

Molecular characterization of bovine leukemia virus from Moldovan dairy cattle

Aneta Pluta¹ · Marzena Rola-Łuszczak¹ · Piotr Kubiś¹ · Svetlana Balov² · Roman Moskalik³ · Bhudipa Choudhury⁴ · Jacek Kuźmak¹

Received: 10 October 2016 / Accepted: 5 January 2017 / Published online: 17 February 2017
© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract Bovine leukemia virus (BLV) is the causative agent of enzootic bovine leukosis (EBL), a disease that has worldwide distribution. Whilst it has been eradicated in most of Western Europe and Scandinavia, it remains a problem in other regions, particularly Eastern Europe and South America. For this study, in 2013, 24 cattle from three farms in three regions of Moldova were screened by ELISA and nested PCR. Of these cattle, 14 which were PCR positive, and these were molecularly characterized based on the nucleotide sequence of the *env* gene and the deduced amino acid sequence of the encoded gp51 protein. Our results demonstrated a low level of genetic variability (0–2.9%) among BLV field strains from Moldova, in contrast to that observed for other retroviruses, including human immunodeficiency virus (HIV) (20–38%) Mason IL (Trudy vologod moloch Inst 146–164, 1970) and equine infectious anemia virus (EIAV) (~40%) Willems L et al (AIDS Res Hum Retroviruses 16(16):1787–1795, 2000), where the envelope gene exhibits high levels of variation Polat M et al (Retrovirology 13(1):4, 2016). Sequence comparisons and phylogenetic analysis revealed that BLV genotype 7 (G7) is predominant in Moldova and that the BLV population in

Moldovan cattle is a mixture of at least three new sub-genotypes: G7D, G7E and G4C. Neutrality tests revealed that negative selection was the major force operating upon the 51-kDa BLV envelope surface glycoprotein subunit gp51, although one positively selected site within conformational epitope G was detected in the N-terminal part of gp51. Furthermore, two functional domains, linear epitope B and the zinc-binding domain, were found to have an elevated ratio of nonsynonymous to synonymous codon differences. Together, these data suggest that the evolutionary constraints on epitopes G and B and the zinc-binding domains of gp51 differ from those on the other domains, with a tendency towards formation of homogenous genetic groups, which is a common concept of global BLV diversification during virus transmission that may be associated with genetic drift.

Introduction

Bovine leukemia virus (BLV), a lymphotropic retrovirus belonging to the genus *Deltaretrovirus* of the family *Retroviridae*, is the etiologic agent of enzootic bovine leukosis (EBL). Infection with BLV may remain clinically silent in an aleukemic form, but about 30% of infected cattle develop persistent lymphocytosis, characterized by polyclonal expansion of B cells, with only a few percent developing lymphoid tumours [1]. BLV infection has a worldwide distribution, and epidemiological studies based on serology show high prevalence in North and South America, some Asiatic and Middle Eastern countries, and Eastern Europe [2–5]. Successful eradication programs have eliminated the disease from most Western European countries, giving these countries a significant trading advantage.

The BLV envelope glycoprotein, encoded by the envelope (*env*) gene, is crucial to the virus life cycle and is

✉ Aneta Pluta
aneta.pluta@piwet.pulawy.pl

¹ OIE Reference Laboratory for EBL, Department of Biochemistry, National Veterinary Research Institute, Pulawy, Poland

² Republican Center for Veterinary Diagnostic, Chisinau, Moldova

³ Scientific Practical Institute for Biotechnologies and Zootechny and Veterinary Medicine, Chisinau, Moldova

⁴ OIE Reference Laboratory for EBL, Department of Virology, Animal and Plant Health Agency, Weybridge, UK

the primary target of neutralizing antibodies. Studies of the genetic diversity of the *env* gene of BLV isolates from different geographical locations initially revealed eight different genotypes [2, 6–9]. Recently, Lee *et al.* [10] and Polat *et al.* [11], using whole-genome next-generation sequencing (NGS) analysis, performed a comparative analysis of the *env* sequences and identified two new genotypes, bringing the total to ten (e.g., G1-G10). These authors reported that the homology amongst BLV isolates was 94.5–97.7%.

The envelope glycoprotein consists of a 51-kDa extracellular surface subunit (SU, gp51) and a 30-kDa transmembrane subunit (TM, gp30). Analysis of the nucleotide sequence and the deduced amino acid sequence of the gp51-encoding region identified some distinct and highly conserved regions common to all field isolates of the virus [12, 13]. Some of the mutations found in gp51 were associated with marked changes in the course of infection, and it was postulated that some of them can lead to increased virulence by allowing escape from host immune surveillance [12, 14].

