Archives of Virology (2018) 163:659–670
https://doi.org/10.1007/s00705-017-3658-9

ORIGINAL ARTICLE

Molecular epidemiology and characterization of bovine leukemia virus in domestic yaks (*Bos grunniens*) on the Qinghai-Tibet Plateau, China

Meng Wang[1] · Yun Wang[2] · Abdul Rasheed Baloch[4] · Yangyang Pan[1] · Fang Xu[1] · Lili Tian[3] · Qiaoying Zeng[1]

Received: 13 August 2017 / Accepted: 9 November 2017 / Published online: 9 December 2017

Abstract

Bovine leukemia virus (BLV) is a member of the genus *Deltaretrovirus* of the family *Retroviridae* and cause a chronic lymphosarcoma, which is extensive in cattle. In yaks (*Bos grunniens*), the distribution, strains and genetic characteristics of BLV have rarely been studied. The aim of our study was to investigate BLV infections in domestic yaks and determine the genetic variability of BLV circulating in a region of the Qinghai Tibet Plateau, China. Blood samples were collected from 798 yaks, which were from different farms from Gansu, Qinghai and Sichuan provinces surrounding the Qinghai-Tibet Plateau. Nested PCR targeting BLV long terminal repeats was used to detect the BLV provirus. The highest prevalence of BLV infection was in Gansu province, where it was 18.93% (39/206) in white yaks from Tianzhu City and 19.14% (31/162) in black yaks from Gannan City. In Qinghai and Sichuan provinces, the prevalence of BLV in black yaks was 14.83% (35/236) and 14.94% (29/194), respectively. The prevalence of BLV was not significantly different in yaks up to one year old than in older animals. Phylogenetic analysis was performed using 16 different *env-gp51* (497-bp) gene sequences from the three provinces and 71 known BLV strains, which revealed that in both Gansu and Qinghai provinces, genotypes 6 and 10 of the BLV strains were at high levels, whereas only genotype 10 was prevalent in Sichuan Province. Phylogenetic analysis and sequence comparisons revealed 95.7-99.8% sequence identity among the full-length *env* genes of 16 strains, nearly full-length genome sequences of six BLV strains, and those of the known genotypes 6 and 10 of BLV. This study provides comprehensive information is regarding the widespread infection of domestic yaks with BLV on the Qinghai-Tibet Plateau of China, and shows that at least two BLV genotypes (genotypes 6 and 10) are circulating in this population.

Introduction

Bovine leukemia virus (BLV) is a member of the genus *Deltaretrovirus* in the family *Retroviridae* [1], and isolates of this virus have been used as a model for research on the pathogenesis of human T-cell leukemia virus type 1 (HTLV-1) [2]. BLV is considered the etiologic agent of enzootic bovine leucosis (EBL), and it infects cattle worldwide, making a severe economic impact on the beef and dairy cattle industries [3]. Most BLV-infected cattle are asymptomatic; however, approximately one-third of infected animals develop persistent lymphocytosis (PL) characterized by the polyclonal proliferation of $CD5^+$ B lymphocytes [3]. Only 1-5% of BLV-infected cattle show malignant monoclonal B-cell lymphosarcoma (LS) [4–6]. The symptoms/signs of BLV-infected cattle also include reduced milk production, weakness, digestive disturbance, loss of appetite, weight loss or general debility, and various neurological manifestations [7]. BLV infections are often overlooked because they are

Handling Editor: Diego G. Diel.

✉ Qiaoying Zeng
zengqy@gsau.edu.cn

Yangyang Pan
panyangyang_2007@126.com

1 College of Veterinary Medicine, Gansu Agricultural University, Lanzhou, China

2 Anning Branch Lanzhou General Hospital, Lanzhou, China

3 China Animal Health and Epidemiology Center, Qingdao, China

4 Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, Research Institute of Fish Culture and Hydrobiology, University of South Bohemia in České Budějovice, Zátiší 728/II, 389 25 Vodňany, Czech Republic

asymptomatic, resulting in a reservoir of infection for both horizontal and vertical transmission [8, 9].

The BLV genome consists of 8714 nucleotides and includes the essential genes *gag, pro, pol*, and *env*, which encode structural proteins and enzymes, the regulatory genes *tax* and *rex*, and the accessory genes *R3* and *G4*, and two identical long terminal repeats (LTRs) [10, 11]. Similar to other members of the family *Retroviridae*, the BLV *env* gene, encodes the envelope glycoprotein precursor $pr72^{env}$, which is important for viral infectivity and syncytium formation [12–14]. The *env* gene also encodes a mature surface glycoprotein (gp51) and a transmembrane protein (gp30) [13]. The glycoprotein gp51 plays a crucial role in the viral life cycle [15, 16]. It influences the capacity of BLV to enter cells and has been identified as a target of specific neutralizing antibodies [16, 17]. Three conformational epitopes are present in the N-terminal half of BLV gp51, F, G, and H [18], which are involved in viral infectivity and syncytium formation, whereas the C-terminal half contains the linear epitopes A, B, D, and E [18]. Therefore, the gp51 region has been widely used for BLV genotyping and phylogenetic studies to identify viral strains isolated in recent years [7, 19–21]. To date, at least ten genotypes of BLV have been identified based on the genetic characteristics of the envelope glycoprotein [7].

BLV has been detected worldwide in cattle, with different levels of prevalence and genotype distribution in different countries [19, 21]. In the past five years, many countries have reported BLV infection in cattle. The range of BLV prevalence is 5.3% - 87.8% in Thailand [20], 4.8% - 9.7% in the Philippines [5], 9.1% in Myanmar [7], and 42.3% - 77.4% in South America [21], with genotypes 1-8 documented, in addition to the novel genotypes 9 and 10, which were also discovered in South America [21] and Myanmar [7]. Previous studies have shown that the distribution of BLV genotypes is not always consistent with the country of origin; however, there is a general correlation between genotype distribution and geographic region [7, 21, 22]. In northern and northeastern China, the seroprevalence of BLV has been reported to be 18.29% in dairy and beef cattle [23] and 21.09% in yaks (*Bos grunniens*) in Gansu Province [24]. However, more information is needed on BLV genotypes in China, particularly regarding yaks on the Qinghai-Tibet Plateau.

Yaks are a unique bovine species that live at high altitudes. They are a valuable, semi-wild animal species of which 95% are distributed in the territory of China. Special white yaks live in Tianzhu City, Gansu Province [24]. Since their domestication at least 5000 years ago by ancestors of the present-day Chinese Tibetan people, yaks have lived exclusively in the cold highlands surrounding the Qinghai-Tibet Plateau (altitudes < 3000 m, average annual temperature < 0 °C) [25], including in the provinces of Qinghai, Gansu and Sichuan in northwestern and southwestern China. Because they live in a relatively secluded and very cold geographic region, only a few pathogens have been discovered in yaks [26, 27]. We have been involved in the surveillance of yak infections in recent years [25, 26] in a study that contributes to the health of animals and the Chinese Tibetan people.

In this study, we investigated the distribution of BLV strains in domestic yaks on the Qinghai-Tibet Plateau of China based on the amplification of BLV LTRs using a combination of BLV-CoCoMoqPCR-2 and nested PCR. The partial and full-length *env-gp51* sequences of 14 different Chinese BLV strains in yaks were used for phylogenetic analysis and compared with isolates from other geographical locations worldwide. The nearly full-length genome sequences of six Chinese BLV strains were obtained, and their genetic variability and genotypes were also analysed and compared with those of 16 whole BLV genome sequences from the NCBI database. This study is the first to identity the BLV genotypes in yaks of China.

Materials and methods

Experimental sample collection and extraction of genomic DNA

Blood samples were collected from 798 apparently healthy domestic yaks from 27 different farms in three provinces between March and December 2016 (Fig. 1 and Table 1). Farms located in the primary domestic-yak-raising areas were chosen for sampling in each province. The yaks were divided by age into ≤ 1 year and > 1 year and by species into white yak and black yak (Table 1). Genomic DNA was extracted from 50 µL of whole blood from each sample, using an E-Z 96® Blood DNA Kit (Omega, Norcross, GA, USA) according to the manufacturer's instructions. The extracted DNA was stored at – 20 °C until required for BLV detection.

