.00081 ± 0.00078	-4.004**
All	452	0.927 ± 0.008	0.0165 ± 0.0084	-23.830**	405	0.916 ± 0.009	0.0156 ± 0.0079	-23.931**

*N*, number of sequences; *Hd*, haplotype diversity; SD, standard deviation;  $\pi$ , nucleotide diversity; \*\**P* < 0.01.

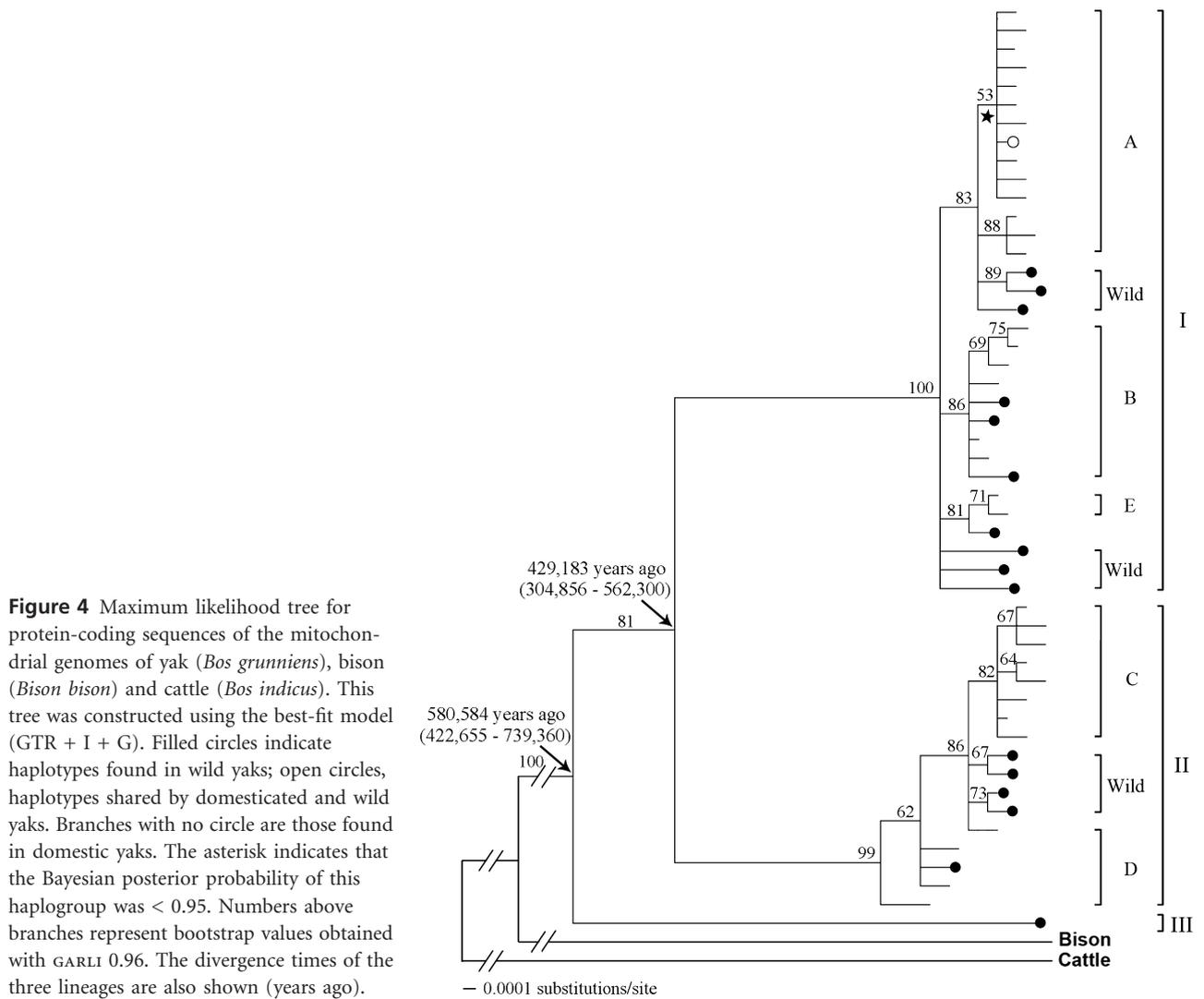
Phylogenetic analyses of these 50 haplotypes based on the best-fit model (GTR + I + G) identified three highly divergent lineages, with each lineage receiving high bootstrap-support values (Fig. 4). The other nodes of the ML tree received moderate to high support values. The Bayesian tree was consistent with the ML tree in its topology and all nodes received high posterior probabilities (> 0.95), with one exception (posterior probability = 0.82, marked with a star in Fig. 4). Most haplotypes clustered into the two lineages I and II, which contained haplotypes from both domesticated