Glycoprotein gp51 contains seven epitopes, named beginning from the N- to the C-terminal end, G, H, F, E-E', B-B', D-D' and A, and three of these (G, H and F) are conformational epitopes that are major determinants of virus neutralization and inhibition of syncytium formation [15]. Additionally, three neutralization domains, ND1, ND2 and ND3, were shown to induce BLV-neutralizing antibodies [16]. The N-terminal and internal regions of gp51 contain five T-cell epitopes that are involved in the induction of proliferative responses in BLV-infected cattle: a CD4⁺ T-cell epitope, a CD8⁺ T-cell epitope, gp51N5, gp51N11 and gp51N12, an immunologic target of cytotoxic T lymphocytes (CTL) [17]. The glycoprotein gp30 contains an extracellular domain encompassing the hydrophobic fusion peptide and a cytoplasmic domain with a YXXL motif that is believed to be involved in signal transduction pathways [18].

We previously showed that while strains from Poland, Russia, Ukraine and Belarus can be assigned mainly to G4 and G7, a few of them contained variations that defined a new genotype designated G8 [8]. Since G8 was a unique genotype that grouped BLV isolates from the Balkan region but no other regions of Eastern Europe, we wanted to determine whether strains from cattle in another Balkan country, Moldova, might also belong to this genotype [6, 8]. EBL was first identified in Moldova in 1965, and subsequent serological surveys confirmed widespread distribution of BLV infection in Moldovan cattle [19, 20]. Moldova launched an eradication program in 2008 that decreased the overall number of BLV-infected cattle from 48.8% to 4.9%. However, serologically positive animals are currently found in the central and western parts of the

country, where the regional herd-level prevalence in dairy farms is estimated to be 20% [21].

In this study, we report the genetic variation in BLV strains from this part of the world. We focused on analysis of an 804-base-pair (bp) fragment of the *env* gene including the entire gp51 coding sequence in fourteen BLV samples from Moldovan dairy cattle. The newly generated sequences were analyzed taking into account all currently known functional domains and epitopes of the gp51 surface glycoprotein. Phylogenetic analysis was performed using all currently known BLV genotypes.

Materials and methods

Sample collection and DNA isolation

Blood samples were obtained from 24 serologically positive cattle that were naturally infected with BLV, with diagnosis confirmed by an ELISA test (INGEZIM BLV Compac 1.2.BLV.K.3, INGENASA, Spain). The animals came from three herds that were located in three geographically distinct regions of Moldova: Riscani, Aneii Noi, and Hincesti, with 4.4%, 1.0%, and 1.8% cattle seropositivity, respectively. Peripheral blood leukocytes were isolated by centrifugation at $1500 \times g$ for 25 minutes, and erythrocytes were haemolysed by osmotic shock with water and 4.5% sodium chloride. After two washes in phosphate-buffered saline (PBS), the supernatant was discarded and the cell pellets were sent to the National Veterinary Research Institute in Pulawy for analysis. Genomic DNA was extracted from 5×10^6 cells using a DNeasy Tissue Kit (QIAGEN) following the manufacturer's instructions. The concentration of genomic DNA was measured using a nanophotometer (Implen), and the samples were stored at -20°C until PCR amplification.

Amplification of proviral DNA by nested PCR

A fragment of 804 bp of the *env* gene encoding the gp51 surface glycoprotein was amplified by seminested PCR using the following primers: AP_4762 (this study) (5'-GC TCTCCTGGCTACTGACC-3'), ZM2_5786 [22] (5'-TCT GATGGCTAAGGGCAGACACGGC-3') and ZM5_5733 [22] (5'-GCTAGGCCTAAGGTCAGGGCCGC-3'). Amplification was performed on 1 μg of genomic DNA using KAPA Taq thermal polymerase (KAPA Biosystems) and a thermal cycler (Biometra) with the following cycle parameters: 3 min at 95°C , 30 s at 95°C , 30 s at 62°C (external primers) or 30 s at 66°C (internal primers), 2 min at 72°C ; after the last (34th) cycle, the samples were incubated at 72°C for 10 min. Reactions were carried out in 50 μl , with 10 μl 10X KAPA Taq Buffer, 1 μl of 10 μM dNTP mix, 1.5 μl of

each primer (10 mM), 3 μ l of 10 mM MgCl₂, and 1.5 U of KAPA Taq DNA Polymerase. Reaction products of all strains were analyzed by 1.5 % agarose gel electrophoresis with ethidium bromide staining (1 μ g/ml), in 1X TAE buffer. To rule out the possibility of Taq polymerase errors, we carried out three independent PCR amplifications for each of the DNA templates.

