



Phylogenetic analysis of the partial sequences of the *env* and *tax* BLV genes reveals the presence of genotypes 1 and 3 in dairy herds of Antioquia, Colombia

Cristina Úsuga-Monroy¹ · F. J. Díaz² · Luis Gabriel González-Herrera¹ · José Julián Echeverry-Zuluaga¹ · Albeiro López-Herrera¹

Received: 15 March 2023 / Accepted: 7 August 2023

© The Author(s) 2023

Abstract

Bovine leukemia virus (BLV) is a retrovirus that primarily infects dairy cows. Although few studies have also used the *tax* gene, phylogenetic studies of BLV use mostly the *env* gene. The aim of this work was to establish the circulating genotypes of BLV in specialized dairy cattle from Antioquia, Colombia. Twenty blood samples from Holstein Friesian cows were collected, and their DNA was isolated. A PCR was performed for a partial region of the *env* and *tax* genes. A phylogenetic analysis was carried out using the maximum likelihood and Bayesian methods for both genes. Nineteen sequences were identified as genotype 1 by *env* and *tax* genes. Only one sequence was clustered with genotype 3 and had the highest proportion of different nucleotide sites compared to other strains. Four amino acid substitutions in the 134 amino acid residue fragment of the Env protein were identified in the Colombian sequences, and three new amino acid substitutions were reported in the 296 amino acid residue fragment of the Tax protein. R43K (Z finger), A185T (Activation domain), and L105F changes were identified in the genotype 3 sample. This genotype has been reported in the United States, Japan, Korea, and Mexico, but so far, not in Colombia. The country has a high rate of imported live animals, semen, and embryos, especially from the United States. Although it is necessary to evaluate samples from other regions of the country, the current results indicate the presence of two BLV genotypes in specialized dairy herds.

Keywords Bovine leukemia · Dairy herds · Holstein Friesian · Strain · Virus

Introduction

The etiologic agent of enzootic bovine leukemia (EBL) is the bovine leukemia virus (BLV) that belongs to the *Deltaretrovirus* genus and the *Retroviridae* family. BLV infects bovine B cells and compromises the humoral immune response. Among 30 to 70% of infected cows have increased blood lymphocyte count; between 0.1 and 10%

of the cows with more than three years of infection develop the tumor phase or lymphosarcomas [19, 32]. The BLV genome has a size of 8714 base pairs and has a *pX* region flanked by two identical long terminal repeats (LTRs), in addition to the *gag*, *pro*, *pol*, and *env* gene regions [1, 28]. The *gag* gene codes for p15, p24, and p12 proteins. The p15 or matrix protein, has 109 amino acids; the p12 or nucleocapsid protein, has 69 amino acids and is strongly bound to the viral RNA; and the virus capsid consists mainly of the p24 protein. The *pol* gene codes for integrase and reverse transcriptase [14, 28]. The *env* gene codes for the gp51 surface glycoprotein and the gp30 transmembrane protein; both proteins are glycosylated polypeptides and are associated through disulfide bridges forming a stable complex. The gp51 protein plays an essential role in the life cycle of the virus since it requires to enter the cell, induces the mass production of neutralizing antibodies in infected animals, and is essential for the infectivity of the virus [4, 14]. Moreover, the N-terminal of the gp51 protein

✉ Albeiro López-Herrera
alherrera@unal.edu.co

Cristina Úsuga-Monroy
cusugam@unal.edu.co

¹ Grupo BIOGEM, Facultad de Ciencias Agrarias, Universidad Nacional de Colombia Sede Medellín, Calle 65 No 59A-110, Medellín, Colombia

² Grupo Inmunovirología, Facultad de Medicina, Universidad de Antioquia, Calle 70 No. 52-21, Medellín, Colombia

has three conformational epitopes (F, G, and H) associated with viral infectivity and syncytia formation [4].

The pX region of the BLV genome contains the genes *tax*, *rex*, *G4*, and *R3* related to regulatory viral transcription functions. The *tax* gene is a transcriptional transactivator that increases transcription factor binding to DNA and upregulates the synthesis of viral proteins. The tax protein acts on the enhancer motif Tax-response element (TxRE) located in the 5'LTR's U3 region of the BLV genome to stimulate viral transcription [2, 43]. The tax protein has oncogenic potential, inhibits DNA repair, increases DNA damage by the accumulation of mutations [14], contributes to the induction of lymphomas in mouse fibroblasts through cooperation with the oncoprotein Ha-Ras [33, 37], modulates the expression of genes related to cell growth, affects the proliferation of B lymphocytes, cell survival and the independent growth of cytokines in B cells [35].

A phylogenetic tree is an ideal approach for reconstructing the evolutionary history of any organism on earth and understanding adaptive evolution at the molecular level. Sometimes, the mutational changes of DNA sequences determine changes in protein synthesis that promote genetic variants. The *env* gene is the most used for phylogenetic studies of BLV, and its characterization has resulted in a classification of BLV in 12 genotypes [34]. According to Nishikaku et al. [24], after the first appearance of BLV in Asia, the progenitor of lineage I reached Europe in 1824, and from there, it reached South America in 1850. From South America, it passed to the United States and Mexico in 1942 [24]. BLV differences at the nucleotide level have been related to geographic location; therefore, the geographic component plays an essential role in BLV distribution worldwide. Nucleotide differences in many *env* sites are related to the geographical location of the strain. Genotypes 7 and 8 have been registered only in Europe, genotype 9 only in Bolivia [29], genotype 10 only in Thailand [18], and the new genotype 12 in Kazakhstan [34]. However, genotypes 1 and 4 are present on several continents. BLV is widely distributed in the Americas. Genotypes 1 and 5 are found in the United States, genotypes 1 and 5 in Costa Rica [42], genotypes 1, 2, and 6 in Brazil [6], genotype 7 in Chile [12], genotype 1 in Uruguay [23], and genotype 9 in Bolivia [29]. On the other hand, reports indicate that the *tax* gene is used to confirm the presence of BLV infection [10, 27] and is also used in BLV phylogenetic studies [8, 9]. The *tax* gene is more divergent than the *pol* gene, therefore its sequence could provide significant information for virus characterization [9]. Accordingly, the aim of this work was to establish the circulating genotypes of BLV in the specialized dairy cattle of Antioquia, Colombia using partial sequences of the *env* and *tax* genes through two phylogenetic inference methods.

Materials and methods

Samples

Blood samples were collected from 20 Holstein Friesian cows between the first and fifth lactation period and from 3 to 7 years of age. Sampling was carried out from February to June 2017. Cows naturally infected with BLV were previously diagnosed by ELISA (SVANOVIR® BLV gp51-Ab) for the gp51 protein and by PCR [40] for the envelope gene. None of the selected cattle showed visible symptoms associated with BLV and were milking. These animals came from 10 municipalities of Antioquia divided into three sub-regions: North (Belmira, Entrerios, San Pedro de los Milagros, Don Matias, and Santa Rosa de Osos), Valle de Aburrá (Bello and Medellín), and East (La Unión, Rionegro, and El Retiro). Two municipality-level samples were selected for sequencing.

DNA isolation

One tube (4 mL) of blood was collected in heparinized syringes by coccygeal venipuncture. Samples were homogenized by inversion and transferred to 15 mL tubes. The *buffy coat* was isolated by centrifugation for 4 min at 3000 rpm and 4 °C. The salting-out technique obtained DNA from the *buffy coat* [20]. DNA was resuspended in TE buffer (1 M Tris HCl, 0.5 M EDTA, pH 8.0) and stored at 4 °C until analysis. The quality and quantity of the DNA were evaluated on a spectrophotometer (NanoDrop2000) and agarose gel (1%).

PCR amplification for the *env* gene

A nested PCR was performed for all samples. A region of the viral *env* gene was amplified to obtain a 444 bp fragment using the previously reported oligonucleotides [5]. The reaction was carried out according to the protocol previously described [40].

PCR amplification for the *tax* gene

A region of the viral *tax* gene was amplified to obtain a 926 bp fragment using the oligonucleotides of this study, *tax*-FW (5'-GGCCCCACTCTCTACATGC-3') and *tax*-RV (5'-CGGGAGAGCCATTCATTTT-3'). PCR amplification was performed in a final volume of 25 µL with 150 ng DNA, 3.0 µL (10 mM) of each oligonucleotide, 0.4 mM dNTPs, 1X buffer PCR (ThermoScientific®), 3 mM MgCl₂, and 1U Taq Polymerase DNA. The reaction was performed in a T3 thermocycler (Biometra®) with the following protocol:

initial denaturation at 94 °C for 5 min, followed by 40 cycles of 98 °C for 30 s, 56 °C for 30 s, and 72 °C for 3 min; the final extension was run at 72 °C for 5 min. The PCR product was analyzed by electrophoresis on 2% agarose gel stained with EZ-VISION (Amresco®) in a Gel Doc (Bio-Rad, California-USA).