and wild yaks. In these two lineages, five haplogroups (A–E) of domesticated yaks, which were in accordance with the D-loop phylogenetic tree, could be identified. We also identified more haplogroups or haplotypes specific to the wild yaks (Fig. 4). One wild haplotype (III) formed a sister lineage to the two main lineages. The D-loop haplotypes identified from this wild sequence were situated at the base of lineage II in the phylogenetic tree of the D-loop region (Fig. 1). The divergent, low-frequency mtDNA haplotypes may result from nuclear mitochondrial pseudogenes (numts) (Parr *et al.*, 2006).

**Table 5** Analyses of molecular variance (AMOVA) for haplotypes of yaks (*Bos grunniens*) in western China.

	Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation	Fixation index
Wild yaks	Among populations	2	1.260	0.01159	2.37	$F_{ST} = 0.024$
	Within populations	39	18.621	0.47747	97.63	
	Total	41	19.881	0.48906		
Domestic yaks	Among regions	5	4.828	0.00293	0.64	$F_{CT} = 0.006$
	Among populations	40	24.902	0.02508	5.46	$F_{SC} = 0.055^{**}$
	Within populations	353	152.398	0.43172	93.91	$F_{ST} = 0.061^{**}$
	Total	398	182.128	0.45973		
All yaks	Among groups	1	1.547	0.00906	1.92	$F_{CT} = 0.019$
	Among populations	47	30.989	0.02596	5.51	$F_{SC} = 0.056^{**}$
	Within populations	392	171.020	0.43627	92.57	$F_{ST} = 0.074^{**}$
	Total	440	203.556	0.47129		

d.f., degrees of freedom;  $F_{CT}$ , correlation of haplotypes within regions relative to total;  $F_{SC}$ , correlation within populations relative to regions;  $F_{ST}$ , correlation within populations relative to total;  $^{**}P < 0.01$ .



However, there are two reasons why we can exclude the possibility for this wild sample. First, this lineage was identified in the phylogenetic analyses of the D-loop haplotypes with two

individuals nested within it. Second, nucleotide alignments of each gene from this sample are highly similar to those of the other haplotypes without any indication of numts.

Based on the assumption that the divergence between yak and bison started around 2.5 Ma, the divergence time of the two main lineages (I and II) was estimated to have occurred 429,183 years ago, with a 95% highest posterior density (HPD) of 304,856–562,300 years ago; the divergence between these and the remote lineage III was estimated to have occurred 580,584 years ago with a 95% HPD of 422,655–739,360 years ago (Fig. 4).

## DISCUSSION

### Phylogeographical structure

Wild yaks are distributed in the northwestern QTP while domestic yaks radiated around this region (Wiener *et al.*, 2003). Despite the parapatric distributions, there are distinct morphological differences between them (Leslie & Schaller, 2009). In the present study, however, the three main haplogroups occurring at high frequencies in domestic yaks form a small proportion of the identified haplotypes in the wild populations. One more lineage (III), as well as many endemic haplogroups and/or haplotypes, were identified to be specific to the wild yaks (Figs 1 & 4). About 40% of the wild yaks were found to have specific haplotypes or haplogroups that did not belong to the six haplogroups recovered from the domestic yaks. It is obvious that there is a distinct phylogeographical difference between wild and domestic yaks. In addition, AMOVA analyses of the D-loop data suggested that about 2% of the total variance could be partitioned between them. However, such difference can be explained by the fact that only a part of the wild populations or individuals were involved in the yak domestication, although the actual ancestral population size is difficult to determine without further nuclear data.

We found that all domestic haplogroups included the haplotypes from the wild yaks. Therefore the genetic diversity of wild yaks is clearly higher than that of domesticated yaks, based on the mitochondrial D-loop sequences and genomic data (Table 1). This reduced diversity is a common feature of domestic animals when compared with their wild counterparts (Bruford *et al.*, 2003). However, within domestic yaks there was no correlation between the identified haplogroups and their geographical distributions (Fig. 2) or the current morphological breeds (Table 3). Samples from an individual population or morphological breed were distributed across different haplogroups, with each haplogroup containing individuals from different populations or breeds. These findings are largely consistent with our previous results (Guo *et al.*, 2006), despite the fact that the present analyses included all morphological breeds and comprised all samples across the QTP and adjacent regions of western China.