DNA sequencing and sequence analysis

The PCR products were purified using a NucleoSpin Extract II Kit (Macherey-Nagel GmbH & Co) and subsequently, three independent PCR products for each sample were sequenced in both directions by the Genomed S A Company (Warsaw, Poland), using a 3730xl DNA Analyzer (Applied Biosystems) and a Big Dye Terminator v3.1 Cycle Sequencing Kit. All raw sequencing data were viewed in BioEdit v7.2.5 (Abbott, Carlsbad CA). The consensus sequences corresponding to the 804-bp fragment of the *env* gene were generated from three repeats of each sample, and they were submitted to the GenBank database under accession numbers KF801457-KF801470 (Table 1). This table also includes 35 BLV reference sequences representing G1-G10, which were used in this study. To estimate the genetic distances among the strains from Moldova (intra-genotype) and between the Moldovan strains and the reference sequences (inter-genotype), analysis based on the maximum composite likelihood model (MEGA 5.2.2) [23, 24] was conducted on the 804-bp sequences. Synonymous and non-synonymous sites were identified using DnaSP 5.0 [25]. Statistical analysis was performed using STATISTICA ver. 10 (StatSoft, Dell Software, USA). A Z-test based on the number of synonymous substitutions per synonymous site (dS) and the number of non-synonymous substitutions per non-synonymous site (dN) was used to test for selection. To assess the relationship between the BLV subtypes identified in this study and those described in previous studies, the 804-bp sequences were aligned to reference sequences using the Geneious Alignment module within Geneious Pro 5.3 Software (Biomatters Ltd) [26], and a phylogenetic tree was constructed by the Bayes method with the GTR substitution model using the tree-builder tool of the Geneious software. For measuring the reliability of the clustering, 1,000 bootstrap replicates were run for each method.

Results

Phylogenetic analysis of the region of BLV *env* encoding gp51

PCR products of the expected size of 804 bp were successfully amplified and sequenced from the proviral DNA of all fourteen strains. Using maximum composite

likelihood model analysis, the nucleotide sequences of the fourteen Moldovan strains were aligned together with 35 reference sequences representing BLV genotypes G1-G10. The percentages of nucleotide substitutions were determined by pairwise comparisons (data not shown). All Moldovan sequences were found to be closely related to each other at the nucleotide level, showing an average degree of variability of 1.1%, with a range of variability of 0% to 2.9%. No substantial divergence at the nucleotide level was noted between strains from the same herds and those representing geographically distant herds. Interestingly, the smallest number of base substitutions per site and the shortest genetic distance was observed between Moldovan sequences and those belonging to G4 and G7, with 0.3%-3.6% and 0.2%-3.0% variations, respectively. Based on these results, we inferred that the Moldovan sequences group within G4 and G7.

To test their affiliation with G4 and G7, phylogenetic trees based on the 804-bp gp51-encoding sequence were constructed using Bayesian methods (Fig. 1). The topology, with high posterior probabilities, indicated that the Moldovan sequences and all reference sequences reported as of August 2016 were clearly classified into ten distinct genotypes (G1 to G10) and confirmed that the fourteen strains from Moldova belong to G4 and G7. For further analysis, we used the tree constructed on 444-bp sub-fragment, since this has been used most often in phylogenetic analysis of BLV and is therefore better for showing the presence of sub-genotypes (Fig. 2).

Twelve of the Moldovan strains grouped together with isolates from Russia, Poland, Chile and Ukraine to form a common branch within G7 with significant posterior probability of 0.94, but they were clearly clustered into distinct groups within this genotype (Fig. 2). Notably, strains 13MD, 15MD, 19MD and 20MD and strains 14MD, 16MD, 17MD and 18MD clustered into two clearly distinguished subtypes. Since their existence was supported by high posterior probabilities of 0.99 and 1.00, we propose that the existing topology of G7 be extended to include two additional new subtypes, G7D and G7E, identified in this study and defined by these two clusters of Moldovan strains. Both subtypes co-existed in the same herd from the Hincesti region of central Moldova and showed a maximum divergence of 2.1% in their nucleotide sequences. Another four Moldovan strains, 1MD, 3MD, 6MD and 7MD, from a single herd located in the Riscani region in the northern part of Moldova, also belonged to G7 but did not form well-separated groups and were located close to isolates from Russia and Ukraine. The remaining two strains, 8MD and 10MD, which were from Riscani and Aneii Noi in the southern part of Moldova and were most closely related to isolates from France and Mongolia, were classified together in G4. However, both of these strains