Sequencing

The PCR products obtained from the samples were sent for sequencing in a commercial facility (Macrogen Inc. Korea). The PCR products were sent in 1.5 mL tubes at 100 ng/μL concentrations in a final volume of 30 μL. Sequencing was done in both senses. The same commercial facility purified the products.

Phylogenetic analysis of *env* nucleotide sequences

Nucleotide sequences were aligned with 52 partial sequences of the *env* gene deposited in GenBank representative of the 12 BLV genotypes. These sequences were from different geographic regions, including several South American countries. Sequences were manually aligned in the software MEGA v11. The Hasegawa-Kishino-Yano with Gamma distribution (HKY + G) was selected as the best substitution model based on the Akaike Information Criterion (AIC). Phylogenetic analyses were performed by two methods: (1) a Maximum Likelihood analysis was performed using MEGA v11 branch support determined with 1000 bootstraps, and (2) a second phylogenetic tree was built utilizing the Bayesian method with the MrBayes V3.2.7 program with the same substitution model. The number of generations was determined until the likelihood of the cold chain (LnL) was stationary. The final tree was edited in FigTree V1.4.0. Mean nucleotide distances within (intra-genotype) and between (inter-genotype) BLV genotypes were estimated and compared using the *p-distance* model, and the nucleotide substitution rate matrix was estimated using the maximum likelihood method in MEGA v11.

Phylogenetic analysis of *tax* nucleotide sequences

For the phylogenetic analysis of the *tax* region, sequences of the 19 BLV strains of this study were aligned with 55 sequences obtained from GenBank. For this analysis, complete genome sequences of BLV and the *pX* region were included. Sequences were manually aligned in the software MEGA v11. Phylogenetic analysis was performed using the same two methods used for *env*. First, a Maximum Likelihood analysis using MEGA v11 was performed using the General Time Reversible with Gamma distribution (GTR + G), selected as the best substitution model based on AIC, bootstrap values were determined with 1000

repetitions. A second phylogenetic tree was built by Bayesian methods using MrBayes V3.2.7 and the evolution model used was General Time Reversible with Gamma distribution (GTR + G) distribution with the same substitution model. The number of generations was determined until the likelihood of the cold chain (LnL) was stationary. Mean nucleotide distances within (intra-genotype) and among (inter-genotype) and the nucleotide substitution rate matrix were estimated in the same way as for the *env* gene.

Amino acid changes in BLV envelope protein (Env) and regulatory protein (Tax)

Alignments of the partial amino acid sequences of Env and Tax proteins from the twenty strains were obtained using MEGA v11 tools. The BLV cell line FLK (GenBank accession EF600696), BLV strains Arg41 (GenBank accession FJ914764), and the complete BLV genome (GenBank accession AF033818) were used as reference sequences for the analysis of amino acid changes.

Results

Phylogenetic analysis of *env* nucleotide sequences

The partial sequences of the *env* gene were analyzed from strains obtained from Holstein Friesian cows from the Department of Antioquia. A PCR product of 444 bp was obtained for the *env* gene, and the samples were identified as follows: OP345001, OP345002, OP345003, OP345004, OP345005, OP345006, OP345007, OP345008, OP345009, OP345010, OP345011, OP345012, OP345013, OP345014, OP345015, OP345016, OP345017, OP345018, OP345019, and MT452066. A Maximum Likelihood phylogenetic tree of the *env* gene (Fig. 1A) sequences of 19 Colombian samples formed a monophyletic group with <70% bootstrap support inside genotype 1. Sequences from Brazil, Uruguay, Argentina, Mexico, Costa Rica, the United States, Korea, and Japan were also in genotype 1. The complete genotype 1 clade had a 73% bootstrap support. The OP328419 Belmira sample was the only one that grouped with the sequences EF065649 (United States), KP201464 (Korea), EF065650 (Japan), and MG678790 (Mexico) in genotype 3 with 90% bootstrap support. Figure 1B shows the resulting Bayesian phylogenetic tree. As in the Maximum Likelihood tree, 19 Colombian sequences were grouped with sequences from Brazil, Uruguay, Paraguay, Mexico, Argentina, the United States, and Korea within genotype 1. On the contrary, the OP328419 Belmira sample grouped with sequences from the EF065649 (United States), KP201464 (Korea), EF065650 (Japan), and MG678790 (Mexico) in genotype 3 with high posterior probabilities (prob = 1).

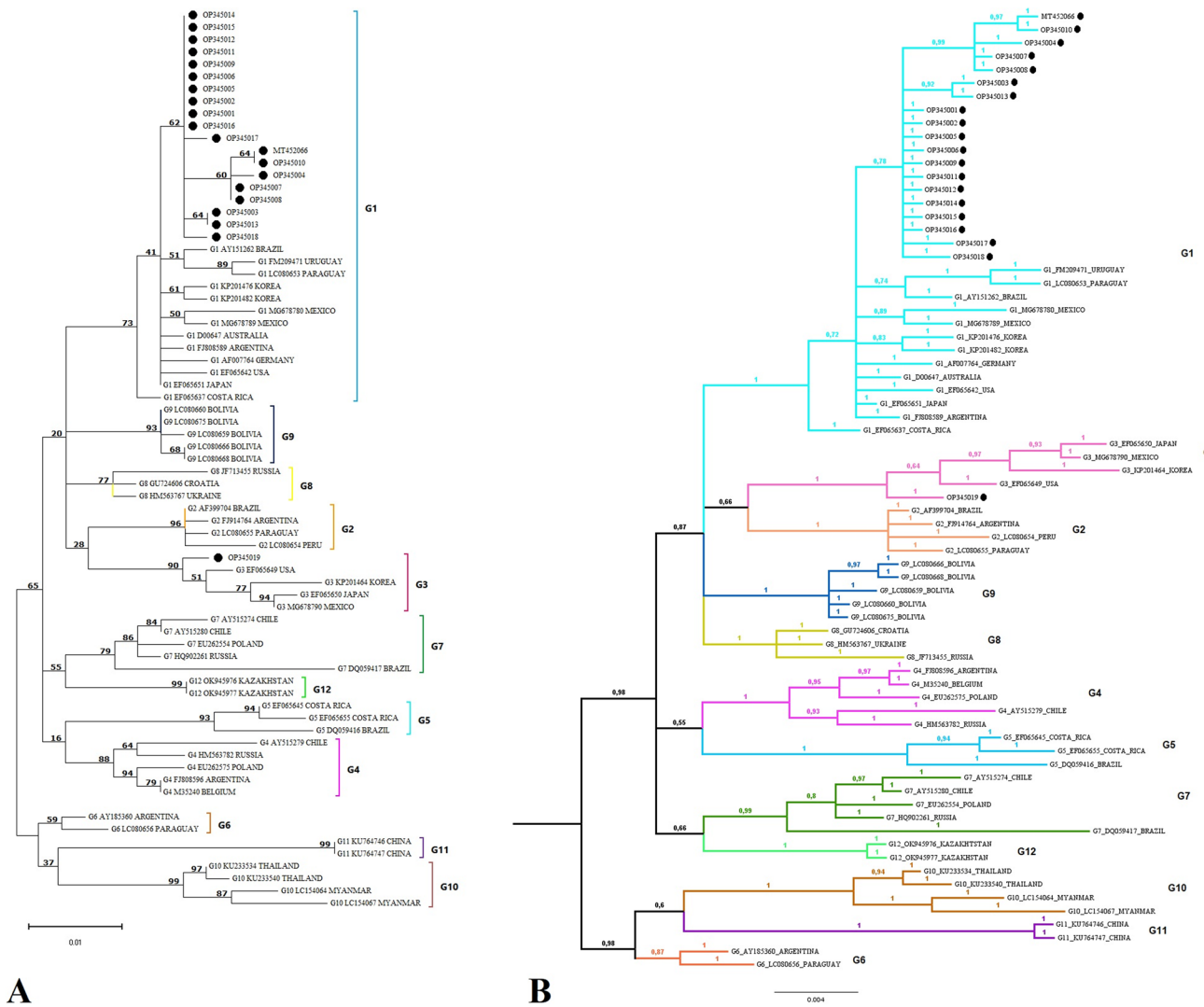


Fig. 1 Molecular phylogenetic analysis of the BLV *env* gene (444 nucleotides) through Maximum Likelihood and Bayesian inference **A**. Maximum Likelihood: The evolutionary history was inferred using the Hasegawa-Kishino-Yano substitution model. A discrete Gamma distribution was used to model evolutionary rate differences among sites. Branch support values came from 1000 bootstrap replicates. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Bar length indicates 0.01 substitutions per site. Twenty partial *env* sequences from Colombian municipalities are indicated with (●). Genotypes (G) are indicated on the right side. There were a total of 444 positions and 72 sequences in the final data-