We further quantified the degree of structuring by computing the components of D-loop variation under the hierarchical AMOVA framework, and the results reflected the low phylogeographical structure within domestic yaks (Table 5): only 0.64% of the variance was attributed among

regions. Because most breeds represent local distribution (Wiener *et al.*, 2003), genetic correlations with morphological breeds may also reflect phylogeographical structure. However, we failed to find such a correlation (Table 3). Similarly, based on nuclear microsatellite alleles, Zhang *et al.* (2008) revealed that 94.4% of the genetic variation was observed within yak breeds, while only 5.6% of the genetic variation existed among breeds.

Although further studies based on nuclear data are needed, we tentatively suggest three possible factors that might account for the limited structures in domestic yak populations. First, these maternal haplogroups recovered from the domesticated yaks had formed in the wild yaks before domestication; this can be inferred from the phylogenetic analyses (Figs 1 & 4). Domestication of yaks from a single wild gene pool may have involved a few haplogroups, and there has been an insufficient amount of subsequent breeding in specific geographical areas and/or breed selection to produce maternal haplogroups through genetic drift. This hypothesis was also supported by the AMOVA analyses showing that only a small percentage of variation existed between the wild populations (2.37%) (Table 5), indicating little between-population genetic differentiation for the wild yaks.

Second, exchanges between different local breeds are frequent, thus preventing segregation of the maternal haplogroups. Domestic yaks were the main transport power resources in the QTP region in early herding history; the high mobility of domestic yaks might have caused extensive interchanges between different maternal lineages. Notably, hybridization between different breeds and sometimes even with wild yaks was widely used in cultivation; therefore these activities would blur the concordance between the current breeds and the identified haplogroups.

Third, morphological diversification and local fixing may result from recent breeding activities and artificial selection for a few beneficial nuclear alleles, thus reducing the strength of natural selection. However, such alleles are not linked to mtDNA variations. It is obvious that further studies, especially based on functional genes, are needed to clarify the genetic mechanisms underlying the morphological and/or geographical differentiation of the current yaks in the QTP and adjacent highlands of western China.

Since the genetic diversity is similar among the geographical regions (Table 2), it is difficult to determine the original sites of domestication. However, both the total number of haplotypes and the number of region-specific haplotypes are higher in Qinghai and Tibet, which are close to the current distribution of wild yaks. This might hint that the earliest domestication occurred in this region. Subsequently, the mitochondrial gene pool spread from the region of origin to other regions. All domesticated yak haplotypes of four of the six haplogroups formed a star-phylogeny encircling the basic haplotypes, and have a wide distribution and high frequency (Fig. 3), suggesting an exponential expansion following initial domestication (Avice, 2000). This hypothesis was further supported by Fu's  $F_S$  test (Table 4).

### Development of divergent lineages within yaks

The D-loop region is highly variable and sometimes subject to parallel mutation (e.g. Tamura & Nei, 1993), thus calibrations based on this region may distort the actual between-lineage divergence (Ingman & Gyllensten, 2001). Previous dating based on the D-loop region suggested that two lineages (lineages I and II here) diverged 100,000–130,000 years ago (Guo *et al.*, 2006) or more recently, 75,000 years ago (Ho *et al.*, 2008). In this study, we used the third codon sites of mitochondrial protein-coding genes to produce a new estimate of the divergence time between the major lineages of wild yaks. Based on this new approach and new calibration point (centred on 2.5 Ma), the divergence times of the three lineages were estimated to have occurred around 420,000 and 580,000 years ago. In addition to the older calibration point based on the bison fossil, the elevated divergence estimates of main lineages in this study seem more reasonable than the previous ones based on the mitochondrial D-loop region (Guo *et al.*, 2006; Ho *et al.*, 2008), and probably reflect the following factors: (1) one more lineage was included in the present analysis, which might be more representative of the total intraspecific phylogeny; (2) a larger genomic dataset allows an improvement in phylogenetic resolution; and (3) the present molecular clock-like dataset might have avoided a few errors in the previous study, such as rate variation among lineages, parallel mutations in the D-loop 'hotspot' mutation region, and time-dependent rates (Ho *et al.*, 2005).