Table 1 Identification and origin of the sequences used for sequence analysis. The 444-bp and 804-bp sequences are indicated by green and violet squares, respectively (colour figure online)

No.	GenBank accession no.	Virus strain	Geographic origin	Genotype	Reference
	KF801467	Sample 1MD ¹	Moldova: Region Riscani	7	this work
	KF801468	Sample 3MD	Moldova: Region Riscani	7	this work
	KF801469	Sample 6MD	Moldova: Region Riscani	7	this work
	KF801470	Sample 7MD	Moldova: Region Riscani	7	this work
	KF801459	Sample 8MD	Moldova: Region Riscani	4 D	this work
	KF801460	Sample 10MD	Moldova: Region Anenii Noi	4 D	this work
	KF801457	Sample 13MD	Moldova: Region Hincesti	7 D	this work
	KF801464	Sample 14MD	Moldova: Region Hincesti	7 E	this work
	KF801461	Sample 15MD	Moldova: Region Hincesti	7 D	this work
	KF801458	Sample 16MD	Moldova: Region Hincesti	7 E	this work
	KF801465	Sample 17MD	Moldova: Region Hincesti	7 E	this work
	KF801466	Sample 18MD	Moldova: Region Hincesti	7 E	this work
	KF801462	Sample 19MD	Moldova: Region Hincesti	7 D	this work
	KF801463	Sample 20MD	Moldova: Region Hincesti	7 D	this work
	EU266063.1	Teheran	Iran	1	Hemmatzadeh, 2007 [10]
	D00647.1	pBLV-A1	Australia	1	Coulston, 1990 [4]
	KU233564	Sk8-12	Thailand	1	Lee et al., 2016 [12]
	M35242.1	FLK-BLV	USA	1	Mamoun, 1990 [14]
	AF399704.3	384	Brazil: Minas Gerais State	2	Camargos, 2002 [3]
	AF257515.1	B19	Argentina	2	Dube, 2000 [6]
	LC080654.1	lima40	Peru	2	Polat et al., 2016 [22]
	KP201465.1	GBGS-12	South Korea	3	Lee, 2015 [11]
	LC007993.1	366-K29	Japan	3	Mekata et al., (2014) ξ
	EF065648.1	USCA-2	USA: California	3	Zhao, 2007 [31]
	JQ686091.1	MC45	Russia: Moscow region	4	Lomakina et al. (2014) γ
	M35238.1	LB59	France	4	Mamoun, 1990 [14]
	LC060794.1	BLV_Mongolia-3	Mongolia	4	Ochirkhuu, 2015 [20]
	LC060795.1	BLV_Mongolia-4	Mongolia	4	Ochirkhuu, 2015 [20]
	HQ902258	1_BY	Belarus	4	Rola-Łuszczak, 2013 [25]
	HQ902259	2_BY	Belarus	4	Rola-Łuszczak, 2013 [25]
	M35240.1	LB285	Belgium	4	Mamoun, 1990 [14]
	EU262583.1	237_PL	Poland	4A	Rola-Łuszczak, 2013 [25]
	EU262575.1	301_PL	Poland	4A	Rola-Łuszczak, 2013 [25]
	EU262576.1	8_PL	Poland	4A	Rola-Łuszczak, 2013 [25]
	EU262582.1	147_PL	Poland	4A	Rola-Łuszczak, 2013 [25]
	HM563778.1	70_PL	Poland	4B	Rola-Łuszczak, 2013 [25]
	HM563775.1	67_PL	Poland	4B	Rola-Łuszczak, 2013 [25]
	HM563774.1	68_PL	Poland	4B	Rola-Łuszczak, 2013 [25]
	HM563755.1	107_PL	Poland	4B	Rola-Łuszczak, 2013 [25]
	HM563776.1	65_PL	Poland	4B	Rola-Łuszczak, 2013 [25]

