

Full Length Research Paper

Genetic diversity and demographic history of wild Yak (*Bos grunniens mutus*) inferred from mtDNA D-loop sequences

Zhijie Ma^{1*}, Jincheng Zhong², Jianlin Han³, Jintao Xu¹, Wenlin Bai⁴, Jiye Li⁵, Dongzhi Cairang¹, Xinyan Jin¹ and Jun Zhang¹

¹Academy of Animal Science and Veterinary Medicine, Qinghai University, Xining 810016, China.

²College of Life Science and Technology, Southwest University for Nationalities, Chengdu 610041, China.

³International Livestock Research Institute, P. O. Box 30709, Nairobi 00100, Kenya.

⁴College of Animal Science and Veterinary Medicine, Shenyang Agricultural University, Shenyang, 110161, China.

⁵Datong Breeding Farm of Qinghai Province, Datong 810102, China.

Accepted 22 October, 2010

The wild yak (*Bos grunniens mutus*) is considered as one of the endangered species. To provide useful information for conservation and utilization of this species, we investigated the genetic diversity and demographic history of wild yak by analyzing a 637 bp hypervariable fragment of 21 mitochondrial DNA (mtDNA) D-loop sequences including six sequenced in this study and 15 downloaded from GenBank. Forty-five variable sites and 15 haplotypes (including three haplotypes present firstly in this study) were identified, and the nucleotide and haplotype diversities were 0.024430 ± 0.012685 and 0.9619 ± 0.0260 , respectively, indicating a rich genetic diversity in wild yak. The NJ phylogenetic tree revealed two distinct lineages and the selective neutrality and mismatch distribution tests all showed no population expansion event in the demographic history of the wild yak.

Key words: Wild yak, mitochondrial DNA, D-loop region, genetic diversity, population expansion.

INTRODUCTION

The wild yak (*Bos grunniens mutus*), a species of *Bovidae* family, is a rare and valuable wildlife found on Qinghai-Tibetan Plateau with an altitude above 4,500 m. At present, there are about 15,000 heads of wild yak only survived in China (Wiener et al., 2003). The wild yak is considered as one of the endangered species. Over the past decades, some researchers have analyzed the genetic diversity of wild yak at morphological, physiological, biochemical and chromosomal levels (Li and Lu, 1990; Li et al., 1998). In recent years, preliminary analysis about the genetic diversity of wild yak has been done at the molecular level (Wang, 2000; Hu, 2001; Guo et al., 2006), which indicating abundant genetic diversity in wild yak population. Moreover, to the problem of demographic

mtDNA control region has been proven to be a powerful tool for investigating intra- or inter-species genetic variation, population structure and demographic history (Wolf et al., 1999). Therefore, in this study, we sequenced six wild yak mtDNA D-loop sequences and downloaded 15 corresponding wild yak sequences. Based on these 21 wild yak mtDNA D-loop sequences, we analyzed the genetic diversity of the wild yak and inferred their population expansion status. The results are useful for the conservation and utilization of wild yak genetic resources.

MATERIALS AND METHODS

Samples collection

Fresh blood samples of six wild yaks were collected from the Datong Breeding Farm of Qinghai province in China. They were captured in 7806 Afr. J. Biotechnol.

*Corresponding author. E-mail: maziwise2004@yahoo.com.
Fax: 86-971-5318783.

history, little is also known only for the wild yak. The

infancy on the Qinghai-Tibetan Plateau, but no precise details of the locations of the captures are available.

Other mtDNA D-loop sequences in this study were downloaded from GenBank which included 15 complete or partial D-loop sequences of the wild yak (Accession numbers: AY749414, AY722118 and DQ139202-DQ139214) and a complete D-loop sequence of the cattle (*B. taurus*) (Accession number: V00654) as an out-group.