set. Evolutionary analyses were conducted in MEGA v11. **B** Bayesian phylogenetic tree of the BLV *env* gene sequences inferred with the Hasegawa-Kishino-Yano model with Gamma distribution (HKY + G). The consensus tree was obtained after discarding (burnin) the initial 25% of the generations. The numbers at the nodes show posterior probabilities, and the bar length indicates 0.003 substitutions per site. Partial *env* gene sequences from Colombian municipalities are indicated with (●). Genotypes (G) names are shown around the figure. A total of 444 positions and 72 sequences were included in the final dataset. Evolutionary analysis was conducted in MrBayes V3.2.7

As shown in Table 1, the mean intra-genotype distances between genotypes 4, 5 and 10 are not different (0.014). Genotypes 5 and 10 (0.056) and 3 and 10 (0.051) are larger than genotypes 1 and 10, and 7 and 10 (0.047). On the other hand, genotypes 8 and 10 have the smallest mean inter-genotype distance (0.022).

The BLV sequences OP345005 (El Retiro), OP345006 (El Retiro), OP345015 (San Pedro), and OP345016 (San Pedro) are identical using the *p-distance* model, although

they are from distant municipalities geographically separated by a valley. OP345019 (Belmira) has a higher proportion of nucleotide sites different from other samples and is farther from the OP345010 (La Unión) and OP345004 (Donmatias) sequences (bold cells) (Table 2).

Substitution pattern and rates were estimated under the General Time Reversible model (+G); the nucleotide frequencies are A = 23.42%, T/U = 26.97%, C = 29.25%, and G = 20.36% for the *env* gene fragment (Table 3).

Table 1 Mean intra-genotype and inter-genotype genetic distances in the *env* gene (444 bp) of BLV genotypes. Distances were estimated with the *p-distance* substitution model

	G.1	G.2	G.3	G.4	G.5	G.6	G.7	G.8	G.9	G.10	G.11	G.12
G.1	0.008*											
G.2	0.031	0.006*										
G.3	0.033	0.030	0.012*									
G.4	0.034	0.028	0.038	0.014*								
G.5	0.036	0.037	0.045	0.037	0.014*							
G.6	0.029	0.030	0.032	0.027	0.038	0.008*						
G.7	0.036	0.032	0.040	0.034	0.040	0.030	0.017*					
G.8	0.025	0.026	0.029	0.028	0.040	0.023	0.031	0.010*				
G.9	0.027	0.028	0.030	0.028	0.041	0.026	0.032	0.022	0.003*			
G.10	0.047	0.048	0.051	0.044	0.056	0.030	0.047	0.040	0.044	0.014*		
G.11	0.054	0.050	0.057	0.048	0.057	0.035	0.050	0.048	0.047	0.051	0.000*	
G.12	0.033	0.034	0.037	0.031	0.039	0.027	0.028	0.028	0.030	0.045	0.047	0.000*

Bold values: *Mean intra-genotype nucleotide distances

Phylogenetic analysis of *tax* nucleotide sequences

Figure 2A presents the Maximum Likelihood phylogenetic tree based on 926 bp of *tax* gene sequences of BLV strains. The sample RIONEGRO_375 was not successful in sequencing. Eighteen Colombian sequences were grouped in genotype 1 with 91% bootstrap support, along with sequences from Japan (AB934282, AP018013), Mexico (MN707614), Uruguay (HE967301), and Argentina (JF288774). The OP345019 (Belmira) sequence from the *tax* gene grouped in genotype 3 with sequences from Japan (LC577639, LC577638, and LC577637); the bootstrap support of this branch was 95%. The topology of the Bayesian phylogenetic tree of the *tax* gene was similar to the Maximum Likelihood tree. Eighteen Colombian sequences were grouped in genotype 1 with high posterior probabilities (prob = 1) and with sequences from Uruguay, Paraguay, Mexico, Argentina, USA, and Japan. The sequences of strains from Bolivia and Paraguay are found within genotype 9. On the other hand, the OP345019 (Belmira) sequence was registered in the same clade as genotype 3 with sequences from Japan (the same sequences used in the Maximum Likelihood tree).

Table 4 shows the mean intra-genotype and inter-genotype nucleotide distance in the *tax* gene (926 bp). The mean intra-genotype distances between genotypes 4 and 10 are not different (0.014). This distance was similar to the analysis of the *env* gene. Genotypes 1 and 6, 3 and 6, and 3 and 10 have the biggest mean inter-genotype distance (0.056). Genotypes 1 and 2 and 2 and 9 have the smallest mean inter-genotype distance (0.022).

BLV sequences from Santa Rosa and El Retiro municipalities are identical using the *p-distance* model, although they are geographically remote municipalities. OP328418 Belmira has a higher proportion of nucleotide sites than

other samples and is farther from the San Pedro, La Unión, Entreríos, and Medellín sequences (bold cells) (Table 5).

Substitution patterns and rates were estimated under the General Time Reversible model (+G). The nucleotide frequencies are A = 18.58%, T/U = 24.34%, C = 37.44%, and G = 19.64% for the *tax* gene fragment (Table 6).

Amino acid sequences of the BLV envelope protein (Env)

Figure 3 presents the amino acid alignment of the Env protein for the 20 strains obtained from Holstein Friesian cows naturally infected with BLV. The envelope protein of 134 amino acids was obtained from the amino acid positions 101 to 234. Four different amino acid substitutions were found in the Colombian sequences concerning reference sequences and previously reported changes. E111K was found in five strains (MT452066, OP345004, OP345007, OP345008, and OP345010), D128H in three (MT452066, OP345010, and OP345017), Q135R in two (OP345003 and OP345013), and Q151L only in OP345004. E111K is in the CD4 + T cell epitope, whereas Q135R is in the neutralization domain ND2.

Amino acid sequences of the BLV regulatory protein (Tax)

Figure 4 shows the amino acid alignment of the partial sequence of the Tax protein for 19 BLV strains. The 296 amino acid residues (positions 15 to 310 in the protein) were deduced from the nucleotide sequence of the *tax* gene obtained in this study. Three new amino acid substitutions were found in the strains relative to the reference sequences

Table 2 Estimates of evolutionary divergence between twenty BLV strains from different municipalities of Antioquia, Colombia. Distances were estimated with the *p-distance* substitution model