Detection of BLV provirus by nested PCR targeting LTR regions

BLV provirus was detected by amplifying the BLV LTR regions in each DNA sample using nested PCR as described previously [5, 7, 28]. The first PCR amplification was performed with primers BLTR256F and BLTR453R (Fig. 2 and Table 2). As an internal control, the BoLA-DRA gene was amplified using primers BDRA488F and BDRA1145R, and in the nested PCR, using primers BLTR306F and BLTR408R. The same master mixes and thermal cycler conditions were used for both PCR and nested PCR. The reactions consisted of 5 picomoles of each primer, 10 µL of

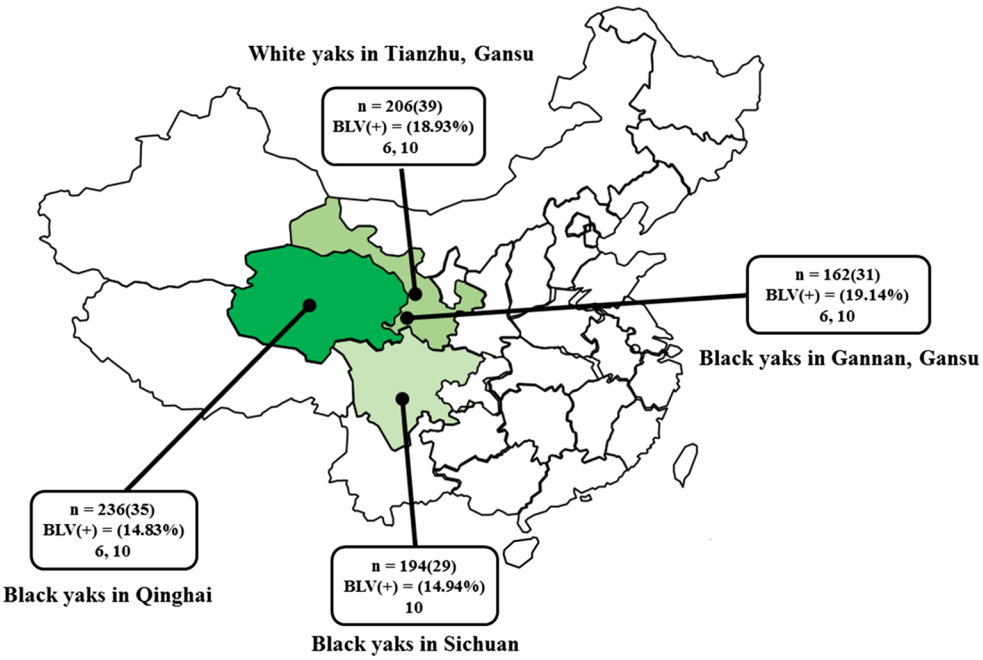

Fig. 1 Map of China showing the number and species of yak from provinces around the Qinghai-Tibet Plateau included in this study. The three provinces in which sampling was performed, yak species and occurrence of BLV genotypes are indicated in the figure. "n" indicates the total number of samples in each province or for each species, and BLV(+) % indicates the BLV prevalence rate

Table 1 Results of BLV detection by nested PCR for three provinces on the Qinghai Tibetan Plateau, China

Location (province)	City[a]	Breed[b]	Age (years)	No. tested[c]	No. positive (%)[d]	Genotype[e]	
						6 (%)	10 (%)
Gansu	Tianzhu	White yak	≤ 1	87	16 (18.39)	10 (62.50)	6 (37.50)
			> 1	119	23 (19.32)	18 (78.26)	5 (21.74)
	Gannan	Black yak	≤ 1	76	13 (17.11)	8 (61.54)	5 (38.46)
			> 1	86	18 (20.93)	12 (66.67)	6 (33.33)
Qinghai		Black yak	≤ 1	111	14 (12.61)	9 (64.29)	5 (35.71)
			> 1	125	21 (16.80)	17 (80.95)	4 (19.05)
Sichuan		Black yak	≤ 1	92	12 (13.04)	0 (0.00)	12 (100.00)
			> 1	102	17 (16.67)	0 (0.00)	17 (100.00)

[a]The cities of Tianzhu and Gannan in Gansu province where samples from white and black yaks were collected

[b]There are only black yaks in Qinghai and Sichuan province

[c]The number of yak samples collected from each province with two age groups in the study

[d]The number and percentage of positive samples in each age group from different provinces

[e]Genotypes based on gp51 sequences: number and percentage of samples in each column

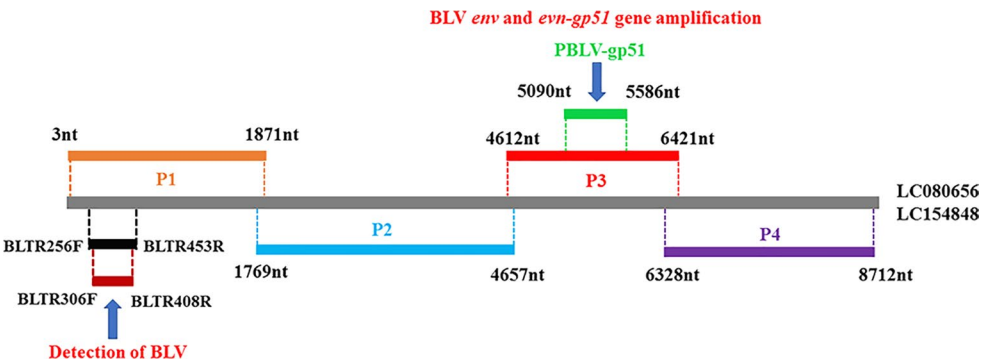

Fig. 2 Strategies and primers used to detect yak BLV and amplify *env* genes and full-length BLV genomes

Table 2 Primers used for detection and full-length genome amplification of BLV in yaks

Primer	Sequence	Binding position[a]	Length[b]	Purpose[c]	Reference
BLTR256-F	5′-GAGCTCTCTTGCTCCCGAGAC-3′	256-453	198	BLV detection	[22], [24]
BLTR453-R	5′-GAAACAAACGCGGGTGCAAGC CAG-3′				
BLTR306-F	5′-ACAACACCCCAAACACCAAT-3′	306-408	103		
BLTR408-R	5′-AGGAAGGGGAGGTAGTGGAA-3′				
BLV-F1	5′-TATGAAAGATCATGCCGGCCTAG-3′	3-1871	1869	Amplification of full-length BLV genomes	This study
BLV-R1	5′-CCTCACTAAGRGAATCTGTTATG-3′				
BLV-F2	5′-CCATTGGAAACGAGACTGTCC-3′	1769-4567	2889		
BLV-R2	5′-CCTCTGAAATGACAGCAAGTGG-3′				
BLV-F3	5′-AGATGGGAGCTACACCATTCA-3′	4612-6421	1810	Amplification of full-length BLV genomes and *env* gene	
BLV-R3	5′-AAGCCAGSGYTCCACCACCGC-3′				
BLV-F4	5′-CTCCCCAYTTCYCTGAAATCTCC-3′	6328-8712	2475	Amplification of full-length BLV genomes	
BLV-R4	5′-GTCTCTCCTGGCCGCTAGAG-3′				
BLV-gp51-F	5′-ATCTACTGGCCCCCCCCRCAAG-3′	5090-5586	497	BLV *env-gp51* gene detection and amplification	
BLV-gp51-R	5′-AAGGACGTGTTGACCCAGAAGAT-3′				

F, forward; R, reverse orientation; Y = C/T; K = G/T; R = A/G; S = C/G

[a]Nucleotide positions are according to the proviral genome sequences of bovine leukemia virus strains par62 (LC080656) and L1 (LC154848)

[b]The amplicon size is given in base pairs

2× PCR mix (Omega), and 6 μL of nuclease-free water, with 3 μL of DNA added to the PCR mixture and 3 μL of PCR product added to the nested PCR mixture. The final volume was 20 μL. The thermal cycle reactions were conducted in a thermocycler (Bio-Rad, USA) with initial denaturation at 94 °C for 5 min, followed by 35 cycles consisting of denaturation at 94 °C for 30 s, annealing at 58 °C for 30 s, and extension at 72 °C for 90 s. The last cycle was run at 72 °C for 5 min. PCR products were confirmed by electrophoresis in a 1.5% agarose gel stained with ethidium bromide and were visualized under ultraviolet light. Sterilized water was used as the PCR negative control.