DNA extraction, PCR amplification and sequencing

DNA was extracted from frozen blood of six wild yaks using a standard phenol-chloroform method (Sambrook and Russell, 2001). The mtDNA D-loop region was amplified using the primers PF 5'-cta cag tct cac cgt caa cc-3' and PR 5'-ggg gtg tag atg ctt gc-3'. The polymerase chain reaction (PCR) reaction mixture contained 50 to 100 ng genomic DNA of wild yak, 20 pM of each primer, 0.50 U ExTaq DNA polymerase (TakaRa, Dalian, China), 10 × ExTaq Buffer (Mg²⁺ free), 0.25 mM dNTP (deoxynucleotides), 2.5 mM MgCl₂ and ddH₂O in a final volume of 25 μl. The following procedures were applied: an initial denaturation at 95°C for 4 min, followed by 35 cycles at 95°C/1 min, 58.5°C/45 s and 72°C/1 min 30 s, and a final extension at 72°C for 5 min. The PCR products were electrophoresed on a 1.0% agarose gel and purified using the DNA Agarose Gel Extraction Kit (Omega) according to the manufacturer's instructions. The purified fragments were cloned into pMD18-T vector and subsequently transformed into *Escherichia coli* JM109. After 15 to 20 h, single colonies were inoculated to obtain recombinant plasmid. The recombinant plasmid DNA was extracted and then sequenced using an ABI 3730 automated sequencer (Applied Biosystems).

Data analysis

All nucleotide sequences of the wild yak were aligned using BioEdit 7.0.9 (Hall 1999) software with the Clustal W multiple alignment program and refined manually. Sequence variation sites, average nucleotide composition, nucleotide diversity (π) and haplotype diversity (h) within population were estimated using DnaSP 4.10.1 (Rozas et al., 2003) and Arlequin 3.11 software (Excoffier et al., 2005). A neighbor-joining tree was constructed for all haplotypes with the cattle sequence (Accession number: V00654) as an out-group based on the Kimura's 2 parameter model using Mega 4.0 (Kumar et al., 2004).

Two different approaches were used to investigate the demographic history of all samples and each lineage. First, Fu's F_s (Fu 1997) and Tajima's D statistics (Tajima 1989) were used to test whether the sequences conformed to the expectations of neutrality because both two test methods are appropriate for short DNA sequences. Fu's F_s statistic was very sensitive to population demographic expansion, which generally leads to a large negative F_s value, and significant D value may be because of factors like population expansion and bottleneck. Secondly, the observed distribution of pair-wise differences between sequences (Slatkin and Hudson, 1991) was examined. Populations that have been stable over time are predicted to have a bimodal or multimodal mismatch distribution, whereas a unimodal distribution is generally found in population having passed through a bottleneck or recent demographic expansion. In addition, the goodness-of-fit of observed distribution with the expected distribution was tested by calculating the sum of squared deviation (SSD) and raggedness index (r) with 1000 bootstrap replicates. The above analysis about demographic history of the wild yak was also carried out using DnaSP 4.10.1 and Arlequin 3.11 software.

RESULTS

Sequence variation and nucleotide diversity

The mtDNA D-loop sequences of six wild yaks were PCR-amplified, sequenced and all sequences were deposited in GenBank (Accession numbers: FJ548840-FJ548845). Combined with the other 15 D-loop sequences of wild yak and taking the sequence of DQ139203 (Hap1, Figure 1) as a reference, all 21 sequences were aligned. A 637-bp fragment of the hypervariable region of mtDNA D-loop sequences was used to analyze the genetic diversity. There were 45 variable sites within the fragment (Figure 1), accounting for 7.06% of the total number of sites. Among these polymorphic sites, ten were singleton variable sites and 34 were parsimony-informative sites. The 45 variable sites were comprised of one in/del, 41 transitions and three transversions. The average nucleotide composition of all sequences was 31.86% A, 27.70% T, 15.45% G and 24.99% C and the average nucleotide content of A + T (59.56%) was obviously higher than that of G + C (40.44%). The nucleotide diversity and the mean number of pair wise differences were 0.024430 ± 0.012685 and 15.561905 ± 7.238573 , respectively, indicating a relatively rich genetic diversity in wild yak.

Haplotype diversity and phylogenetic tree

Fifteen haplotypes were identified from 21 wild yak D-loop sequences based on nucleotide variation (Figure 1). The number of sequences with each haplotypes varied: the haplotypes with the largest number of sequences (Hap2 and Hap11) consisted of three sequences each; the Hap3 and Hap13 included two sequences each; and the rest haplotypes included only one sequence each. Compared with the previous study (Guo et al., 2006), three haplotypes (Hap4, Hap8 and Hap12) were firstly detected in wild yak populations in this study. Haplotype diversity (h) was 0.9619 ± 0.0260 , indicating a rich genetic diversity in wild yak population. Based on Kimura's two-parameter model, the pair-wise genetic distances between haplotypes fell in the range of 0.010 ± 0.004 to 0.046 ± 0.009 and the mean pair-wise genetic distances among all the haplotypes was 0.025 ± 0.004 .