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1 BELLO_006 OP345001																				
2 BELLO_364 MT452066	0.008																			
3 BELMIRA_108 OP345019	0.023	0.030																		
4 BELMIRA_9911 OP345002	0.000	0.008	0.023																	
5 DONMATIAS_026 OP345003	0.003	0.010	0.025	0.003																
6 DONMATIAS_053 OP345004	0.008	0.005	0.030	0.008	0.010															
7 ELRETIRO_6177 OP345005	0.000	0.008	0.023	0.000	0.003	0.008														
8 ELRETIRO_P16 OP345006	0.000	0.008	0.023	0.000	0.003	0.008	0.000													
9 ENTRERRIOS_126 OP345007	0.005	0.003	0.028	0.005	0.008	0.003	0.005	0.005												
10 ENTRERRIOS_177 OP345008	0.005	0.003	0.028	0.005	0.008	0.003	0.005	0.005	0.000											
11 LAUNION_5682 OP345009	0.000	0.008	0.023	0.000	0.003	0.008	0.000	0.000	0.005	0.005										
12 LAUNION_5719 OP345010	0.008	0.000	0.030	0.008	0.010	0.005	0.008	0.008	0.003	0.003	0.008									
13 MEDELLIN_421 OP345011	0.000	0.008	0.023	0.000	0.003	0.008	0.000	0.000	0.005	0.005	0.000	0.008								
14 MEDELLIN_P14 OP345012	0.000	0.008	0.023	0.000	0.003	0.008	0.000	0.000	0.005	0.005	0.000	0.008	0.000							
15 RIONEGRO_375 OP345013	0.003	0.010	0.025	0.003	0.000	0.010	0.003	0.003	0.008	0.008	0.003	0.010	0.003	0.003						
16 RIONEGRO_P17 OP345014	0.000	0.008	0.023	0.000	0.003	0.008	0.000	0.000	0.005	0.005	0.000	0.008	0.000	0.000	0.003					
17 SANPEDRO_118 OP345016	0.000	0.008	0.023	0.000	0.003	0.008	0.000	0.000	0.005	0.005	0.000	0.008	0.000	0.000	0.003	0.000				
18 SANPEDRO_137 OP345015	0.000	0.008	0.023	0.000	0.003	0.008	0.000	0.000	0.005	0.005	0.000	0.008	0.000	0.000	0.003	0.000	0.000			
19 SANTAROSA_1750 OP345017	0.003	0.005	0.025	0.003	0.005	0.010	0.003	0.003	0.008	0.008	0.003	0.005	0.003	0.003	0.005	0.003	0.003	0.003	0.003	0.003

Table 2 (continued)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
20 SANTAROSA_1758 OP345018	0.003	0.010	0.025	0.003	0.005	0.010	0.003	0.003	0.008	0.008	0.003	0.010	0.003	0.003	0.005	0.003	0.003	0.003	0.003	0.005

Bold values indicate the evolutionary divergence between OP345002 and other BLV strains

Table 3 Maximum Likelihood estimates of the substitution matrix (*env* gene)

	A	T/U	C	G
A	–	<i>1.52</i>	<i>3.46</i>	18.38
T/U	<i>1.32</i>	–	23.72	<i>1.09</i>
C	<i>2.77</i>	21.88	–	<i>1.34</i>
G	21.15	<i>1.45</i>	<i>1.92</i>	–

Rates of different transitional substitutions are shown in bold, and those of transversional substitutions are observed in italics

and previously reported amino acid changes. R43K and A185T were found in OP328418 (Belmira strain), previously reported by Ohnuki et al. [30]. However, a new change in L105F was found in this same strain. The V146A amino acid change (without a previous report) was registered in three strains (OP328401, OP328414, and OP328415). The A185T and L168V changes (without a previous report) are recorded in the leucine-rich activation domain.

Discussion

The BLV genotypes were initially established from the differences found in the restriction site analysis, and the BLV strains were classified into three genotypes according to the restriction patterns of Beier et al. [5]. Currently, BLV is classified into twelve genotypes described through the partial sequence of the *env* gene. The topology used in the phylogenetic analyzes of the *env* gene with high values of branch support (> 70% in bootstrap and 1.0 in posterior probabilities) by the maximum likelihood and Bayesian methods agrees with the topology published in other papers for the 12 current genotypes [18, 28, 42]. Through these analyses, genotypes 1 and 3 were identified in the samples of the municipalities of Antioquia. Genotype 1 has previously been identified in South America in Brazil, Uruguay, Argentina, and Mexico (in Holstein Friesian Cows) [6, 21, 23, 30]. Studies have found a correlation between the genotype and the geographical origin of the sample [24]. Genotype 1 is distributed worldwide [15], but some genotypes have been found in specific regions. Genotypes 6 and 9 are found in South America only [29]; Genotype 8 solely in Europe [31]; Genotype 10 only in Southeast Asia [18]; and recently, a new genotype was discovered in Kazakhstan [34]. The origin of BLV is located in Asia, and its distribution throughout the world is due to its ability to infect cattle (related to primary production and food security). The introduction of cows in South America and later to the United States (1950) and the crosses with local populations positively influenced the rapid spread of the virus [24]. In addition, geographic isolation

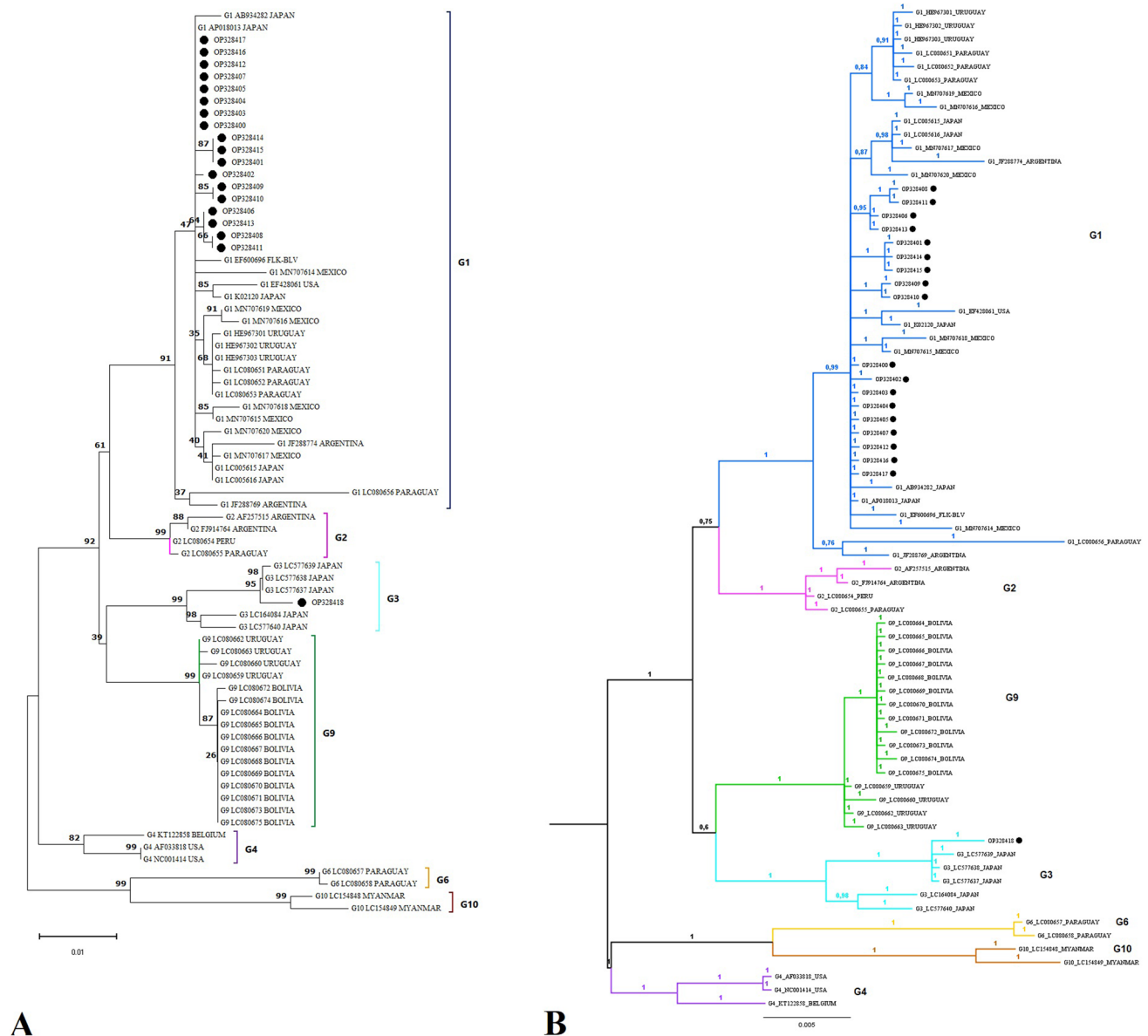


Fig. 2 Phylogenetic tree of the BLV *tax* gene by Maximum Likelihood and Bayesian inference. **A** Maximum Likelihood method: The evolutionary history was inferred using the General Time Reversible substitution model with a discrete Gamma distribution of rate differences among sites (GTR+G). Bootstrap values (1000 replicates) are shown close to the nodes. The tree is drawn to scale, with branch lengths measured as the number of substitutions per site (Bar: 0.01 substitutions per site). The analysis involved 55 partial *tax* gene sequences from the GenBank database. Nineteen *tax* gene sequences from Colombian municipalities are indicated with (●). Genotypes (G) are indicated by the numbers to the right of the clades. There

was a total of 926 positions in the final dataset. Evolutionary analyses were conducted in MEGA v11. **B** Bayesian phylogenetic tree using the General Time Reversible model with Gamma distribution (GTR+G). The numbers at the nodes are posterior probabilities, and the bar length indicates 0.005 substitutions per site. The analysis involved 74 partial *tax* gene sequences. Partial *tax* gene sequences from Colombian municipalities are indicated with (●). Numbers around the figure indicate genotypes. There was a total of 926 positions in the final dataset. Evolutionary analyses were conducted in the MrBayes V3.2.7 program

has also contributed to the diversification of the virus [18, 29, 34].