Our current estimates might not fully reflect the actual diversification time-scales of yaks because they are based exclusively on mitochondrial data. However, these estimates provide a preliminary timeframe to trace the intraspecific differentiation of the maternal lineages of yaks. Two alternative hypotheses are suggested to explain the development of divergent lineages within a species: the extremely large ancestral population size; or allopatric differentiation in the past (Avice, 2000).

Under the first scenario, we followed Luikart *et al.* (2001) to compute the possible effective size ( $N_e$ ) of an ancestral population necessary to maintain three divergent mtDNA lineages we observed today. To perform this calculation, we assumed the coalescent time of all lineages in the ancestral population to be around 580,000 years ago, or 290,000 generations of 2 years. The expectation of this time is  $2N_e(1-1/k)$  generations, where  $k$  is 72 mt protein-coding sequences, implying an  $N_e$  value of 147,000. A generation time of 3 years would lead to  $N_e = 98,000$ . The population census size would have to be far larger than this estimated  $N_e$  (Avice, 2000). Wild yaks have been declining ever since modern humans reached the QTP (Schaller, 1998); based on the limited number of wild yaks currently present on the QTP (Leslie & Schaller, 2009), it is unlikely that a population of this size was present at any time (Wiener *et al.*, 2003). Such a large ancestral population is also unlikely for any other ungulate species (Luikart *et al.*, 2001).

Therefore the second scenario is more likely: namely that the glacial climate drove the wild yaks into a number of refugia,

where the populations diverged, they then mixed again when the climate improved. The divergence times of the three lineages of yaks estimated here are highly consistent with the large glaciation events recorded for the QTP during the middle–late Pleistocene (Zheng *et al.*, 2002). The ice-sheet size of the Last Glacial Maximum (LGM) (which started around 20,000 years ago) in the QTP remains disputed (Shi *et al.*, 1998; Kaufmann, 2005; Kuhle, 2007). However, it is commonly agreed that the most extensive glacial activity on the QTP occurred between 800,000 and 600,000 years ago in the Pleistocene: the Naynayxungla Glaciation formed an ice sheet around 500,000 km<sup>2</sup> (Shi *et al.*, 1990, 1998; Wu *et al.*, 2001; Zheng *et al.*, 2002; Owen *et al.*, 2006). In addition, the subsequent and relatively extensive Nebraskan glaciation occurred between 500,000 and 400,000 years ago (Shi *et al.*, 1998; Owen *et al.*, 2006), as occurred in North America (Dort, 2008). Although temperatures rose during the interglacial stages, development of glaciers and/or extremely low temperatures at high elevations (> 4500 m) throughout the Pleistocene may have continued to impede the gene flow of wild yaks, eventually resulting in the genetic divergence of yaks in the allopatric regions. Recently, a few other animal and plant species that have been studied on the QTP have been found to exhibit similar intraspecific divergences, which largely correspond to the Quaternary glaciations during the mid- and late-Pleistocene: for example, the white-rumped snowfinch (*Onychostruthus taczanowskii*, Yang *et al.*, 2006), the schizothoracine fish (*Schizopygopsis pylzovi*, Qi *et al.*, 2007), the toad-headed lizard (*Phrynocephalus vlangalii*, Jin *et al.*, 2008), the plateau zokor (*Eospalax baileyi*, Tang *et al.*, 2010) and an alpine herb (*Aconitum gymnantrum*, Wang *et al.*, 2009). Within the toad-headed lizard and the plateau zokor, a series of divergences in several lineages were found to be highly associated with these extensive glacial advances (Jin *et al.*, 2008; Tang *et al.*, 2010). However, the diverged lineages of these other species are allopatrically distributed across their current ranges. Presumably, it is the high mobility of yaks and the deteriorating habitats on the QTP that might have resulted in a reunion of the allopatrically divergent lineages of the wild yak at the end of the LGM into a single gene pool within their current limited distributional range. Therefore later domestication(s) from this gene pool around the start of the Holocene (Schaller, 1998) might have further led to the presence of two divergent lineages within domestic yaks, although a third remote lineage remained undomesticated. Despite the fact that mobility and domestication might have complicated the intraspecific maternal divergences of yaks, our analyses highlight the roles of past extensive glaciations on the recent evolution of this species.

### ACKNOWLEDGEMENTS

We thank Matthew T. Webster, Jennifer A. Leonard, Kangshan Mao and Greger Larson for discussion. We also thank Bin Tian and Xinmin Tian for their help with collecting samples, and Guili Wu for her help with sequencing. We are grateful to the

editor and two anonymous referees for their helpful suggestions and comments on the manuscript. This study was supported by grants from the National Natural Science Foundation of China (30725004), the '111' project from the foreign Ministry of China, and the Australian Research Council.