Table 1 continued

	EF065643.1	CRLV	Costa Rica	5	Zhao, 2007 [31]
	EF065645.1	CRAG-1	Costa Rica: Alajuela-Grecia	5	Zhao, 2007 [31]
	KJ668817.1	PCC4	Philippines	6	Polat, 2015 [21]
	FJ808582.1	PL-1238	Argentina	6	Rodriguez, 2009 [24]
	AY185360	151	Brazil	6	Camargos <i>et al.</i> (2004)§
	LC080658	par91	Paraguay	6	Polat <i>et al.</i> , 2016 [22]
	KU233544.1	Ns94-D4	Thailand	6	Lee <i>et al.</i> , 2016 [12]
	JN695882.1	4V	Russia: Vologda region	7	Lomakina <i>et al.</i> (2014) γ
	JN695879.1	9V	Russia: Vologda region	7	Lomakina <i>et al.</i> (2014) γ
	JQ686119.1	NK13	Russia: Novosibirsk region	7	Lomakina <i>et al.</i> (2013) ϕ
	S83530.1	I2	Italy	7	Molteni, 1996 [17]
	AY515274.1	14	Chile: IX Region	7	Felmer, 2005 [8]
	EU262555.1	151_PL	Poland	7C	Rola-Łuszczak, 2013 [25]
	HM563763	160_PL	Poland	7C	Rola-Łuszczak, 2013 [25]
	EU262554.1	146_PL	Poland	7C	Rola-Łuszczak, 2013 [25]
	HM563750.1	2_RU	Russia	7A	Rola-Łuszczak, 2013 [25]
	HM563748	3_RU	Russia	7A	Rola-Łuszczak, 2013 [25]
	HM563749	5_RU	Russia	7A	Rola-Łuszczak, 2013 [25]
	HM563760	3_3_UA	Ukraine	7B	Rola-Łuszczak, 2013 [25]
	HM563761	3_1_UA	Ukraine	7B	Rola-Łuszczak, 2013 [25]
	HM563759	3_18_UA	Ukraine	7B	Rola-Łuszczak, 2013 [25]
	HM563758	3_15_UA	Ukraine	7B	Rola-Łuszczak, 2013 [25]
	JQ675760.1	MKC3511	Russia: Moscow region	8	Lomakina <i>et al.</i> (2013) ϕ
	JQ675759.1	MKC2137	Russia: Moscow region	8	Lomakina <i>et al.</i> (2013) ϕ
	HM563764.1	4_6_UA	Ukraine	8	Rola-Łuszczak, 2013 [25]
	HM563767.1	3_43_UA	Ukraine	8	Rola-Łuszczak, 2013 [25]
	HM563768.1	2_48_UA	Ukraine	8	Rola-Łuszczak, 2013 [25]
	GU724606.1	M1/ELG_Cro/08	Croatia	8	Balic, 2012 [1]
	JN990074.1	ELG_Cro/BIO/10	Croatia	8	Balic, 2012 [1]
	LC080668.1	por28	Bolivia	9	Polat, 2016 [22]
	LC080675.1	por108	Bolivia	9	Polat, 2016 [22]
	LC080659	mon1	Bolivia	9	Polat, 2016 [22]
	KU233543	Ns80-D3	Thailand	10	Lee <i>et al.</i> , 2016 [12]
	KU233547	Pa51-A3	Thailand	10	Lee <i>et al.</i> , 2016 [12]
	KU233555	Pr67-C5	Thailand	10	Lee <i>et al.</i> , 2016 [12]

ξ Mekata *et al.*, 2014, ϕ Lomakina *et al.*, 2013, γ Lomakina *et al.*, 2014, § Camargos *et al.*, 2004, direct submission to GenBank

formed a separate branch within G4, supported by a significant of posterior probability value 0.97. Therefore, we propose to add sub-genotype G4C to the existing G4A and G4B sub-genotypes. These subtypes showed high posterior probabilities, reflecting variability within G4, with a maximum divergence of 1.6% of nucleotide sequences. Taken

together, these results demonstrated that the most prevalent genotype (85.7%) was G7, with the remaining Moldovan strains clustering in G4 (14.3%). Most of the strains formed unique and distinct subtypes correlating with their geographical origin, which we propose to be designated as new sub-genotypes G7D, G7E and G4C.