The current research identified one sample of OP328418 (by the *env* gene) and OP345019 (by the *tax* gene) from Belmira as genotype 3. This genotype has previously only

been reported in Korea [17], the United States, and Japan [42] and was recently identified in Mexico [7]. The Belmira strain is the first genotype 3 reported in South America, and its entry into the country may also be related to the import of live animals, semen, and embryos, especially from the

Table 4 Mean intra-genotype and inter-genotype genetic distances in the *tax* gene (926 bp). Distances were estimated with the *p-distance* substitution model

	G.1	G.2	G.3	G.4	G.6	G.9	G.10
G.1	0.006*						
G.2	0.023	0.004*					
G.3	0.032	0.030	0.011*				
G.4	0.034	0.029	0.037	0.010*			
G.6	0.056	0.052	0.056	0.044	0.001*		
G.9	0.028	0.023	0.031	0.029	0.055	0.002*	
G.10	0.052	0.049	0.056	0.043	0.044	0.051	0.010*

Bold values: *Mean intra-genotype nucleotide distances

United States. In 2017, 108 breeding cattle, 118 embryos, and 331,057 semen straws were imported to Colombia. The United States has been one of the primary sources of genetic material (semen or embryo samples) for genetic breeding programs of cattle herds in Colombia. The BLV genome has been found in fresh and frozen semen samples [3, 11]. Semen, like other fluids such as blood, are sources for the horizontal transmission of BLV [11], and its infectivity potential has been demonstrated in the sheep model [16].

The transcript of the *env* gene is a 5.1 kb RNA; the resulting products are the surface protein (SU) or gp51 and the transmembrane protein (TM) or gp30. Both proteins are glycosylated polypeptides and are associated through disulfide bridges forming a stable complex [14]. The gp51 protein is responsible for binding the virus to the host cell. This interaction allows the refolding of the gp30 protein, which activates its potential to fuse with the plasma membrane of the cell through a fusion peptide. The gp51 protein induces the production of antibodies in infected animals and is essential for the infectivity of the virus [14]. On the other hand, the gp51 protein has three conformational epitopes (F, G, and H) in the N-terminal region, which play an important role in viral infectivity and syncytia formation [4]. The Env protein has four important regions: CD4 + epitope, zinc-binding peptide, CD8 + T cell epitope, and E epitope. This study found four new amino acid substitutions: E111K substitution was the most frequent among the Colombian strains, followed by D128H, Q135R, and Q151L. The change of amino acid E111K was registered in the CD4 + T cell epitope, whereas the change of amino acid Q135R was recorded in the neutralization domain ND2. CD4 + T cell epitope has been related to the protection against infections and cancer [39]. The substitution of glutamine (polar amino acid uncharged) by arginine (an amino acid with a positive charge) can change the structure or function of this domain. Substitutions between residues 137 and 156 affect the fusion of the gp51 protein and in vivo infectivity [13]. It also decreases the immunoreactivity of the ND2 epitope [41]. Only one strain showed a change in this region: the change of glutamine (polar amino acid uncharged) by leucine (hydrophobic amino acid) in position 151. However,

the effect of this substitution on the protein structure is not known. CD8 + T cell epitope, E epitope, and ND3 epitope were conserved in all the strains from Colombia.

BLV clusters have generally been elaborated through phylogenetic analysis of a 444 bp fragment of the envelope gene (*env*); however, other studies have also used the viral *tax* gene [8, 9]. The phylogenetic analyses carried out in the current study through the Bayesian and Maximum Likelihood of the *env* and *tax* genes of Colombian strains are consistent in the genotype 1 and genotype 3 clustering. The mean inter-genotypic nucleotide distances in the *env* and *tax* genes are similar. This suggests that *tax* is a potential gene for phylogenetic studies. The *tax* gene is a transcriptional transactivator that increases transcription factors binding to DNA and improves the synthesis of viral proteins. It is located in the *pX* region between the *env* and the 3'LTR regions [2]. The *tax* protein targets T cells through epitopes 110–130/131–150 and B cells through epitopes 261–280. It modulates the expression and function of many proteins involved in cell growth, apoptosis, tumor suppression, and DNA repair enzymes [2, 14]. These characteristics make *tax* an essential gene in the pathogenesis of BLV. Previously, different functional areas have been described in the *tax* protein as a putative zinc finger motif between amino acids 30 to 53, the transactivation domain between amino acids 157 and 197, and two phosphorylation sites in residues 106 and 293 [41]; these sites are critical for modulating transcriptional activation. Amino acid substitutions in the *tax* protein can trigger structural and functional damage. It has been reported that C30G and A53H changes in the putative zinc finger motif region were critical and generated the loss of function as a viral transactivator [41].

The sequences of the strains in the current study were conserved between amino acids 30 to 53, except for the OP328418 Belmira strain that showed the R43K change, similar to what Ohnuki et al. [25] found. Although the R43K and A185T changes of genotype 3 are found in the putative zinc finger and leucine-rich activation domains, respectively, no data demonstrate to date the functional significance of these changes [36]. Both residues (K and T) are hydrophilic and belong to the charged polar amino

Table 5 Estimates of evolutionary divergence between nineteen BLV strains from different municipalities of Antioquia, Colombia. Distances were estimated with the *p-distance* substitution model

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
1 BELLO_006 OP328400																				
2 BELLO_364 OP328401	0.002																			
3 BELMIRA_108 OP328418	0.028	0.030																		
4 BELMIRA_9911 OP328402	0.001	0.003	0.029																	
5 DONMATIAS_026 OP328403	0.000	0.002	0.028	0.001																
6 DONMATIAS_053 OP328404	0.000	0.002	0.028	0.001	0.000															
7 ELRETIRO_6177 OP328405	0.000	0.002	0.028	0.001	0.000	0.000														
8 ELRETIRO_P16 OP328406	0.001	0.003	0.029	0.002	0.001	0.001	0.001													
9 ENTRERRIOS_126 OP328407	0.000	0.002	0.028	0.001	0.000	0.000	0.000	0.001												
10 ENTRERRIOS_177 OP328408	0.002	0.005	0.030	0.003	0.002	0.002	0.002	0.001	0.002											
11 LAUNION_5682 OP328409	0.002	0.005	0.030	0.003	0.002	0.002	0.002	0.003	0.002	0.005										
12 LAUNION_5719 OP328410	0.002	0.005	0.030	0.003	0.002	0.002	0.002	0.003	0.002	0.005	0.000									
13 MEDELLIN_421 OP328411	0.002	0.005	0.030	0.003	0.002	0.002	0.002	0.001	0.002	0.000	0.005	0.005								
14 MEDELLIN_P14 OP328412	0.000	0.002	0.028	0.001	0.000	0.000	0.000	0.001	0.000	0.002	0.002	0.002	0.002							
15 RIONEGROP17 OP328413	0.001	0.003	0.029	0.002	0.001	0.001	0.001	0.000	0.001	0.001	0.003	0.001	0.001	0.000						
16 SANPEDRO_118 OP328414	0.002	0.000	0.030	0.003	0.002	0.002	0.002	0.003	0.002	0.005	0.005	0.005	0.005	0.002	0.003	0.003				
17 SANPEDRO_137 OP328415	0.002	0.000	0.030	0.003	0.002	0.002	0.002	0.003	0.002	0.005	0.005	0.005	0.005	0.002	0.003	0.003	0.000			
18 SANTAROSA_1750 OP328416	0.000	0.002	0.028	0.001	0.000	0.000	0.000	0.001	0.000	0.002	0.002	0.002	0.002	0.000	0.001	0.001	0.002	0.002		
19 SANTAROSA_1758 OP328417	0.000	0.002	0.028	0.001	0.000	0.000	0.000	0.001	0.000	0.002	0.002	0.002	0.002	0.000	0.001	0.001	0.002	0.002	0.002	0.000