Amplification and sequencing of partial BLV *env-gp51* and complete *env* genes

Samples that were BLV positive by nested PCR targeting LTR regions were used for amplification of partial BLV *env-gp*51 and complete *env* genes, which were also amplified by nested PCR using primers P3 and PBLV-gp51 (Fig. 2 and Table 2). The reaction mixture contained 7 μL (initial PCR) or 6 μL (second PCR) of distilled water, 10 μL of 2× PCR mix, 0.5 μL of each primer, and 2 μL of DNA (initial PCR) or initial PCR product (second PCR) in a final volume of 20 μL. Conditions for PCR amplification were as follows: 94 °C for 5 min, followed by 35 cycles of denaturation at 94 °C for 35 s, annealing at 58 °C for 30 s, and extension at 72 °C for 90 s (initial PCR) or 30 s (second PCR). The last cycle was run at 72 °C for 5 min. PCR with the external primers resulted in amplification of a 1810-bp DNA fragment that contained the complete *env* gene, and the internal primers amplified a 497-bp fragment of the gp51 region of the *env* gene.

Positive PCR products from each of the two rounds were purified using a QIAquick Gel Extraction Kit (QIAGEN, Germany) and sequenced (TaKaRa, Dalian, China). The sequences included a 1810-bp portion of the *env* gene and a 497-bp portion of the *env-gp51* gene, corresponding to nucleotide positions 4612-6421 and 5090-5589 of the proviral DNA sequence of bovine leukemia virus, strain par62 and L1, respectively (GenBank accession numbers LC080656 and LC154848) [7, 21]. Editing, alignment, and identification of nucleotide sequences were performed using MEGA 7.1 software [29].

Amplification and sequencing of the whole BLV provirus genome

Because the strains all showed high sequence similarity to one another based on analysis of the partial BLV *env-gp51* gene and the complete *env* gene, three strains of each genotype were selected randomly for subsequent sequencing. Four overlapping genomic fragments covering the complete BLV genome sequence were obtained by PCR amplification from six Chinese yak strains using the primers listed in Table 2 and Fig. 2. The final reaction mixture (25 μL) contained 12.5 μL of 2× PCR mix, 1 μL of each primer (each at 10 pmol), 3 μL of DNA template, and 7.5 μL of distilled

water. PCR amplification was performed as follows: 95 °C for 5 min, followed by 35 cycles of denaturation at 95 °C for 30 s, annealing at 60 °C for 20 s, and extension at 72 °C for 30 s per kilobase. The last cycle was run at 72 °C for 10 min. The four different BLV provirus genome PCR amplicons from each individual were purified using a QIAquick Gel Extraction Kit (QIAGEN, Germany) and sequenced (TaKaRa, Dalian, China). The sequences were assembled and edited using MEGA 7.1 [29] and DNAMAN 9.0 software to produce the final sequences of the viral genomes. The genome sequences included the essential structural and enzymatic *gag, pro, pol*, and *env* genes, regulatory genes *tax* and *rex*, and the accessory genes *R3* and *G4*. Only two nucleotides were lacking from each of the identical long terminal repeats (LTRs), which were located at the initiation and termination sites and did not influence the analysis of the structure and function of the BLV genome.

Phylogenetic analysis of the partial BLV *env-gp51* gene, the complete *env* gene, and the whole BLV provirus genome

The 16 different partial BLV *env-gp51* sequences and 16 complete *env* sequences from all positive samples from China were, aligned with 71 partial BLV *env-gp51* sequences and 26 complete *env* sequences, respectively, from GenBank, which were representative of the ten known BLV genotypes, using MEGA 7.1 software [29, 30]. Phylogenetic trees were constructed using the maximum-likelihood (ML) algorithm with the K2+G model of nucleotide substitution in MEGA 7.1. The reliability of the phylogenetic relationships was evaluated by nonparametric bootstrap analysis with 1000 replicates. The six complete genome sequences (8710 bp) of Chinese BLV strains were also aligned with 16 previously determined complete BLV sequences. Phylogenetic analysis of a 497-bp portion of BLV *env-gp51*, the complete *env* gene, and the whole BLV provirus genome was conducted using MEGA 7.1 [29]. Prediction of the protein sequences of the partial BLV env-gp51 genes through translation of nucleotide sequences to amino acid sequences was also performed using MEGA.

Results

The prevalence of BLV infection in domestic yaks on the Qinghai-Tibet Plateau of China

Blood samples from 798 yaks were tested for BLV by nested PCR targeting the LTR regions. In Gansu Province, the blood samples of yaks were collected from two different species in two different cities, Tianzhu and Gannan. Sixteen of 87 white yaks ≤ 1 year old (18.39%) and 23 of 119 white yaks > 1 year old (19.32%) tested positive for BLV provirus (Fig. 1 and Table. 1). In Gannan, the positive rates of BLV in black yaks were 17.11% (13/76) for yaks ≤ 1 year old and 20.93% (18/86) for yaks > 1 year old. Samples from 14 of 111 black yaks ≤ 1 year old (12.61%) and 21 of 125 black yaks > 1 year old (16.80%) from Qinghai Province were BLV positive. Of 194 samples collected in Sichuan, samples from 12 yaks ≤ 1 year old (12/92; 13.04%) and 17 yaks > 1 year old (17/102; 16.67%) were positive for BLV provirus.

Phylogenetic analysis and genotyping

In recent phylogenetic studies of the *env* gene of BLV strains isolated from cattle, this virus was classified into ten genotypes [5, 7, 19–21, 24]. Therefore, to gain insight into the number and frequency of genotypes of BLV strains in China, 16 different partial BLV env-gp51 sequences and 16 complete env sequences from all positive samples in China were aligned with 71 partial BLV *env-gp51* sequences and 26 complete *env* sequences, respectively, from GenBank. As shown in Figs. 3 and 4, BLVs in this study were classified into two closely related genotypes, genotype 10 and genotype 6. Both genotypes were detected in black and white yak samples from Gansu and Qinghai provinces; however, only genotype 10 was found from yaks in Sichuan Province. Notably, in Gansu and Qinghai provinces, genotype 10 was less prevalent than genotype 6 (Table 2).

Sequence comparison of BLV strains from China

To compare the sequences and determine the genotypes of the BLV strains, 16 amplified fragments including full-length *env* genomes and the nearly full-length BLV genomes of six strains were sequenced and deposited in the GenBank database (accession numbers MF574053-MF574068 and MF580990-MF580995). These sequences were used to construct an ML phylogenetic tree (Figs. 4 and 5). The nucleotide sequence similarity ranged from 98.6% to 99.8% among the full-length *env* genomes of the eight strains and the nearly full-length BLV genome sequences of the three strains in genotype 6 that were discovered in this study, and the similarity with corresponding genotype strains whose sequences were obtained from GenBank was from 97.4% to 99.8%. The nucleotide sequence similarity of BLV strains of genotype 10 ranged from 97.5% to 99.8%, and the similarity with strains of the corresponding genotype obtained from GenBank ranged from 95.7% to 99.0%.