Taking the cattle (*B. taurus*) counterpart as an out-group (Accession number: V00654), a neighbor-joining tree of the 15 haplotypes from 21 D-loop sequences of the wild yak was constructed based on Kimura's two-parameter genetic distances (Figure 2), demonstrating all haplotypes could be classified into two lineages (termed as lineages A and B) that included eight and seven haplotypes, representing 11 and 10 D-loop sequences, respectively.

Population expansion analysis

Population expansion analysis of whole population and
Ma et al. 7807

Site	572883534011808703458913028305628956031	Genbank Numbers
Hap1	ATTATA-TACATACTAAGAATGCTCACCGCCTTCTAATCCTTAAT	[DQ139203]
Hap2-.....G.....	[FJ548840 FJ548842 DQ139205]
Hap3	..A...-...A.....G.....	[DQ139204 DQ139210]
Hap4	..AG.-...AG.C...G...G.....C.....	[FJ548843]
Hap5	G.A...-A...A.....G.....C..C	[AY722118]
Hap6	G.A...-AG..AG.....G.....G...C..C	[AY749414]
Hap7	..AGCG-. TGA..C...GGC.....T.T.CC.....TT..G..	[DQ139202]
Hap8	..AGCG-. GT. AGTCG.....A.CT...A.TCC.CGG...C..G.	[FJ548845]
Hap9	..AGCGA.GT. AGTCG.....A.CT...A.TCC.CGG...C..G.	[DQ139206]
Hap10	..AGCG-. GT. AG.CG.....A.CT..TA.T.C.C.G...C..G.	[DQ139208]
Hap11	..AGCGA.GT. AGTCGG.....ATCT...A.T.C.CGG...C..G.	[FJ548841 DQ139209 DQ139213]
Hap12	..AGCGA.GT. AGTCGG.....ATCT.T.A.T.C.CGG...C..G.	[FJ548844]
Hap13	.CAGCGA.GT. AG.CGG.....ATCT...ATT.C.CGG...C..G.	[DQ139211 DQ139212]
Hap14	..AGCG-. GT. AGTCG.A...A.CT...A.T.CT..GC..C..G.	[DQ139214]
Hap15	..AGCG-. GT. AGTCG.A...ATCT...A.T...CGG...C..G.	[DQ139207]

Figure 1. Polymorphic sites within 15 mtDNA D-loop region haplotypes of wild yak. The genbank numbers that share the same haplotype are listed in the right column. Dots (.) denote the nucleotide identical to that of reference sequences (Hap1, DQ139203). Horizontal lines (-) denote the nucleotide in/del. The top three rows of numbers represent the concrete polymorphic positions and should be read from up to down.