Bold values indicate the evolutionary divergence between OP345002 and other BLV strains

Table 6 Maximum Likelihood estimate of the substitution matrix (*tax* gene)

	A	T/U	C	G
A	–	<i>2.10</i>	<i>6.17</i>	23.60
T/U	<i>1.60</i>	–	22.20	<i>0.22</i>
C	<i>3.06</i>	14.43	–	<i>1.38</i>
G	22.32	<i>0.28</i>	<i>2.63</i>	–

Rates of different transitional substitutions are shown in bold, and those of transversal substitutions are shown in italics

acids group; glutamic acid is negatively charged, and lysine is positively charged. This may influence protein structure and its capacity as a viral transactivator. Two substitutions were found in the transactivation domain (L177V and A185T) of the OP328418 strain. The substitution of L177V may not represent damage to the structure and function of the tax protein because both residues are an apolar aliphatic chain and have a similar structure. However, substituting the alanine residue (apolar aliphatic chain) with threonine (polar without charge) can influence this domain. The region between residues 240 and 265 of the tax protein is directly related to the transcriptional activity of the virus. Its functions include regulating BLV-LTR activation, expression, in vitro but not in vivo propagation of BLV, in vitro but not

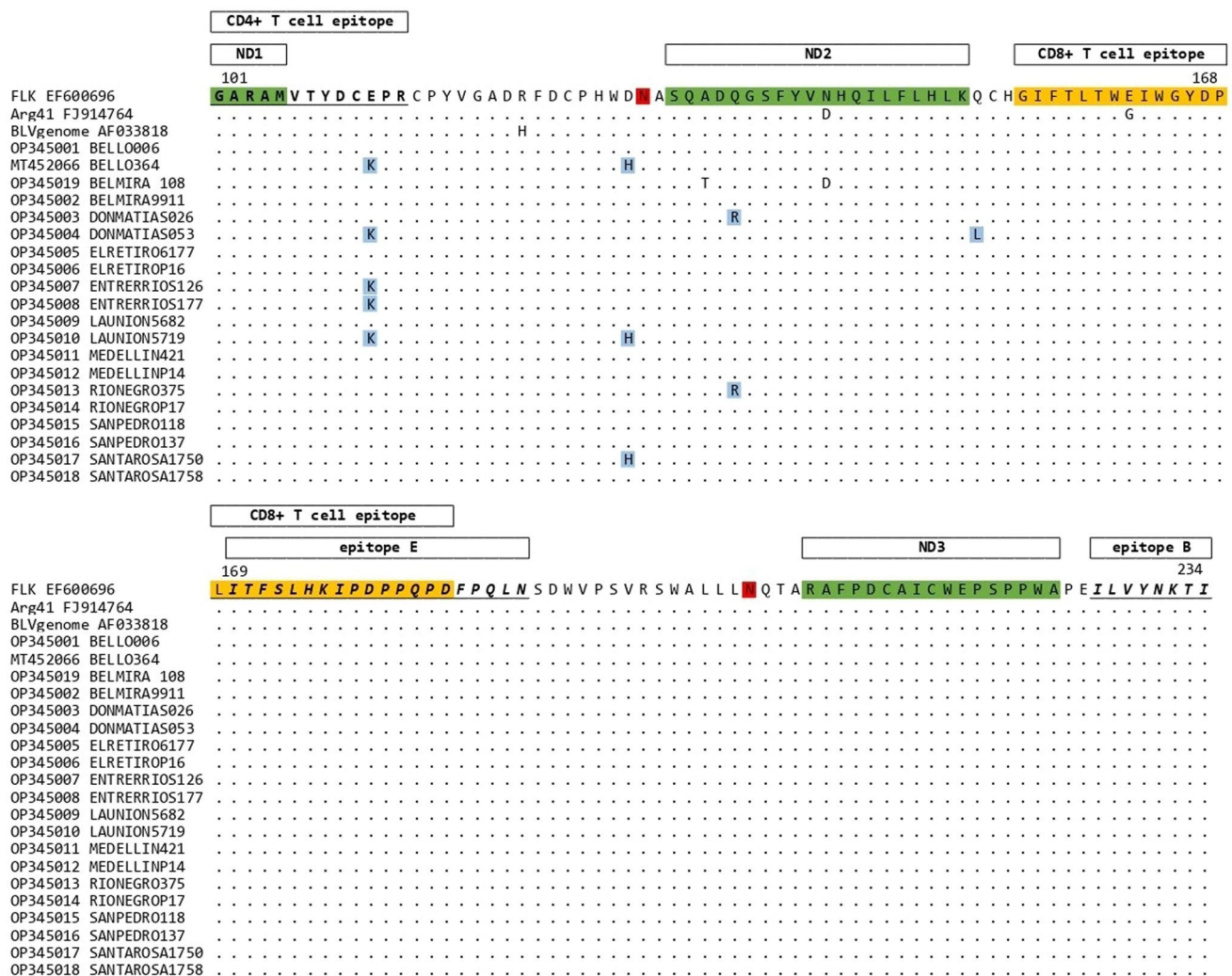


Fig. 3 Alignment of the partial amino acid sequences of the envelope protein (Env) of twenty strains from Colombia. The homologous residues of the FLK (EF600696), Arg41 (FJ914764), and the complete BLV genome (AF033818) reference strains were also included for comparison purposes. Amino acids involved in neutralization domains are shown in green. The CD4+T epitopes are in bold

and underlined. The CD8+T epitopes are shown in orange. Linear epitopes, B and E, are shown in italics and underlined. N-glycosylation sites are in red. Newly found substitutions are marked with blue. Epitopes and glycosylation sites were obtained from previous reports [22, 29, 31]

Zn finger

15 94

FLK EF600696 CPALVLSNDVTDAW**CPLCGPHERLQFERIDTTLTCETH**RINWTDGRPCGLNGTLFPRLVHSETRPQGPRLWNCPLP

Arg41 FJ914764 P T F A

BLVgenome AF033818 T F

OP328400 BELLO006 .

OP328401 BELLO364 .

OP328418 BELMIRA108 K T F

OP328402 BELMIRA9911 .

OP328403 DONMATIAS026 .

OP328404 DONMATIAS053 .

OP328405 ELRETIRO6177 .

OP328406 ELRETIRO6177 .

OP328407 ENTRERRIOS126 .

OP328408 ENTRERRIOS177 .

OP328409 LAUNIONS5682 .

OP328410 LAUNIONS5719 .

OP328411 MEDELLIN421 .

OP328412 MEDELLINP14 .

OP328413 RIONEGROP17 .

OP328414 SANPEDRO 118 .

OP328415 SANPEDRO 137 .

OP328416 SANTAROSA1750 .

OP328417 SANTAROSA1758 .

Leucine-rich activation domain

95 174

FLK EF600696 AVRAQPGPVSL**LP**FERSFPFQPYQCQLPSASSDGCPIIGHGLLPWNNLVTHPVLKVLILNQ**MANFSLLP**SFDTLLVDPLR

Arg41 FJ914764 V G P

BLVgenome AF033818 Q G P

OP328400 BELLO006 . G

OP328401 BELLO364 . A

OP328418 BELMIRA108 F G P

OP328402 BELMIRA9911 . G V

OP328403 DONMATIAS026 . G

OP328404 DONMATIAS053 . G

OP328405 ELRETIRO6177 . G

OP328406 ELRETIRO6177 . G

OP328407 ENTRERRIOS126 . G

OP328408 ENTRERRIOS177 . G

OP328409 LAUNIONS5682 . G

OP328410 LAUNIONS5719 . G

OP328411 MEDELLIN421 . G

OP328412 MEDELLINP14 . G

OP328413 RIONEGROP17 . G

OP328414 SANPEDRO 118 . A

OP328415 SANPEDRO 137 . A

OP328416 SANTAROSA1750 . G

OP328417 SANTAROSA1758 . G

Leucine-rich activation domain

Multiple functional domain

175 254

FLK EF600696 **LSVFAPDTRGAIRYLSTLLTLC**PATCILPLGEPFSPNVPICRFPRDSNEPPLSEFELPLIQT**GLSWSVPAIDLFLTGPP**

Arg41 FJ914764 V T S

BLVgenome AF033818 .

OP328400 BELLO006 .

OP328401 BELLO364 .

OP328418 BELMIRA108 T S P

OP328402 BELMIRA9911 .

OP328403 DONMATIAS026 .

OP328404 DONMATIAS053 .