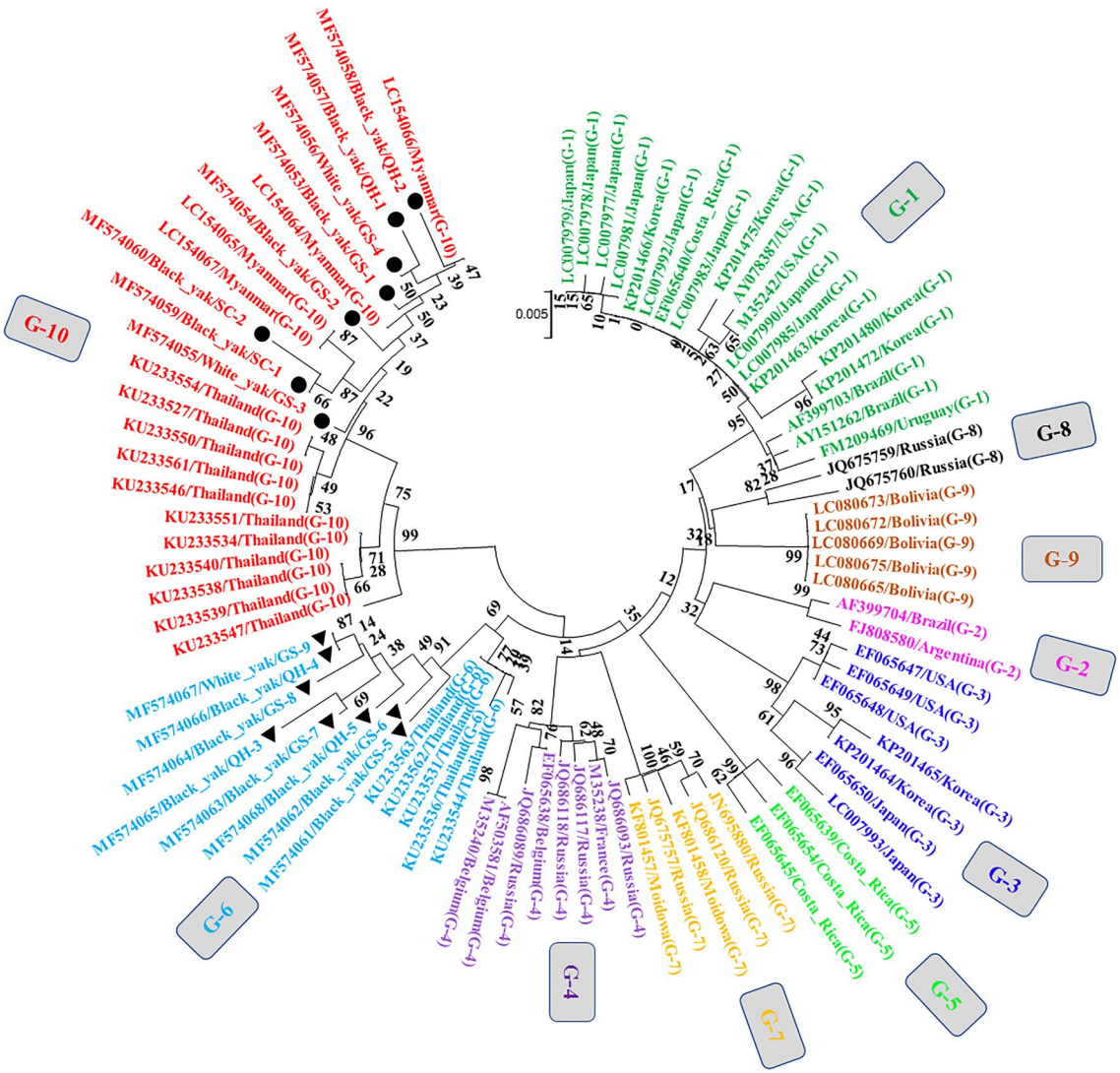

Fig. 3 Phylogenetic tree of partial BLV *env* sequences from different geographical locations worldwide. The phylogenetic tree was constructed from 16 distinct 497-bp BLV *env* sequences generated in this study and 71 sequences from known BLV strains, representing the ten different BLV genotypes. Bootstrap values expressed as percentages of 1000 replications are given at the branch points. The Chinese BLV strains identified in this study are indicated by GenBank accession numbers and breeds and strains. Other isolates are indicated by accession number and country of origin. The 16 newly identified yak BLV strains of genotype 10 and genotype 6 described in the present study (GenBank accession numbers MF574053-MF574068) are indicated by "●" and"▼", respectively. Genotypes are indicated by numbers around the circumference of the figure. Genotypes 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 are shown in dark green, pink, mazarine, violet, green, pastel blue, orange, black, black-yellow and red, respectively. The scale bar indicates nucleotide substitutions per site

Nucleotide and amino acid substitutions in BLV *env-gp51* from strains isolated in yak of China

Six nucleotide sequences of each genotype were selected and aligned with that of a reference sequence of the same genotype obtained from GenBank (Figs. 6A and 7A). The Chinese strains shared 16 nucleotide substitutions (nucleotides 359, 442, 481, 484, 493, 517, 542, 545, 548, 558, 576, 577, 600, 618, 630 and 717) with the genotype-10 strains from Myanmar, although some contained 14 nucleotide mutations (Fig. 7A). In genotype-6, 13 nucleotide substitutions (nucleotides 306, 390, 420, 439, 457, 465, 507, 516, 570, 618, 615, 717 and 727) were shared between Chinese strains and those from Paraguay, although these nucleotide mutations only occurred in some of the strains (Fig. 6A).

To identify the amino acid changes in the *env* proteins of the Chinese BLV strains, we aligned the deduced amino acid sequences of the eight partial env-gp51 sequences with those of representative genotype-10 sequences from Myanmar and genotype-6 sequences from Paraguay. As shown in Figs. 6B and 7B, amino acid substitutions occurred primarily in the middle region of gp51 of some of Chinese