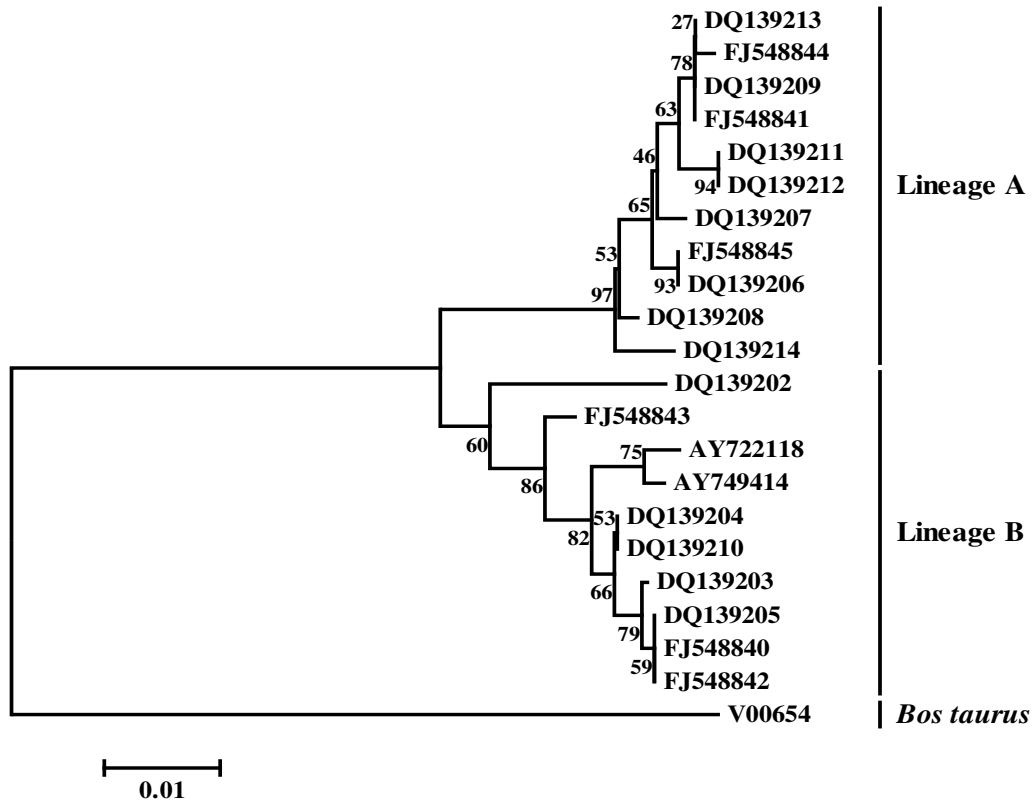


Figure 2. Neighbor-joining tree and bootstrap support values of haplotypes of the partial mtDNA D-loop sequences in wild yak.

Table 1. Number of sequences (n), number of haplotypes (H), Fu's Fs, Tajima's D, mismatch distribution test (SSD) and raggedness index (r) in wild yak population.

Lineage	n	H	Fs	D	SSD	r
A	11	8	-1.6151 (0.1430)	-0.5178 (0.3250)	0.0238 (0.3780)	0.0707 (0.3310)
B	10	7	0.1011 (0.4820)	-1.1960 (0.1320)	0.0203 (0.8230)	0.0435 (0.8130)
Total	21	15	-0.6957 (0.3810)	0.9280 (0.8450)	0.0162 (0.3250)	0.0138 (0.7580)

P-values are in parentheses; SSD, sum of squared deviations.

each lineage was taken to investigate demographic history of the wild yak. The result showed that neither the selective neutrality test nor the mismatch distribution test supported the hypothesis that the wild yak had passed through a population expansion. Firstly, no statistical significance for Fu's Fs or Tajima's D values was observed for whole population or for each lineage ($P > 0.10$) (Table 1). Secondly, the mismatch distribution revealed that the shapes of the distributions were all ragged and multimodal which showed whole population and each lineage of the wild yak has not suffered from a population expansion (Figure 3). At the same time, the model of recent population expansion was also rejected by significant SSD and r values for all tested datasets (Table 1). Thus we concluded that the wild yak population had not passed through a population expansion during the past decades and the size of the population may have kept stable.

DISCUSSION

Many kinds of molecular markers have provided an opportunity for assessing accurately the genetic diversity of wild yak. In this study, the analysis of the 21 wild yak mtDNA D-loop sequences showed that the D-loop control region is an A+T-rich region of the mtDNA genome which is consistent with the result from vertebrate mtDNA D-loop sequences (Brown et al., 1986). Besides, the haplotype diversity and nucleotide diversity of the wild yak in our study are higher ($h = 0.9619 \pm 0.0260$ and $\pi = 0.024430 \pm 0.012685$), indicating a rich genetic diversity in wild yak population. The result is also similar with the results from wild yak based on microsatellite markers (Wang 2000; Hu 2001) and partial mtDNA D-loop sequences (Guo et al., 2006). However, compared with the previous study (Guo et al., 2006), this study identified three new haplotypes (Hap4, Hap8 and Hap12) in wild yak. This indicated that there was more abundant genetic information existing in wild yak population.