OP328405 ELRETIRO6177 .

OP328406 ELRETIRO6177 .

OP328407 ENTRERRIOS126 .

OP328408 ENTRERRIOS177 .

OP328409 LAUNIONS5682 .

OP328410 LAUNIONS5719 .

OP328411 MEDELLIN421 .

OP328412 MEDELLINP14 .

OP328413 RIONEGROP17 .

OP328414 SANPEDRO 118 .

OP328415 SANPEDRO 137 .

OP328416 SANTAROSA1750 .

OP328417 SANTAROSA1758 .

Multiple functional domain

255 310

FLK EF600696 **SPCDRLHVWSS**PQALQRFLHDPTLTWSELVASRKIRLD**PLK**LQLENEWLSR**L**F*

Arg41 FJ914764 Y L

BLVgenome AF033818 G L

OP328400 BELLO006 .

OP328401 BELLO364 .

OP328418 BELMIRA108 . L

OP328402 BELMIRA9911 .

OP328403 DONMATIAS026 .

OP328404 DONMATIAS053 .

OP328405 ELRETIRO6177 .

OP328406 ELRETIRO6177 .

OP328407 ENTRERRIOS126 .

OP328408 ENTRERRIOS177 .

OP328409 LAUNIONS5682 .

OP328410 LAUNIONS5719 .

OP328411 MEDELLIN421 .

OP328412 MEDELLINP14 .

OP328413 RIONEGROP17 .

OP328414 SANPEDRO 118 .

OP328415 SANPEDRO 137 .

OP328416 SANTAROSA1750 .

OP328417 SANTAROSA1758 .

Fig. 4 Alignment of the partial amino acid sequences of the regulatory Tax protein of the 19 Colombian BLV strains obtained in this study. The homologous sequences of FLK (EF600696), Arg41 (FJ914764), and the complete BLV genome (AF033818) reference strains were included. The location of a putative zinc finger domain (blue), a Leucine-rich activation domain (yellow), a multiple functional domain that influences the transactivation activity of the Tax protein (green), and phosphorylation sites (red) are highlighted. Newly found substitutions are marked with pink. Domains and multiple functional domain sites were obtained from previous reports [22, 29, 31]

alive induction of apoptosis, and restrictions on the potential of other retroviral enhancers [26, 36, 38]. Through two phylogenetic reconstruction methodologies, genotypes 1 and 3 were identified in specialized dairy herds in Antioquia, the first report of genotype 3 in Colombian dairy cows. Samples obtained from infected Holstein Friesian cows without clinically visible symptoms show the genetic variability of BLV within close geographic regions associated with nucleotide changes of the virus. Identifying new mutations in the *tax* gene (L105F, V146A, and L168V) indicates the evolutionary changes of the virus. Currently, there is no relationship between genotypes and the virulence or pathogenicity of BLV. However, these results suggest the need to conduct new analyses to establish the effect of amino acid changes on the structure and function of the envelope protein and, especially, on the tax oncogene.

Acknowledgements This research was supported by the Grupo BIOGEM (Universidad Nacional de Colombia, Medellín headquarters). Thanks to the dairy farmers who participated in the study.

Funding Open Access funding provided by Colombia Consortium. This project was funded by the Doctoral scholarship 647 of 2014 from COLCIENCIAS and Universidad Nacional de Colombia, Medellín headquarters.

Data availability All data analyzed during the current study are available from the corresponding author upon request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All animal procedures were approved by the Ethics Committee of the Universidad Nacional de Colombia, sede Medellín [CEMED-022, July 13, 2015], following the institutional guidelines for animal care.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in

the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Aida Y, Murakami H, Takahashi M, Takeshima S. Mechanisms of pathogenesis induced by bovine leukemia virus as a model for human T-cell leukemia virus. *Front Microbiol.* 2013;4:1–11. <https://doi.org/10.3389/fmicb.2013.00328>.
2. Arainga M, Takeda E, Aida Y. Identification of bovine leukemia virus tax function associated with host cell transcription, signaling, stress response and immune response pathway by microarray-based gene expression analysis. *BMC Genomics.* 2012;13:1–18. <https://doi.org/10.1186/1471-2164-13-121>.
3. Asadpour R, Jafari R. Detection of bovine leukosis provirus in blood and semen samples of bulls. *Comp Clin Pathol.* 2012;21:187–91. <https://doi.org/10.1007/s00580-010-1083-5>.
4. Bai L, Otsuki H, Sato H, Kohara J, Isogai E, Takeshima S, Aida Y. Identification and characterization of common B cell epitope in bovine leukemia virus via high-throughput peptide screening system in infected cattle. *Retrovirology.* 2015;12:1–14. <https://doi.org/10.1186/s12977-015-0233-x>.
5. Beier D, Blankenstein P, Marquard O, Kuzmak J. Identification of different BLV provirus strains by PCR-RFLP and DNA sequencing. *Berl Munch Tierarztl Wochenschr.* 2001;114:252–6.
6. Camargos M, Stancek D, Rocha M, Lessa L, Reis J, Leite R. Partial sequencing of *env* gene of bovine leukaemia virus from Brazilian samples and phylogenetic analysis. *J Vet Med Ser B.* 2002;497:325–31. <https://doi.org/10.1046/j.1439-0450.2002.00582.x>.
7. Cerón F, Gonzáles AS, Tórtora JL, Loza-Rubio E, Ramírez A. Lack of association between amino acid sequences of the bovine leukemia virus envelope and varying stages of infection in dairy cattle. *Virus Res.* 2020;278:197866. <https://doi.org/10.1016/j.virusres.2020.197866>.
8. Corredor-Figueroa AP, Salas S, Olaya-Galán NN, Quintero JS, Fajardo Á, Soñora M, Gutiérrez MF. Prevalence and molecular epidemiology of bovine leukemia virus in Colombian cattle. *Infect Genet Evol.* 2020;80:104171. <https://doi.org/10.1016/j.meegid.2020.104171>.
9. Dao TD, Bui VN, Omatsu T, Katayama Y, Mizutani T, Ogawa H, Imai K. Application of the SureSelect target enrichment system for next-generation sequencing to obtain the complete genome sequence of bovine leukemia virus. *Adv Virol.* 2018;163:3155–9. <https://doi.org/10.1007/s00705-018-3957-9>.
10. Dao TD, Nguyen HT, Bui VN, Ogawa H, Imai K. Bovine leukemia virus genotype 1 and 6 are circulating among dairy and beef cattle of small and medium holding farms in northern Vietnam. *Jpn J Vet Res.* 2019;67:83–92. <https://doi.org/10.14943/jjvr.67.1.83>.
11. Dus Santos J, Trono K, Lager I, Wigdorovitz A. Development of a PCR to diagnose BLV genome in frozen semen samples. *Vet Microbiol.* 2007;119:10–8. <https://doi.org/10.1016/j.vetmic.2006.08.030>.
12. Felmer R, Muñoz G, Recabal M. Molecular analysis of a 444 bp fragment of the bovine leukaemia virus gp51 *env* gene reveals a high frequency of non-silent point mutations and suggests the presence of two subgroups of BLV in Chile. *Vet Microbiol.* 2005;108:39–47. <https://doi.org/10.1016/j.vetmic.2005.04.005>.