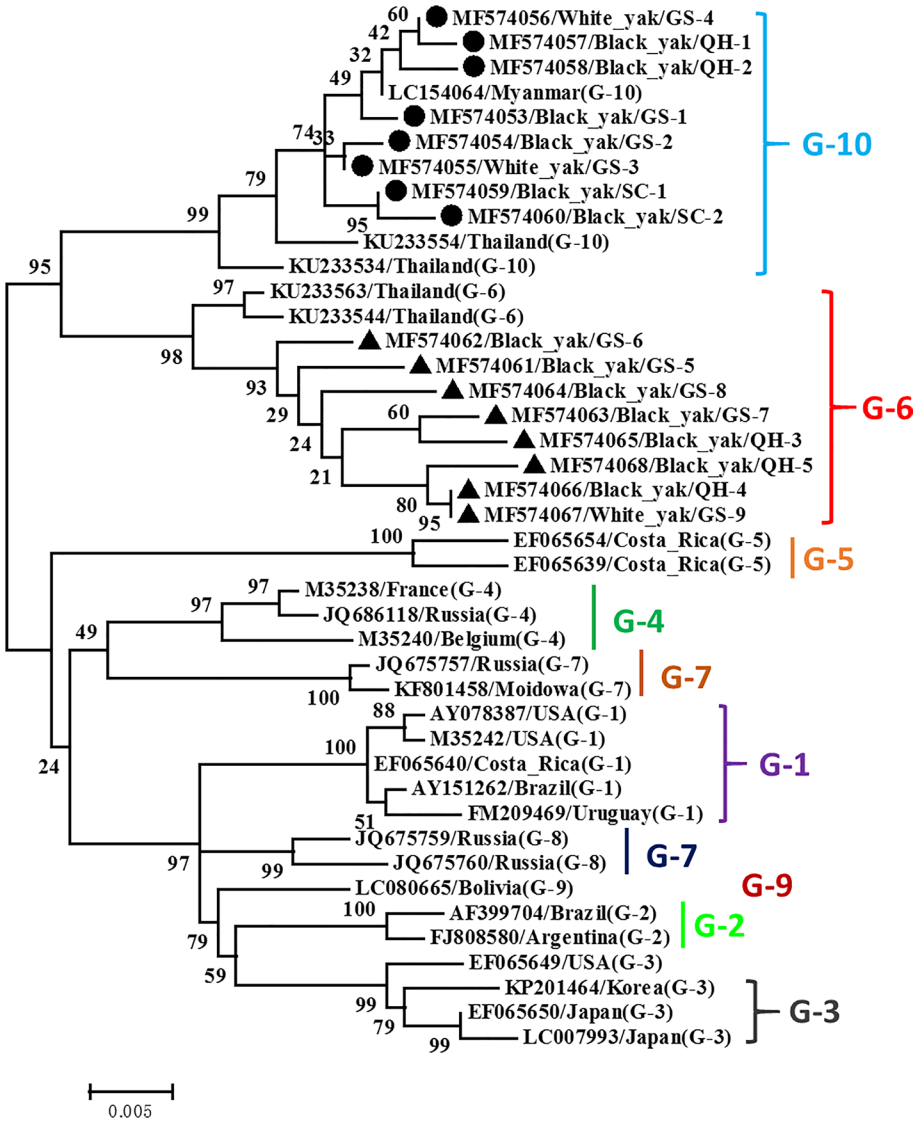

Fig. 4 Maximum-likelihood (ML) phylogenetic tree showing complete BLV *env* sequences from different geographical locations worldwide. The ML phylogenetic tree was constructed from complete (1448 bp) BLV *env* sequences from Chinese BLV strains (submitted to the GenBank nucleotide sequence database and assigned accession numbers MF574053-MF574068) and 26 sequences from known BLV strains (representing ten different BLV genotypes from different locations). These reference sequences were obtained from the GenBank nucleotide sequence database. The Chinese BLV strains identified in this study are indicated by GenBank accession numbers and breeds and strains. Other isolates are indicated by accession number and country of origin. The 16 newly identified yak BLV strains of genotype10 and genotype 6 described in the present study are indicated by "●" and"▼", respectively. Genotypes are indicated by numbers to the right of the figure with different colours. Genotype 10 (G-10) is highlighted in grey. The bar at the bottom of the figure denotes evolutionary distance

BLV strains. As expected, the Chinese strains had 10 amino acid substitutions that aligned with those of the Myanmar BLV strains, which were located in functional domains and included substitutions of histidine to arginine at residue 121 in the G-epitope region and histidine to lysine at residue 147 within the second neutralizing domain (2nd ND) region, eight substitutions (glutamic acid 160 glutamine, isoleucine 161 leucine, tyrosine 164 asparagine, arginine 182 glutamine, proline 183 histidine, aspartic acid 184 valine, glutamine 187 histidine, and proline 194 alanine) within the $CD8^+$ T-cell epitope and E-epitope regions, and substitution of glutamic acid to aspartic acid at residue 224 in the third neutralizing domain (3rd ND) region (Fig. 7B). Furthermore, a substitution of arginine to glutamine at residue 182 occurred in all Chinese genotype-10 strains when compared with BLV strains from Myanmar. By contrast, alignment with sequences from Paraguay revealed only two amino acid substitutions in the genotype 6 strains: histidine to tyrosine at residue 152 and glutamic to aspartic acid within the 3rd ND region (Fig. 6B).

Discussion

Four primary conclusions were reached based on the results of this study. First, the comprehensive investigation demonstrated widespread distribution of BLV in domestic yaks in provinces surrounding the Qinghai-Tibet Plateau of China. As determined by nested PCR targeting the LTR regions, a total of 798 samples collected from different provinces and cities located on the Qinghai-Tibet Plateau of China showed a relatively high rate of BLV infection compared with the infection rates in other countries [5, 7]. Second and most importantly, the genotypes of BLV

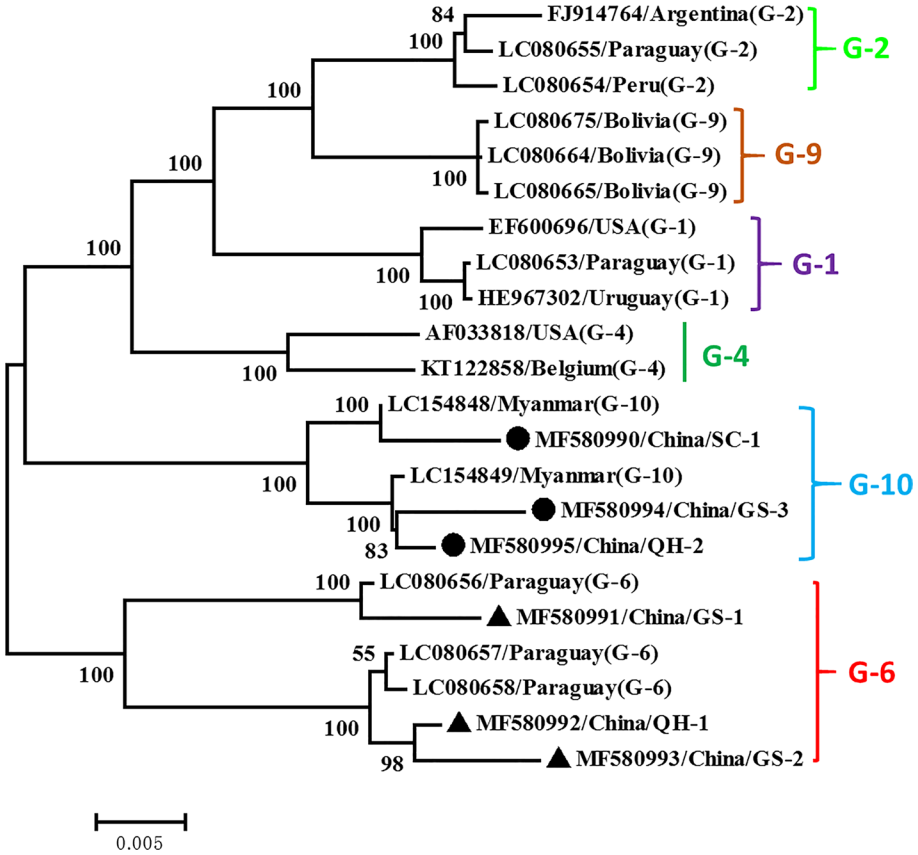

Fig. 5 Maximum-likelihood (ML) phylogenetic tree constructed from complete BLV genomic sequences. The ML phylogenetic tree was constructed using the complete BLV genomic sequences from six Chinese BLV strains (submitted to the GenBank nucleotide sequence database and assigned accession numbers MF580990-MF580995) and 16 reference sequences obtained from the GenBank nucleotide sequence database. One thousand replications were performed to calculate bootstrap values (indicated on the tree). The strains identified in this study are indicated by GenBank accession numbers and breeds and strains. Reference sequences are indicated by accession number and country of origin. The 16 newly identified yak BLV strains of genotype 10 and genotype 6 described in the present study are indicated by "●" and"▼", respectively. Genotypes are indicated by numbers to the right of the figure with different colours. The bar at the bottom of the figure denotes evolutionary distance

in yaks were identified for first the time in China, and a phylogenetic analysis based on partial and complete *env-gp51* sequences revealed that Chinese BLV strains clustered with genotype 10 and genotype 6 of the ten distinct BLV genotypes found worldwide [7, 20–22, 31, 32]. Third, an ML phylogenetic tree based on 497 bp of the BLV *env* sequence clearly showed that Chinese BLV strains in yaks belonged to other genotypes, including a new genotype found in Southeast Asia (Myanmar and Thailand) (Fig. 3), which is very close to China [7, 20]. Finally, several nucleotide and amino acid substitutions were discovered in the full BLV genome sequences and in the *gp51* gene sequences of the Chinese strains, particularly for genotype 10. Additionally, most of the substitutions were observed in structural genes, such as the $CD8^+$ T-cell epitope, the second ND and third ND, and the G and E epitopes, and the substitutions varied according to genotype, which was similar to the genetic characteristics of BLV in the Philippines [5].