## REFERENCES

- Achilli, A., Olivieri, A., Pellecchia, M. *et al.* (2008) Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Current Biology*, **18**, 157–158.
- Avise, J.C. (2000) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, MA.
- Bandelt, H., Forster, P. & Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48. Available at: <http://www.fluxus-engineering.com/sharenet.htm>.
- Beja-Pereira, A., England, P.R., Ferrand, N., Jordan, S., Bakhiet, A.O., Abdalla, M.A., Mashkour, M., Jordana, J., Taberlet, P. & Luikart, G. (2004) African origins of the domestic donkey. *Science*, **304**, 1781.
- Brown, W.M., George, M., Jr & Wilson, A.C. (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences USA*, **76**, 1967–1971.
- Bruford, M.W., Bradley, D.G. & Luikart, G. (2003) DNA markers reveal the complexity of livestock domestication. *Nature Review Genetics*, **4**, 900–910.
- Diamond, J. (2002) Evolution, consequences and future of plant and animal domestication. *Nature*, **418**, 700–707.
- Dort, W., Jr (2008) Stadial subdivisions of early Pleistocene glaciations in central United States – a developing chronology. *Boreas*, **1**, 55–61.
- Drummond, A.J. & Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.
- Endicott, P. & Ho, S.Y.W. (2008) A Bayesian evaluation of human mitochondrial substitution rates. *American Journal of Human Genetics*, **82**, 895–902.
- Ewing, B. & Green, P. (1998) Base-calling of automated sequencer traces using *phred*. II. Error probabilities. *Genome Research*, **8**, 186–194.
- Excoffier, L., Smouse, P.E. & Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.
- Excoffier, L., Laval, G. & Schneider, S. (2005) ARLEQUIN (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.
- Flerow, C.C. (1980) On the geographic distribution of the genus *Poephagus* during the Pleistocene and Holocene. *Quaternary Paleontology (East Berlin)*, **4**, 123–126.
- Fu, Y.X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.
- Gilbert, M.T.P., Drautz, D.I., Lesk, A.M. *et al.* (2008) Intra-specific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proceedings of the National Academy of Sciences USA*, **105**, 8327–8332.
- Giles, R.E., Blanc, H., Cann, H.M. & Wallace, D.C. (1980) Maternal inheritance of human mitochondrial DNA. *Proceedings of the National Academy of Sciences USA*, **77**, 6715–6719.
- Gordon, D., Abajian, C. & Green, P. (1998) Consed: a graphical tool for sequence finishing. *Genome Research*, **8**, 195–202.
- Green, P. (2010) *The Phred/Phrap/Consed system home page. Phrap Assembler*. Available at: <http://www.phrap.org/phredphrap/phrap.html>.
- Gu, Z.L., Zhao, X.B., Li, N. & Wu, C.X. (2005) Complete sequence of the yak (*Bos grunniens*) mitochondrial genome and its evolutionary relationships with other ruminants. *Molecular Phylogenetics and Evolution*, **42**, 248–255.
- Guo, S.C., Savolainen, P., Su, J.P., Zhang, Q., Qi, D.L., Zhou, J., Zhong, Y., Zhao, X.Q. & Liu, J.Q. (2006) Origin of mitochondrial DNA diversity of domestic yaks. *BMC Evolutionary Biology*, **6**, 73.
- Gyllensten, U., Wharton, D. & Wilson, A.C. (1985) Maternal inheritance of mitochondrial DNA during backcrossing of two species of mice. *Journal of Heredity*, **76**, 321–324.
- Ho, S.Y.W. & Phillips, M.J. (2009) Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology*, **58**, 367–380.
- Ho, S.Y.W., Phillips, M.J., Cooper, A. & Drummond, A.J. (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution*, **22**, 1561–1568.
- Ho, S.Y.W., Larson, G., Edwards, C.J., Heupink, T.H., Lakin, K.E., Holland, P.W.H. & Shapiro, B. (2008) Correlating Bayesian date estimates with climatic events and domestication using a bovine case study. *Biology Letters*, **4**, 370–374.
- Huelsenbeck, J.P. & Rannala, B. (1997) Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science*, **276**, 227–232.
- Huelsenbeck, J.P. & Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogeny. *Bioinformatics*, **17**, 754–755.
- Ingman, M. & Gyllensten, U. (2001) Analysis of the complete human mtDNA genome: methodology and inferences for human evolution. *Journal of Heredity*, **92**, 454–461.
- Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. (2000) Mitochondrial genome variation and the origin of modern humans. *Nature*, **408**, 708–713.
- 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. *Proceedings of the National Academy of Sciences USA*, **99**, 10905–10910.
- Jin, Y.T., Brown, R.P. & Liu, N.F. (2008) Cladogenesis and phylogeography of the lizard *Phrynocephalus vlangalii* (Agamidae) on the Tibetan plateau. *Molecular Ecology*, **17**, 1971–1982.