13. Gatot G, Callebaut I, Van Lint C, Demonté D, Kerkhofs P, Portetelle D, Kettmann R. Bovine leukemia virus SU protein interacts with zinc, and mutations within two interacting regions differently affect viral fusion and infectivity in vivo. *J Virol*. 2002;76:7956–67. <https://doi.org/10.1128/JVI.76.16.7956-7967.2002>.
14. Gillet N, Florins A, Boxus M, Burteau C, Nigro A, Vandermeers F, Willems L. Mechanisms of leukemogenesis induced by bovine leukemia virus: prospects for novel anti-retroviral therapies in human. *Retrovirology*. 2007;418:1–32. <https://doi.org/10.1186/1742-4690-4-18>.
15. Inoue E, Matsumura K, Maekawa K, Nagatsuka K, Nobuta M, Hirata M, Okazaki K. Genetic heterogeneity among bovine leukemia viruses in Japan and their relationship to leukemogenicity. *Adv Virol*. 2011;156:1137–41. <https://doi.org/10.1007/s00705-011-0955-6>.
16. Konnai S, Usui T, Ikeda M, Kohara J, Hirata T, Okada K, Onuma M. Tumor necrosis factor-alpha genetic polymorphism may contribute to progression of bovine leukemia virus-infection. *Microbes Infect*. 2006;8:2163–71. <https://doi.org/10.1016/j.micinf.2006.04.017>.
17. Lee E, Kim EJ, Joung HK, Kim BH, Song JY, Cho IS, Lee KK, Shin YK. Sequencing and phylogenetic analysis of the gp51 gene from Korean bovine leukemia virus isolates. *Virol J*. 2015;12:1–13. <https://doi.org/10.1186/s12985-015-0286-4>.
18. Lee E, Kim E, Rathanophart J, Vitoonpong R, Kim B, Cho I, Shin Y. Molecular epidemiological and serological studies of bovine leukemia virus BLV infection in Thailand cattle. *Infect Genet Evol*. 2016;41:245–54. <https://doi.org/10.1016/j.meegid.2016>.
19. Malatestinic A. Bilateral exophthalmos in a Holstein cow with lymphosarcoma. *Can Vet J*. 2003;44:664–6.
20. Miller S, Dykes D, Polesky H. A simple salting out procedure for extracting ADN from human nucleated cells. *Nucleic Acids Res*. 1988;16:1215.
21. Montero Machuca N, Tórtora Pérez JL, González Méndez AS, García-Camacho AL, Marín Flamand E, Ramírez Álvarez H. Genetic analysis of the pX region of bovine leukemia virus genotype 1 in Holstein Friesian cattle with different stages of infection. *Adv Virol*. 2022;167:45–56. <https://doi.org/10.1007/s00705-021-05252-2>.
22. Moratorio G, Fischer S, Bianchi S, Tomé L, Rama G, Obal G, Cristina J. A detailed molecular analysis of complete bovine leukemia virus genomes isolated from B-cell lymphosarcomas. *Vet Res*. 2013;44:1–11. <https://doi.org/10.1186/1297-9716-44-19>.
23. Moratorio G, Obal G, Dubra A, Correa A, Bianchi S, Buschiazzo A, Pritsch O. Phylogenetic analysis of bovine leukemia virus strain in South America reveals diversification in seven distinct genotypes. *Adv Virol*. 2010;155:481–9. <https://doi.org/10.1007/s00705-010-0606-304.010>.
24. Nishikaku K, Yonezawa T, Nishibori M, Harada M, Kawaguchi F, Sasazaki S, Torii Y, Imakawa K, Kawai K, Liu J, Mannen H, Kobayashi T. Phylogenomics and spatiotemporal dynamics of bovine leukemia virus focusing on Asian native cattle: insights into the early origin and global dissemination. *Front Microbiol*. 2022;13:917324. <https://doi.org/10.3389/fmicb.2022.917324>.
25. Ohnuki N, Kobayashi T, Matsuo M, Nishikaku K, Kusama K, Torii Y, Satou Y. A target enrichment high throughput sequencing system for characterization of BLV whole genome sequence, integration sites, clonality and host SNP. *Sci Rep*. 2021;11:4521. <https://doi.org/10.1038/s41598-021-83909-3>.
26. Panei C, Serena M, Metz G, Bravi M, González E, Echeverría M. Analysis of the pX region of bovine leukemia virus in different clinical stages of Enzootic Bovine Leukemia in Argentine Holstein cattle. *Virus Res*. 2013;171:97–102. <https://doi.org/10.1016/j.virusres.2012.08.001>.
27. Polat M, Moe HH, Shimogiri T, Moe KK, Takeshima SN, Aida Y. The molecular epidemiological study of bovine leukemia virus infection in Myanmar cattle. *Adv Virol*. 2017;162:425–37. <https://doi.org/10.1007/s00705-016-3118-y>.
28. Polat M, Takeshima SN, Aida Y. Epidemiology and genetic diversity of bovine leukemia virus. *Virology Journal*. 2017;14:209. <https://doi.org/10.1186/s12985-017-0876-4>.
29. Polat M, Takeshima S, Hosomichi K, Kim J, Miyasaka T, Yamada K, Aida Y. A new genotype of bovine leukemia virus in South America identified by NGS based whole genome sequencing and molecular evolutionary genetic analysis. *Retrovirology*. 2016;13:1–23. <https://doi.org/10.1186/s12977-016-0239-z>.
30. Rodríguez S, Golemba M, Campos R, Trono K, Jones L. Bovine leukemia virus can be classified into seven genotypes: evidence for the existence of two novel clades. *J Gen Virol*. 2009;90:2788–97. <https://doi.org/10.1099/vir.0.011791-0>.
31. Rola-Łuszczak M, Pluta A, Olech M, Donnik I, Petropavlovskiy M, Gerilovych A, Kuźmak J. The molecular characterization of bovine leukaemia virus strains from eastern europe and siberia and its impact on phylogeny. *PLoS ONE*. 2013;8:e58705. <https://doi.org/10.1371/journal.pone.0058705>.
32. Ruiz V, Porta NG, Lomónaco M, Trono K, Alvarez I. Bovine leukemia virus infection in neonatal calves. Risk factors and control measures. *Front Vet Sci*. 2018;5:267. <https://doi.org/10.3389/fvets.2018.00267>.
33. Safari R, Hamaidia M, de Brogniez A, Gillet N, Willems L. Cis-drivers and trans-drivers of bovine leukemia virus oncogenesis. *Curr Opin Virol*. 2017;26:15–9. <https://doi.org/10.1016/j.coviro.2017.06.012>.
34. Sultanov A, Rola-Łuszczak M, Mamanova S, Ryło A, Osiniski Z, Saduakassova MA, Kuźmak J. Molecular characterization of bovine leukemia virus with the evidence of a new genotype circulating in cattle from Kazakhstan. *Pathogens*. 2022;11:180. <https://doi.org/10.3390/pathogens11020180>.
35. Szylnal M, Cleuter Y, Beskorwayne T, Bagnis C, Van Lint C, Kerkhofs P, Van den Broeke A. Disruption of B-cell homeostatic control mediated by the BLV-Tax oncoprotein: association with the upregulation of Bcl-2 and signaling through NF-kappaB. *Oncogene*. 2003;22:4531–42. <https://doi.org/10.1038/sj.onc.1206546>.
36. Tajima S, Aida Y. The region between amino acids 245 and 265 of the bovine leukemia virus BLV tax protein restricts transactivation not only via the BLV enhancer but also via other retrovirus enhancers. *J Virol*. 2000;74:10939–49. <https://doi.org/10.1128/JVI.74.23.10939-10949.2000>.
37. Tajima S, Takahashi M, Takeshima S, Konnai S, Ai Yin S, Watarai S, Aida Y. A mutant form of the tax protein of bovine leukemia virus BLV, with enhanced transactivation activity, increases expression and propagation of BLV in vitro but not in vivo. *J Virol*. 2003;77:1894–903. <https://doi.org/10.1128/JVI.77.3.1894-1903.2003>.
38. Takahashi M, Tajima S, Okada K, Davis W, Aida Y. Involvement of bovine leukemia virus induction and inhibition of apoptosis. *Microbes Infect*. 2005;7:19–28. <https://doi.org/10.1016/j.micinf.2004.09.014>.
39. Tran E, Turcotte S, Gros A, Robbins P, Lu Y. Cancer immunotherapy based on mutation specific CD4+ T cells in a patient with epithelial cancer. *Science*. 2014;344:641–5. <https://doi.org/10.1126/science.1251102>.
40. Úsuga-Monroy C, Echeverri J, López-Herrera H. Diagnóstico molecular del virus de leucosis bovina en una población de vacas Holstein, Colombia. *Archivos de Zootecnia*. 2015;64:383–8. <https://doi.org/10.21071/az.v64i248.424>.
41. Willems L, Kettmann R, Burny A. The amino acid 157–197 peptide segment of bovine leukemia virus p34tax encompass a leucine-rich globally neutral activation domain. *Oncogene*. 1991;6:159–63.
42. Zhao X, Buehring G. Natural genetic variations in bovine leukemia virus envelope gene: possible effects of selection and escape. *Virology*. 2007;366:150–65. <https://doi.org/10.1016/j.virol.2007.03.058>.

43. Zyrianova IM, Kovalchuk SN. Bovine leukemia virus tax gene/Tax protein polymorphism and its relation to Enzootic Bovine Leukosis. *Virulence*. 2020;111:80–7. <https://doi.org/10.1080/21505594.2019.1708051>.

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