The BLV infection rate was 16.79% (134/798) in the yaks surrounding the Qinghai-Tibet Plateau of China as determined by PCR. In neighbouring countries, the percentage of BLV infection is 5.3% to 87.8% in Thailand [20], 9.1% in Myanmar[7], and 4.8% to 9.7% in Philippines [5]. In the present study, the percentage of BLV infection in Chinese yaks was lower than that in cattle of other countries such as Korea (35%) [33] and Tanzania (36%) [34] but higher in Chinese yaks than the 9.1% in Myanmar and the 4.8% to 9.7% in the Philippines. However, the prevalence of BLV in yaks was significantly lower than that in cattle in South America (42.3% to 77.4%) [21]. In China, the seroprevalence of BLV is 18.29% in dairy and beef cattle [23], whereas in Gansu Province, the prevalence in yaks was 21.09%. Furthermore, 24.26% of black yaks and 19.10% white yaks tested positive

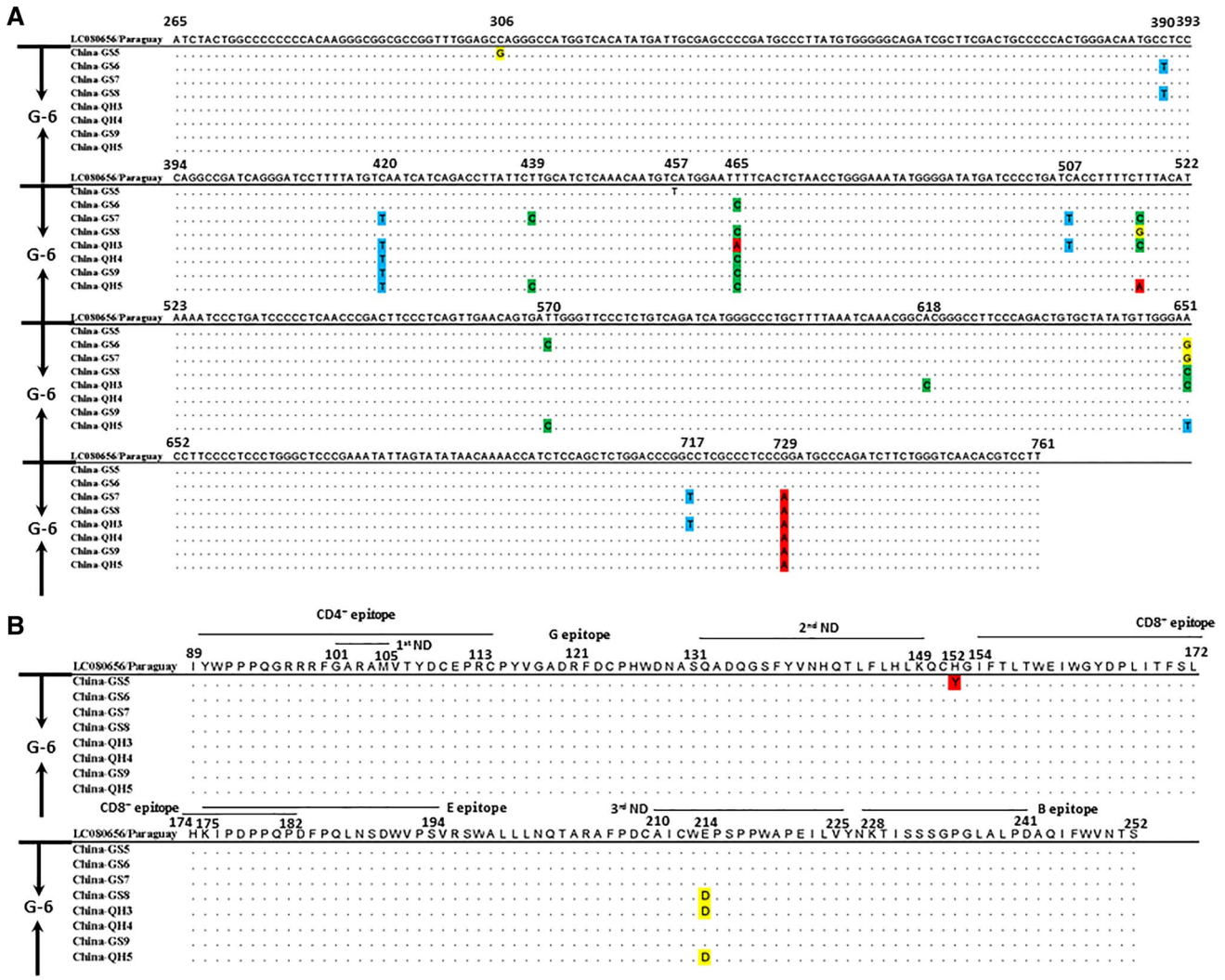

Fig. 6 Alignment of a partial nucleotide sequences and deduced amino acid sequences of the BLV *env* gene from strains of genotype 10 in China. Alignments were performed for eight typical nucleotide sequences (A) and eight unique deduced amino acid sequences (B) of the *env* gene from 134 Chinese BLV strains. Chinese BLV strains are identified by the sample ID at the left side of the figure. The numbers indicate the location of substitutions. The first, second and third neutralizing domains (ND) and other epitopes are shown at the top of the alignment in B. Numbers above the sequences are amino acid residue numbers that indicate the start and end of each domain. Genotypes are indicated by the black bars at the far left of the figure. The strain with accession number LC154849 from Myanmar was used as a reference

in an earlier study [24], which were higher values than those we reported in this study. The percentage of BLV infection in black and white yaks was 18.93% and 19.14%, respectively, and the difference was not significant, which is consistent with the results of serological detection [24]. Based on the methods of nested PCR in this study and serological detection in a previous study [24], the differences of BLV prevalence between yaks > 1 year old and yaks ≤ 1 year old were also not significant. Another notable result was that Gansu showed the highest prevalence of BLV infection, including 18.93% in white yaks from Tianzhu City and 19.14% in black yaks from Gannan City. The prevalence of BLV in black yaks from Qinghai and Sichuan was 14.83% and 14.94%, respectively. These results indicate that age and breed are not risk factors for the presence of BLV in yaks; however, differences in the percentage of BLV infection are likely to occur between provinces and locations within the same country.

The most interesting data in this study were from the molecular characterization of BLV in yaks confirmed by phylogenetic analysis using six new and 16 previously reported BLV whole genome sequences (Fig. 5), which were consistent with those of phylogenetic analysis of 16 partial *env-gp51* sequences and full-length *env* sequences obtained by the ML method (Figs. 3 and 4). Two genotypes (genotype 10 and genotype 6) were identified in BLV