Fig. 1 Phylogenetic relationship of the gp51-encoding 804-bp fragment of the *env* gene nucleotide sequences of new BLV subtypes found in this study (bold text) and all known BLV genotypes (n = 49), as inferred by Bayesian analysis. Numbers at nodes indicate posterior probabilities of sampling the node among 11,000 trees. Genotypes and subtypes as well as new subtypes found in this study are indicated at the right by vertical lines

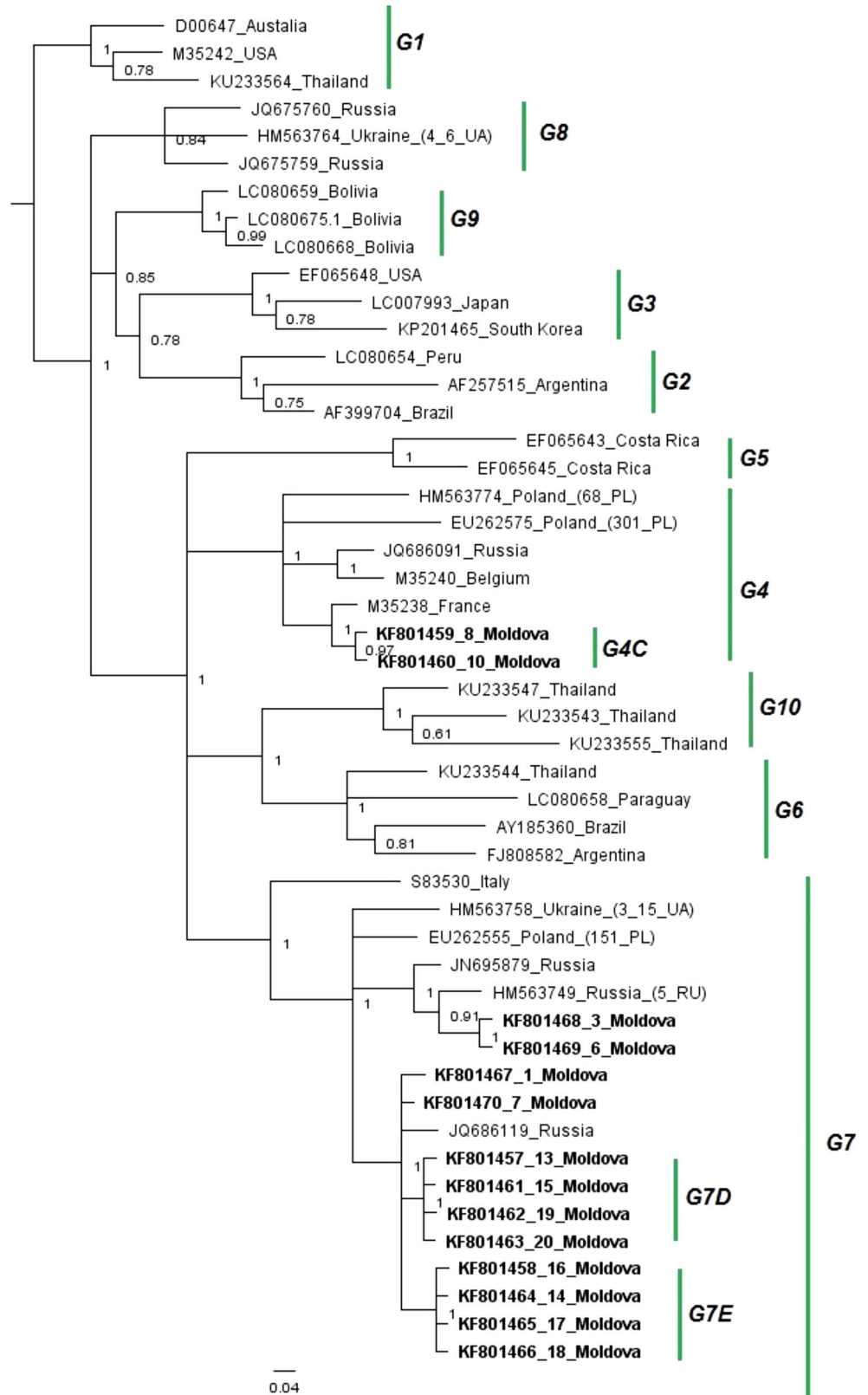
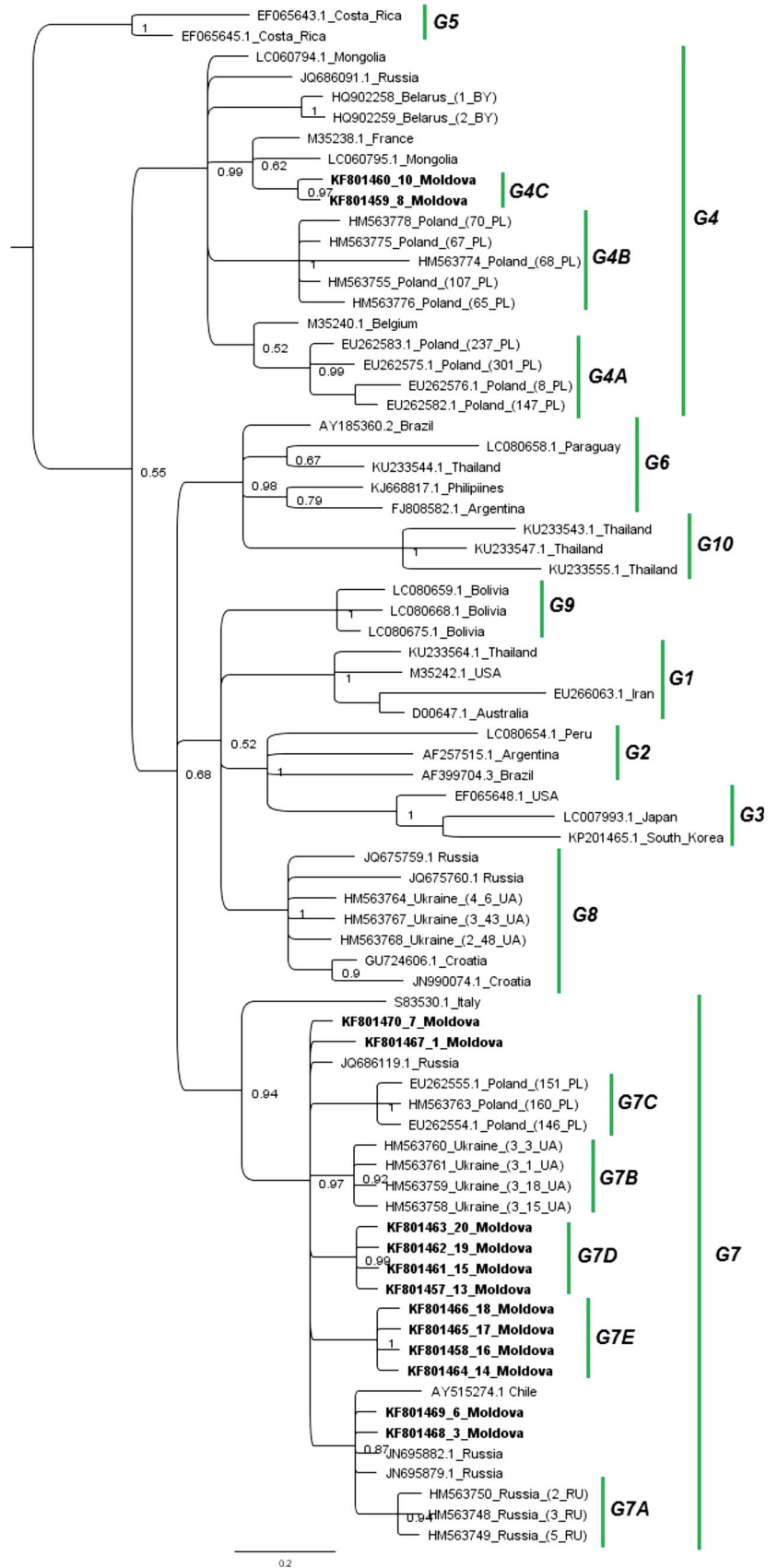