In our study, 15 D-loop haplotypes of wild yak were classified into two lineages which consistent with the result implied by Guo et al. (2006). They also classified 10 D-loop haplotypes of wild yak into two lineages (one lineage with 4 haplotypes and another lineage with 6 haplotypes). Based on mtDNA D-loop sequences of 31 domestic yaks, Lai et al. (2005) analyzed the population

expansion of domestic yak and showed that domestic yak had not passed through population expansion in the past. But the research by Guo et al. (2006) to 278 domestic yak and 13 wild yaks showed that every clades in yak had undergone population expansion. However, the demographic history of wild yak has not been investigated independently and no information is available. In this study, our research suggested firstly no population expansion events occurring in the demographic history of the wild yak.

The wild yak is one of the wild species only lived in Qinghai-Tibetan Plateau. Our study showed that the genetic diversity of wild yak population was relatively higher. Thus, if proper protection and utilization measures are taken, it is possible to conserve the genetic diversity and increase their population size. So, the following strategies for conservation and utilization of this species should be considered. First, to avoid inbreeding among related individuals and a decline in population genetic diversity, appropriate genetic markers should be used to identify the pedigree and build the pedigree records of wild yak population. Secondly, because some haplotypes (such as hap1,4 to 10, 12, 14 to 15) were represented by only one individual in this study, these living individuals should be given more chances to produce offspring in order to preserve these rare alleles. Thirdly, according to the results of this study, wild yak population could be divided into two lineages which implies that the wild yak maybe derived from two independent populations or be domesticated from two independent domestication places, or from a domesticated hybrid population which underwent a certain stress condition within the population resulted in the differentiation of two lineages. Thus, the wild yak population can be considered as two kinds of gene banks to be protected in the future. Fourth, collection of more samples to carry out population genetics research on wild yak is necessary for providing the theoretical guidance for the protection and utilization of this species.

ACKNOWLEDGEMENTS

We thank Wei Y P, He X L, Tong Z B, Chen S M, Wu F D and La H who helped us for sample collection. This study was supported by the Young and Middle-aged Scientific Research Fund Project of Qinghai University (no. 2008-Ma et al. 7809