- Kaufmann, G. (2005) Geodetic signatures of a Late Pleistocene Tibetan ice sheet. *Journal of Geodynamics*, **39**, 111–125.
- Kierstein, G., Vallinoto, M., Silva, A., Schneider, M.P., Iannuzzi, L. & Brenig, B. (2004) Analysis of mitochondrial D-loop region casts new light on domestic water buffalo (*Bubalus bubalis*) phylogeny. *Molecular Phylogenetics and Evolution*, **30**, 308–324.
- Kivisild, T., Shen, P., Wall, D.P., Do, B., Sung, R., Davis, K., Passarino, G., Underhill, P.A., Scharfe, C., Torroni, A., Scozzari, R., Modiano, D., Coppa, A., de Knijff, P., Feldman, M., Cavalli-Sforza, L.L. & Oefner, P.J. (2006) The role of selection in the evolution of human mitochondrial genomes. *Genetics*, **172**, 373–387.
- Kuhle, M. (2007) Critical approach to the methods of glacier reconstruction in High Asia (Qinghai-Xizang Tibet) Plateau, West Sichuan Plateau, Himalaya, Karakorum, Pamir, Kuenlun, Tienshan) and discussion of the probability of a Qinghai-Xizang (Tibetan) inland ice. *Journal of Mountain Science*, **4**, 91–123.
- Lai, S.J., Chen, S.Y., Liu, Y.P. & Yao, Y.G. (2007) Mitochondrial DNA sequence diversity and origin of Chinese domestic yak. *Animal Genetics*, **38**, 77–80.
- Larson, G., Dobney, K., Albarella, U., Fang, M., Matisoo-Smith, E., Robins, J., Lowden, S., Finlayson, H., Brand, T., Willerslev, E., Rowley-Conwy, P., Andersson, L. & Cooper, A. (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science*, **307**, 1618–1621.
- Lau, C.H., Drinkwater, R.D., Yusoff, K., Tan, S.G., Hetzel, D.J.S. & Barker, J.S.F. (1998) Genetic diversity of Asian water buffalo (*Bubalus bubalis*): mitochondrial DNA D-loop and cytochrome *b* sequence variation. *Animal Genetics*, **29**, 253–264.
- Leslie, D.M., Jr & Schaller, G.B. (2009) *Bos grunniens* and *Bos mutus* (Artiodactyla: Bovidae). *Mammalian Species*, **836**, 1–17.
- Li, J.J., Shi, Y.F. & Li, B.Y. (1995) *Uplift of the Qinghai-Xizang (Tibet) Plateau and global change*. Lanzhou University Press, Lanzhou, China.
- Luikart, G., Gellie, L., Excoffier, L., Vigne, J.D., Bouvet, J. & Taberlet, P. (2001) Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proceedings of the National Academy of Sciences USA*, **98**, 5927–5932.
- Mason, I.L. (1984) *Evolution of domesticated animals*. Longman, New York.
- Meadows, J.R.S., Cemal, I., Karaca, O., Gootwine, E. & Kijas, J.W. (2007) Five ovine mitochondrial lineages identified from sheep breeds of the Near East. *Genetics*, **175**, 1371–1379.
- Naderi, S., Rezaei, H., Pompanon, F., Blum, M.G.B., Negrini, R., Naghash, H., Balkiz, Ö., Mashkour, M., Gaggiotti, O.E., Ajmone-Marsan, P., Kence, A., Vigne, J. & Taberlet, P. (2008) The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proceedings of the National Academy of Sciences USA*, **105**, 17659–17664.
- Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P. & Nieves-Aldrey, J.L. (2004) Bayesian phylogenetic analysis of combined data. *Systematic Biology*, **53**, 47–67.
- Olivo, P.D., Van de Walle, M.J., Laipis, P.J. & Hauswirth, W.W. (1983) Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop. *Nature*, **306**, 400–402.
- Olsen, S.J. (1990) Fossil ancestry of the yak, its cultural significance and domestication in Tibet. *Proceedings of the Academy of Natural Sciences of Philadelphia*, **142**, 73–100.
- Owen, L.A., Finkel, R.C., Haizhou, M. & Barnard, P.L. (2006) Late Quaternary landscape evolution in the Kunlun Mountains and Qaidam Basin, Northern Tibet: a framework for examining the links between glaciation, lake level changes and alluvial fan formation. *Quaternary International*, **154–155**, 73–86.
- Parr, R.L., Maki, J., Reguly, B., Dakubo, G.D., Aguirre, A., Wittock, R., Robinson, K., Jakupciak, J.P. & Thayer, R.E. (2006) The pseudo-mitochondrial genome influences mistakes in heteroplasmy interpretation. *BMC Genomics*, **7**, 185.
- Pedrosa, S., Uzun, M., Arranz, J.J., Gutiérrez-Gil, B., San Primitivo, F. & Bayón, Y. (2005) Evidence of three maternal lineages in Near Eastern sheep supporting multiple domestication events. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 2211–2217.
- Posada, D. & Crandall, K.A. (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.
- Qi, D.L., Guo, S.C., Zhao, X.Q., Yang, J. & Tang, W.J. (2007) Genetic diversity and historical population structure of *Schizopygopsis pylzovi* (Teleostei: Cyprinidae) in the Qinghai-Tibetan Plateau. *Freshwater Biology*, **52**, 1090–1104.
- Qi, X.B., Han, J.L., Lkhagva, J.B., Chekarova, I., Badamdorj, D., Rege, J.E.O. & Hanotte, O. (2005) Genetic diversity and differentiation of Mongolian and Russian yak populations. *Journal of Animal Breeding and Genetics*, **122**, 117–126.
- Qian, Y.W. (1979) *The origin of domesticated animal: biohistory*. Science Press, Beijing, China.
- Rambaut, A. & Drummond, A.J. (2010) *Tracer v1.5*. Available at: <http://beast.bio.ed.ac.uk/Tracer> (accessed 10 January 2010).
- Sambrook, J. & Russell, D.W. (2001) *Molecular cloning: a laboratory manual*, 3rd edn. Cold Spring Harbor Laboratory Press, New York.
- Schaller, G.B. (1998) *Wildlife of the Tibetan steppe*. University of Chicago Press, Chicago.
- Schaller, G.B. & Liu, W.L. (1996) Distribution, status, and conservation of wild yak *Bos grunniens*. *Biological Conservation*, **76**, 1–8.
- Shi, Y.F., Zheng, B.X. & Li, S.J. (1990) Last glaciation and maximum glaciation in Qinghai-Xizang (Tibet) plateau. *Journal of Glaciology and Geocryology*, **12**, 1–15.
- Shi, Y.F., Li, J.J. & Li, B.Y. (1998) *Uplift and environmental changes of Qinghai-Tibetan plateau in the Late Cenozoic*. Guangdong Science and Technology Press, Guangzhou (in Chinese).