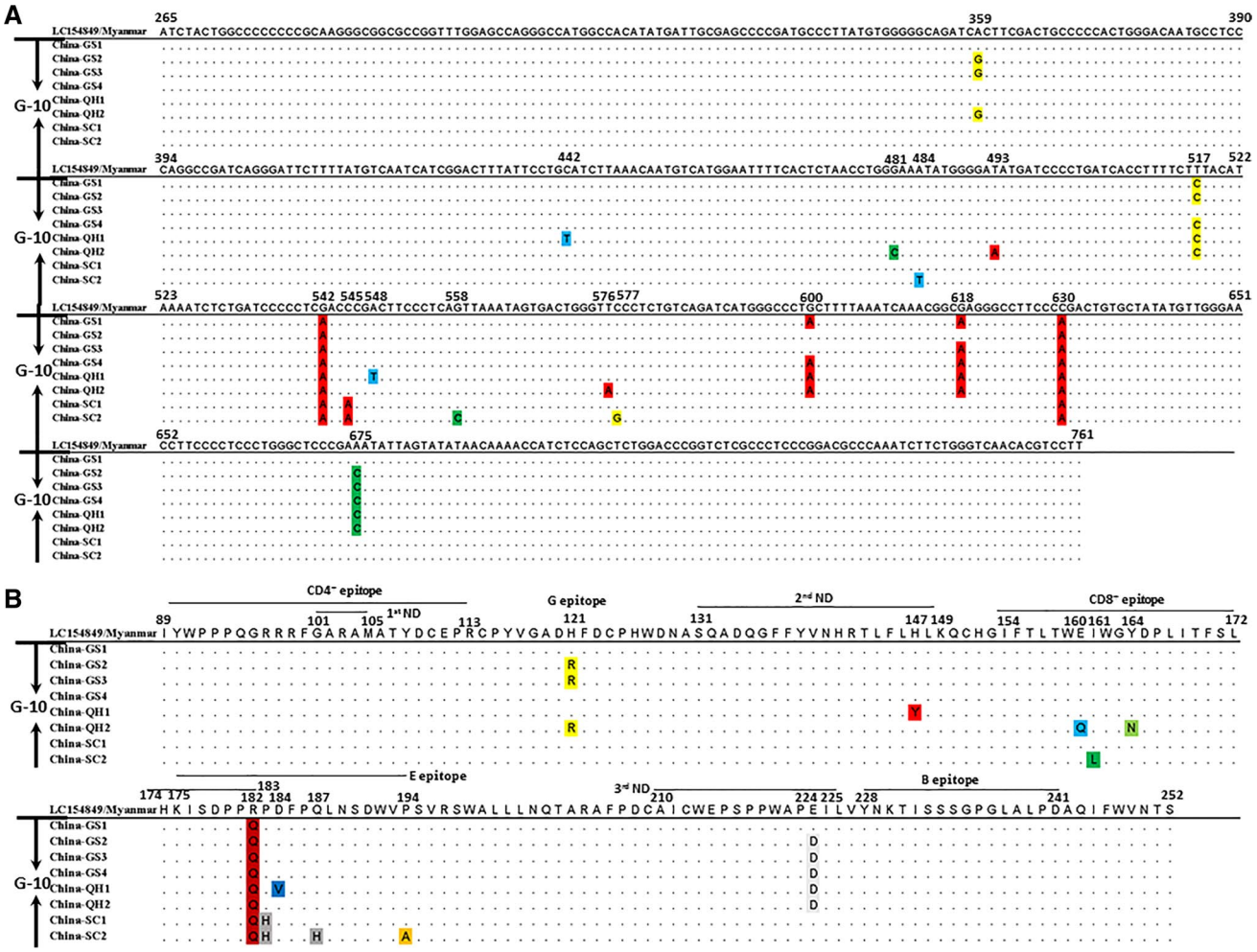

Fig. 7 Alignment of a partial nucleotide sequence and deduced amino acid sequence of the BLV *env* gene from strains of genotype 6 in China. Alignments were performed for eight typical nucleotide sequences (A) and eight unique deduced amino acid sequences (B) of the *env* gene from 134 Chinese BLV strains. Chinese BLV strains are identified by the sample ID at the left side of the figure. The numbers indicate the locations of substitutions. The first, second and third neutralizing domains (ND) and other epitopes are shown at the top of the alignment in B. Numbers above the sequences are amino acid residue numbers that indicate the start and end of each domain. Genotypes are indicated by the black bars at the far left of the figure. The strain with accession number LC080656 from Paraguay was used as a reference

strains from yaks in China. The presence of more than one genotype in the same herd has been observed previously in open herds [35]. The genotype 10 BLV strains in China and the strains from Myanmar, with the Chinese BLV strains identified in samples located close to one another, were detected in all three provinces in this study. By contrast, genotype 6 BLV strains were detected in Gansu and Qinghai provinces but were not found in Sichuan Province, and these strains were also located close to the strains from Thailand. Genotype 10 is a new genotype of BLV that was discovered in recent years [7, 20], and it was detected in all provinces surrounding the Qinghai-Tibet Plateau of China. Collectively, our study is the first to show that BLV strains isolated from yaks in China were most similar to BLV strains from Myanmar and Thailand and that the infection of yaks with BLV might have occurred in recent years.

Our analysis of partial of BLV env-gp51 sequences identified 16 nucleotide substitutions in genotype 10; six of which were silent substitutions and 11 of which resulted in amino acid substitutions. Thirteen nucleotide substitutions were identified in genotype 6 among the Chinese strains, two of which resulted in amino acid substitutions. More substitutions were found in genotype 10 than in genotype 6, and these substitutions also occurred in the same genotype found in this study. This result indicated that the sequences of the genotype-6 BLV strains in yaks were more conserved than those of the genotype-10 BLV strains. Furthermore, almost all substitutions were located in functional epitopes or NDs,

which is consistent with a previous finding that most substitutions in env-gp51 are found within epitopes rather than at random locations [17, 22, 33, 36, 37]. Amino acid substitutions in the second ND could affect the interaction between gp51 and a receptor expressed on host-cell membranes, suggesting that these substitutions could affect viral fusion and infectivity *in vivo* [38]. The $CD4^+$ epitope, $CD8^+$ epitope, E epitope and G epitope are conformational epitopes and are targets for monoclonal antibodies that induce neutralization and inhibit syncytium formation [38, 39]. Therefore, the biological effects of these substitutions should be investigated.

The present study reveals the wide prevalence of BLV infection in yaks surrounding the Qinghai-Tibet Plateau of China, and the molecular characteristics of BLV in yaks were determined for the first time. BLV genotypes 10 and 6 were prevalent in domestic yaks in China. These results will be very important for clarifying the source, circulation pattern, zoonotic potential, and public health risk of this virus. The mechanisms of the pathogenicity, transmission, evolution, and persistence of this virus also require urgent clarification.

Compliance with ethical standards

Funding This work was supported by a Grant from the National Natural Science Foundation of China (no. 31260616); Fuxi Foundation of Exceptional Talent at Gansu Agricultural University; Scientific Research Foundation for the New Scholars, Gansu Agricultural University (no. GSAU-RCZX201702) and the Innovation Foundation of the College of Veterinary Medicine, Gansu Agriculture University (no. JYCX-KX017).

Conflict of interest The authors declare that there are no competing interests regarding the publication of this paper.

Ethical approval All animals were handled with the assistance of veterinarians from the local veterinary institute and animal health and epidemiology centre. All procedures in this study were approved in strict accordance with good animal practice following the guidelines of the Animal Care and Use Committee of Gansu Agricultural University and performed in accordance with animal welfare and ethics.