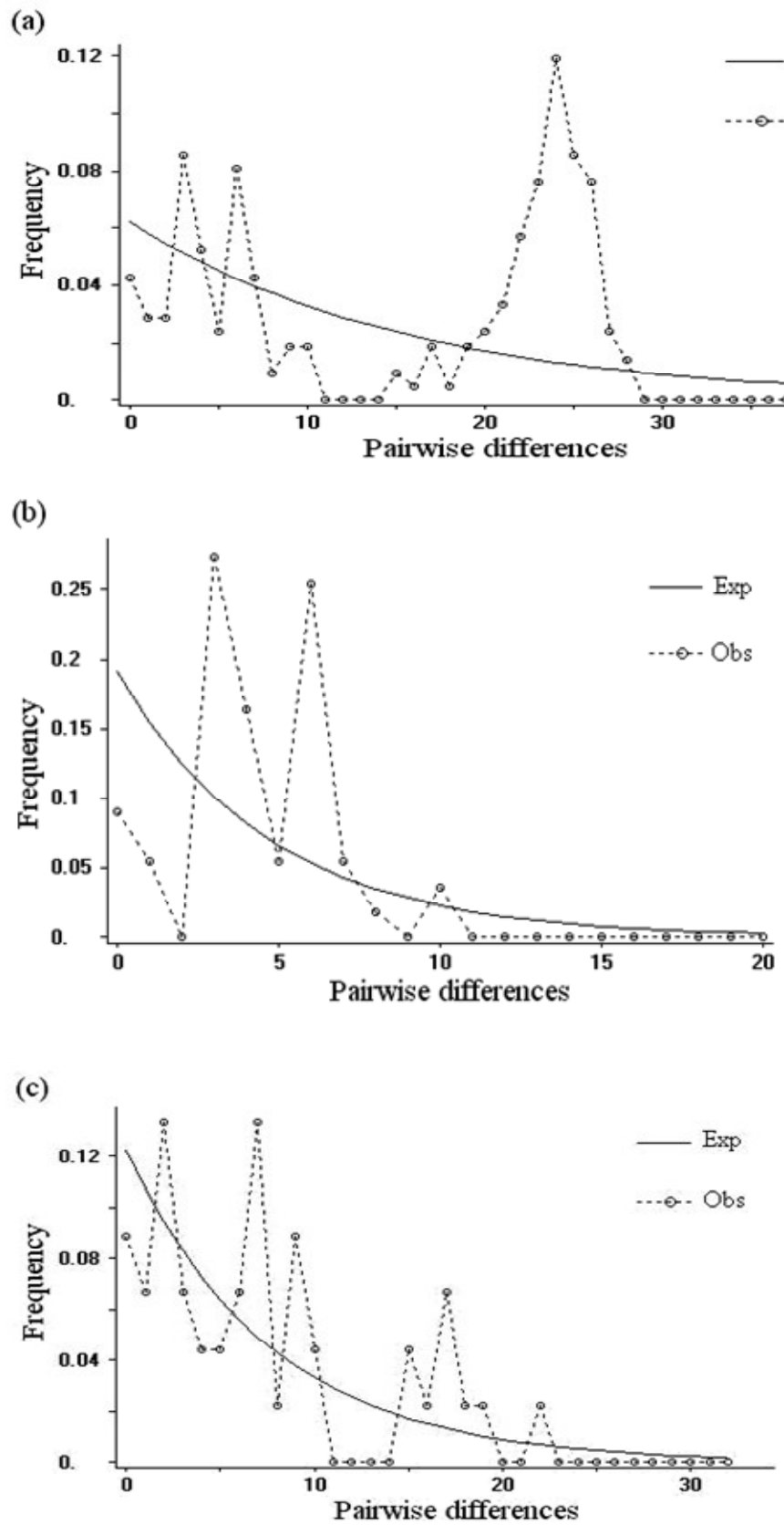


Figure 3. Expected (solid lines) and observed (broken lines) mismatch distribution for whole population and two lineages of wild yak: (a) whole population; (b) lineage A; (c) lineage B.

QN-11), the International Scientific and Technological Cooperation project of National Science and Technology Ministry (no. 2008DFA31100) and the National Natural Science Foundation of China (no.30860213/C120209).

REFERENCES

- Brown GG, Gadaleta G, Pepe G, Saccone C, Sbisà E (1986). Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. *J. Mol. Biol.* 192: 503-511. doi: 10.1016/0022-2836(86)90272-X.
- Excoffier L, Laval G, Schneider S (2005). Arlequin ver. 3.11: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*. 1: 47-50.
- Fu YX (1997). Statistical test of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147: 915-925.
- Guo SC, Savolainen P, Su JP, Zhang Q, Qi DL, Zhou J, Zhong Y, Zhao XQ, Liu JQ (2006). Origin of mitochondrial DNA diversity of domestic yaks. *BMC Evol. Biol.* 6: p. 73. doi: 10.1186/1471-2148-6-73.
- Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41: 95-98.
- Hu J (2001). Study on the biological characteristics and genetic diversities of wild yak and domestic yak and their hybrid. Master's dissertation, Gansu Agriculture University, Lanzhou, China.
- Kumar S, Tamura K, Nei M (2004). MEGA: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics*, 5: 150-163. doi:10.1093/bib/5.2.150.
- Li KL, Lu ZL (1990). The selected paper of yak scientific research. Lanzhou, China, pp. 79-207.
- Li JX, Qi DL, Zhang WX (1998). Study on the chromosome types of the wild yaks. *Gansu Anim. Vet. Sci.* 28: 8-10.
- Lai SJ, Wang L, Liu YP, Li XW (2005). Study on mitochondrial DNA genetic polymorphism of some yak breeds in China. *Acta Genetica Sinica*, 32: 463-470.
- Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003). DNASP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, 19: 2496-2497.
- Sambrook J, Russell DW (2001). *Molecular cloning: A laboratory manual*. New York, America.
- Slatkin M, Hudson RR (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, 129: 555-562.
- Tajima F (1989). Statistical methods for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123: 585-595.
- Wiener G, Han JL, Long RJ (2003). *The Yak*. Regional Office for Asia and the Pacific of the Food and Agriculture Organization of the United Nations, Bangkok, Thailand, pp. 9-11.
- Wang MQ (2000). Study on the partial characteristics of Datong yaks. PhD dissertation, Northwest A&F University, Yangling, China.
- Wolf C, Rentsch J, Hübner P (1999). PCR-RFLP analysis of mitochondrial DNA: a reliable method for species identification. *J. Agric. Food Chem.* 47(4): 1350-1355. doi: 10.1021/jf9808426.