- Subramanian, S., Denver, D.R., Millar, C.D., Heupink, T., Aschrafi, A., Emslie, S.D., Baroni, C. & Lambert, D.M. (2009) High mitogenomic evolutionary rates and time dependency. *Trends in Genetics*, **25**, 482–486.
- Swofford, D.L. (2002) *PAUP\*: phylogenetic analysis using parsimony (\*and other methods)*. Version 4. Sinauer Associates, Sunderland, MA.
- Tamura, K. & Nei, M. (1993) Estimation of the number of nucleotide substitutions in the D-loop of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.
- Tang, L.Z., Wang, L.Y., Cai, Z.Y., Zhang, T.Z., Ci, H.X., Lin, G.H., Su, J.P. & Liu, J.Q. (2010) Allopatric divergence and phylogeographic structure of the plateau zokor (*Eospalax baileyi*), a fossorial rodent endemic to the Qinghai-Tibetan Plateau. *Journal of Biogeography*, **37**, 657–668.
- Tedford, R.H., Flynn, L.J., Zhanxiang, Q., Opdyke, N.D. & Downs, W.R. (1991) Yushe Basin, China: paleomagnetically calibrated mammalian biostratigraphic standard for the late Neogene of eastern Asia. *Journal of Vertebrate Paleontology*, **11**, 519–526.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997) The Clustal\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **24**, 4876–4882.
- Troy, C.S., MacHugh, D.E., Bailey, J.F., Magee, D.A., Loftus, R.T., Cunningham, P., Chamberlain, A.T., Sykes, B.C. & Bradley, D.G. (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature*, **410**, 1088–1091.
- Tu, Z.C., Zhang, Y.P. & Qiu, H. (1997) Genetic diversity and divergence in Chinese yak (*Bos grunniens*) populations inferred from blood protein electrophoresis. *Biochemical Genetics*, **35**, 13–16.
- Tu, Z.C., Qiu, H. & Zhang, Y.P. (2002) Polymorphism in mitochondrial DNA (mtDNA) of yak (*Bos grunniens*). *Biochemical Genetics*, **40**, 187–193.
- Wakeley, J. (1993) Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. *Journal of Molecular Evolution*, **37**, 613–623.
- Wang, L.Y., Abbott, R.J., Zheng, W., Chen, P., Wang, Y.J. & Liu, J.Q. (2009) History and evolution of alpine plants endemic to the Qinghai-Tibetan Plateau: *Aconitum gymnantrum* (Ranunculaceae). *Molecular Ecology*, **18**, 709–721.
- Wiener, G., Han, J.L. & Long, R.J. (2003) *The Yak*, 2nd edn. Regional Office for Asia and the Pacific, Food and Agriculture Organization of the United Nations, Bangkok, Thailand.
- Wu, G.S., Yao, Y.G., Qu, K.X., Ding, Z.L., Li, H., Palanichamy, M.G., Duan, Z.Y., Li, N., Chen, Y.S. & Zhang, Y.P. (2007) Population phylogenomic analysis of mitochondrial DNA in wild boars and domestic pigs revealed multiple domestication events in East Asia. *Genome Biology*, **8**, R245.
- Wu, Y.Q., Cui, Z.J. & Liu, G.N. (2001) Quaternary geomorphological evolution of the Kunlun Pass areas and uplift of the Qinghai-Xizang (Tibet) plateau. *Geomorphology*, **36**, 203–216.
- Yang, S.J., Lei, F.M., Qu, Y.H. & Yin, Z.H. (2006) Intraspecific phylogeography of the white-rumped snowfinch (*Onychostreuthus taczanowskii*) endemic to the Tibetan Plateau based on mtDNA sequences. *Journal of Zoology*, **268**, 187–192.
- Yonezawa, T., Nikaido, M., Kohno, N., Fukumoto, Y., Okada, N. & Hasegawa, M. (2007) Molecular phylogenetic study on the origin and evolution of Mustelidae. *Gene*, **396**, 1–12.
- Zhang, G.X., Chen, W.S., Xue, M., Wang, Z.G., Chang, H., Han, X., Liao, X.J. & Wang, D.L. (2008) Analysis of genetic diversity and population structure of Chinese yak breeds (*Bos grunniens*) using microsatellite markers. *Journal of Genetics and Genomics*, **35**, 233–238.
- Zheng, B.X., Xu, Q.Q. & Shen, Y.P. (2002) The relationship between climate change and Quaternary glacial cycles on the Qinghai-Tibetan Plateau: review and speculation. *Quaternary International*, **97–98**, 93–101.
- Zwickl, D.J. (2010) *GARLI, Genetic Algorithm for Rapid Likelihood Inference, Version 0.96*. Available at: [http://www.nescent.org/wg\\_garli/Main\\_Page](http://www.nescent.org/wg_garli/Main_Page).

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Appendix S1** The samples selected for complete mitochondrial sequencing.

**Appendix S2** Sampled breeds of domestic yak (*Bos grunniens*) and collection information from each province in western China.

**Appendix S3** Primers used for amplifying the complete mitochondrial genome.

**Appendix S4** Primers used for sequencing the complete mitochondrial genome.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be reorganized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

## BIOSKETCHES

The authors collaborate to study the population dynamics of alpine animals and plants on the Qinghai-Tibetan Plateau.

Author contributions: J.Q.L. conceived the research and manuscript outline; Z.F.W., X.S., B.L., J.P.S., S.C.G. and Y.Y. collected the data; Z.F.W., T.Y. and S.Y.W.H. analysed the data; and J.Q.L., Z.F.W., T.Y., M.H. and C.V. wrote the manuscript.

Editor: Pauline Ladiges