References

1. Kettmann R, Portetelle D, Mammerickx M, Cleuter Y, Dekegel D, Galoux M, Ghysdael J, Burny A, Chantrenne H (1976) Bovine leukemia virus: an exogenous RNA oncogenic virus? Modern trends in human leukemia II. Springer, Berlin, pp 375–389
2. Willems L, Burny A, Collete D, Dangoisse O, Dequiedt F, Gatot J-S, Kerkhofs P, Lefebvre L, Merezak C, Peremans T (2000) Genetic determinants of bovine leukemia virus pathogenesis. AIDS Res Hum Retroviruses 16(16):1787–1795
3. Moratorio G, Obal G, Dubra A, Correa A, Bianchi S, Buschiazzo A, Cristina J, Pritsch O (2010) Phylogenetic analysis of bovine leukemia viruses isolated in South America reveals diversification in seven distinct genotypes. Arch Virol 155(4):481–489
4. Burny A, Cleuter Y, Kettmann R, Mammerickx M, Marbaix G, Portetelle D, Van den Broeke A, Willems L, Thomas R (1988) Bovine leukaemia: facts and hypotheses derived from the study of an infectious cancer. Vet Microbiol 17(3):197–218
5. Polat M, Ohno A, S-n Takeshima, Kim J, Kikuya M, Matsumoto Y, Mingala CN, Onuma M, Aida Y (2015) Detection and molecular characterization of bovine leukemia virus in Philippine cattle. Arch Virol 160(1):285–296
6. Llames L, Goyache J, Domenech A, Arjona A, Suarez G, Gomez-Lucia E (2001) Evaluation of virus excretion by cells persistently infected with the bovine leukaemia virus (BLV) using monoclonal antibodies. J Clin Virol 22(1):31–39
7. Polat M, Moe HH, Shimogiri T, Moe KK, S-n Takeshima, Aida Y (2017) The molecular epidemiological study of bovine leukemia virus infection in Myanmar cattle. Arch Virol 162(2):425–437
8. Schwartz I, Levy D (1994) Pathobiology of bovine leukemia virus. Vet Res 25(6):521–536
9. Gutiérrez G, Alvarez I, Politzki R, Lomónaco M, Santos MJD, Rondelli F, Fondevila N, Trono K (2011) Natural progression of bovine leukemia virus infection in Argentinean dairy cattle. Vet Microbiol 151(3):255–263
10. Aida Y, Murakami H, Takahashi M, Takeshima S-N (2013) Mechanisms of pathogenesis induced by bovine leukemia virus as a model for human T-cell leukemia virus. Front Microbiol 4:328
11. Gillet N, Florins A, Boxus M, Burteau C, Nigro A, Vandermeers F, Balon H, Bouzar A-B, Defoiche J, Burny A (2007) Mechanisms of leukemogenesis induced by bovine leukemia virus: prospects for novel anti-retroviral therapies in human. Retrovirology 4(1):18
12. Mamoun R, Astier T, Guillemain B, Duplan J (1983) Bovine lymphosarcoma: expression of BLV-related proteins in cultured cells. J Gen Virol 64(9):1895–1905
13. Zarkik S, Decroly E, Wattiez R, Seidah NG, Burny A, Ruysschaert J-M (1997) Comparative processing of bovine leukemia virus envelope glycoprotein gp72 by subtilisin/kexin-like mammalian convertases. FEBS Lett 406(1–2):205–210
14. Bai L, Otsuki H, Sato H, Kohara J, Isogai E, S-n Takeshima, Aida Y (2015) Identification and characterization of common B cell epitope in bovine leukemia virus via high-throughput peptide screening system in infected cattle. Retrovirology 12(1):106
15. Callebaut I, Voneche V, Mager A, Fumiere O, Krchnak V, Merza M, Zavada J, Mammerickx M, Burny A, Portetelle D (1993) Mapping of B-neutralizing and T-helper cell epitopes on the bovine leukemia virus external glycoprotein gp51. J Virol 67(9):5321–5327
16. Johnston ER, Radke K (2000) The SU and TM envelope protein subunits of bovine leukemia virus are linked by disulfide bonds, both in cells and in virions. J Virol 74(6):2930–2935
17. Mamoun R, Morisson M, Rebeyrotte N, Busetta B, Couez D, Kettmann R, Guillemain B (1990) Sequence variability of bovine leukemia virus env gene and its relevance to the structure and antigenicity of the glycoproteins. J Virol 64(9):4180–4188
18. Bruck C, Mathot S, Portetelle D, Berte C, Franssen J-D, Herion P, Burny A (1982) Monoclonal antibodies define eight independent antigenic regions on the bovine leukemia virus (BLV) envelope glycoprotein gp51. Virology 122(2):342–352
19. Khudhair YI, Hasso SA, Yaseen NY, Al-Shammari AM (2016) Serological and molecular detection of bovine leukemia virus in cattle in Iraq. Emerg Microb Infect 5(6):e56
20. Lee E, Kim E-J, Ratthanophart J, Vitoonpong R, Kim B-H, Cho I-S, Song J-Y, Lee K-K, Shin Y-K (2016) Molecular epidemiological and serological studies of bovine leukemia virus (BLV) infection in Thailand cattle. Infect Genet Evol 41:245–254
21. Polat M, S-n Takeshima, Hosomichi K, Kim J, Miyasaka T, Yamada K, Arainga M, Murakami T, Matsumoto Y, Diaz VB (2016) A new genotype of bovine leukemia virus in South America

identified by NGS-based whole genome sequencing and molecular evolutionary genetic analysis. Retrovirology 13(1):4
22. Rodriguez SM, Golemba MD, Campos RH, Trono K, Jones LR (2009) Bovine leukemia virus can be classified into seven genotypes: evidence for the existence of two novel clades. J Gen Virol 90(11):2788–2797
23. Sun W-W, Lv W-F, Cong W, Meng Q-F, Wang C-F, Shan X-F, Qian A-D (2015) *Mycobacterium avium* subspecies paratuberculosis and bovine leukemia virus seroprevalence and associated risk factors in commercial dairy and beef cattle in northern and northeastern China. BioMed Res Int 2015:315173. https://doi.org/10.1155/2015/315173
24. Ma J-G, Zheng W-B, Zhou D-H, Qin S-Y, Yin M-Y, Zhu X-Q, Hu G-X (2016) First report of bovine leukemia virus infection in yaks (*Bos mutus*) in China. BioMed Res Int 2016:9170167. https://doi.org/10.1155/2016/9170167
25. Xu F, Pan Y, Wang M, Wu X, Tian L, Baloch AR, Zeng Q (2016) First detection of ungulate tetraparvovirus 1 (bovine hokovirus 1) in domestic yaks in northwestern China. Arch Virol 161(1):177–180
26. Xu F, Pan Y, Baloch AR, Tian L, Wang M, Na W, Ding L, Zeng Q (2014) Hepatitis E virus genotype 4 in yak, northwestern China. Emerg Infect Dis 20(12):2182
27. Zhu W, Dong J-B, Zhang J, Uchida K, K-i Watanabe, Goto Y, Haga T (2013) *Bos grunniens* papillomavirus type 1: a novel deltapapillomavirus associated with fibropapilloma in yak. J Gen Virol 94(1):159–165
28. Tajima S, Ikawa Y, Aida Y (1998) Complete bovine leukemia virus (BLV) provirus is conserved in BLV-infected cattle throughout the course of B-cell lymphosarcoma development. J Virol 72(9):7569–7576
29. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10):2731–2739
30. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4):406–425
31. Balić D, Lojkić I, Periškić M, Bedeković T, Jungić A, Lemo N, Roić B, Čač Ž, Barbić L, Madić J (2012) Identification of a new genotype of bovine leukemia virus. Arch Virol 157(7):1281–1290
32. Matsumura K, Inoue E, Osawa Y, Okazaki K (2011) Molecular epidemiology of bovine leukemia virus associated with enzootic bovine leukosis in Japan. Virus Res 155(1):343–348
33. Suh GH, Lee JC, Lee CY, Hur TY, Son DS, Ahn BS, Kim NC, Lee CG (2005) Establishment of a bovine leukemia virus-free dairy herd in Korea. J Vet Sci 6(3):227–230
34. Schoepf K, Kapaga A, Masami H, Hyera J (1997) Serological evidence of the occurrence of enzootic bovine leukosis (EBL) virus infection in cattle in Tanzania. Trop Anim Health Prod 29(1):15–19
35. Licursi M, Inoshima Y, Wu D, Yokoyama T, González ET, Sentsui H (2002) Genetic heterogeneity among bovine leukemia virus genotypes and its relation to humoral responses in hosts. Virus Res 86(1):101–110
36. Portetelle D, Couez D, Bruck C, Kettmann R, Mammerickx M, Van Der Maaten M, Brasseur R, Burny A (1989) Antigenic variants of bovine leukemia virus (BLV) are defined by amino acid substitutions in the NH2 part of the envelope glycoprotein gp51. Virology 169(1):27–33
37. Zhao X, Buehring GC (2007) Natural genetic variations in bovine leukemia virus envelope gene: possible effects of selection and escape. Virology 366(1):150–165
38. Gatot J-S, Callebaut I, Van Lint C, Demonté D, Kerkhofs P, Portetelle D, Burny A, Willems L, Kettmann R (2002) Bovine leukemia virus SU protein interacts with zinc, and mutations within two interacting regions differently affect viral fusion and infectivity in vivo. J Virol 76(16):7956–7967
39. Bruck C, Portetelle D, Burny A, Zavada J (1982) Topographical analysis by monoclonal antibodies of BLV-gp51 epitopes involved in viral functions. Virology 122(2):353–362