Fig. 2 Phylogenetic relationship of a 444-bp fragment of the *env* gene nucleotide sequences of new BLV subtypes found in this study (bold text) and all known BLV genotypes and/or subtypes (n = 75, updated as of August, 2016), as inferred by Bayesian analysis. Numbers at nodes indicate posterior probabilities of sampling the node among 11,000 trees. Subtype terminology is based on that of Rola-Luszczak et al. [8]. Genotypes and/or subtypes and new subtypes found in this study are indicated at the right by vertical lines



Nucleotide and amino acid sequence diversity of the region of BLV *env* encoding gp51

The mutations found in BLV strains from Moldovan cattle were mapped in the alignment shown in Fig. 3, with antigenic determinants indicated for comparison and clarity. Comparison to the consensus sequence revealed that 31 out of 804 nucleotide sites (3.9%) had a point variation in at least one Moldovan strain. The nucleotide variability (number of variations per site) in an individual strain varied from 1.6% to 2.1%, with an average value of 1.96%. Since nucleotide sequence variability was found throughout the *env* sequences, we analyzed the significance of these variations with respect to selection pressure on the *env* gene and individual sub-regions of gp51. To perform this analysis, the dN/dS ratio was calculated for each site, and the Z-test was used for selection testing. The results are shown in Table 2. While Z-test results showed extremely high dN/dS (>1) values for neutralizing epitope G, indicating its strong positive selection, negative selection (dN<dS) was found for the gp51-encoding sequence of the *env* gene as a whole and for most of its sub-regions. Although two functional domains, linear epitope B and the zinc-binding domain, showed somewhat higher dN/dS ratios (0.781 and 0.552, respectively), these were not high enough to indicate positive selection. Generally, these results indicated that the regions within *env* encoding functional domains and epitopes of gp51 of Moldovan strains were constrained by strong purifying negative selection.

Next, we analyzed whether nucleotide substitutions affected the amino acid (aa) composition along the full length of mature gp51, encompassing aa positions 34 to 301 and excluding aa 1–33, which comprise the cleaved signal peptide. A comparison of predicted aa sequences with consensus aa sequence revealed the presence of eight aa substitutions (3%), indicated in Fig. 3 by filled rectangles. While aa substitutions were found over the length of gp51, they mainly clustered in the segment associated with antigenic determinants. Some of the substitutions were observed in multiple samples. Interestingly, four out of eight substitutions occurred within the variable conformational epitopes G and H. We found substitutions of serine to phenylalanine at residue 56 within epitope H, alanine to proline at residue 73 within epitope G, arginine to histidine at residue 121 within epitope G, and isoleucine to threonine at residue 144 within the T-cell epitope CD8+N11, which also overlapped with the zinc-binding peptide and neutralizing domain 2. All of these substitutions were specifically linked to the strains 8MD and 10MD, belonging to the new G4D subtype. Four other aa substitutions were found to be characteristic of strains

belonging to G7. In four strains (14MD, 16MD, 17MD and 18MD) classified as subtype G7E, a unique substitution of tyrosine to cysteine at residue 229 was observed within epitope B; this substitution was only observed in Moldovan strains. Similarly, in the four strains grouped together in subtype G7D (13MD, 15MD, 19MD and 20MD), a substitution of asparagine to histidine at residue 141 within epitope CD8+N11 was observed. In addition, the twelve Moldovan strains belonging to G7 shared a substitution of alanine to valine at residue 291, within linear epitope A, and ten of these strains also shared a substitution of threonine to isoleucine at residue 48, within the neutralizing epitope G. These two last substitutions are not specific for Moldovan samples, having been previously described for strains from Russia, Italy and Argentina.

Discussion

In this study, the portion of the *env* gene encoding the gp51 glycoprotein of BLV was isolated from seropositive cattle in Moldova, and their nucleotide and amino acid sequences were analyzed. Genetic distance analysis of 14 sequences representing BLV isolated in Moldova and 35 sequences representing each of the known genotypes (G1–G10) revealed 95.3% to 99.8% sequence identity at the nucleotide level. Thus, our data confirmed the high degree of conservation of the gp51 *env* gene sequence among geographically different strains, as has been reported previously for many BLV isolates [7, 8, 10, 27–30].

The initial phylogenetic analysis was conducted with an 804-bp fragment of the BLV *env* gene. We aimed to include in this analysis as many geographically diverse strains as possible; however, this was limited by the number of available sequences representing the full-length gp51. Therefore, we needed to confirm these results for inferring the phylogenetic tree using a 444-bp fragment of gp51-encoding sequences which has been used most often in phylogenetic analysis of BLV [14, 30–32]. Another advantage of the use of the 444-bp fragment was that we were able to fully match phylogenetic data already generated in our previous study with BLV obtained from Eastern Europe and Siberia [8]. The topology of this phylogenetic tree clearly showed that Moldovan strains represent well-delineated groups within G4 and G7. Surprisingly, none of the strains from Moldova clustered in G8, the genotype that was thought to be specific for this geographical area. Instead, the Moldovan strains grouped primarily with the strains from the neighbouring countries of Ukraine and Russia [8]. Most of the Moldovan strains (85%) were clustered with G7 strains, and we showed the existence of two new well-defined homogenous sub-

Table 2 DnaSP analysis for evidence of selection in functional domains of BLV gp51 *env* sequences within the Moldovan population

Functional domains		dS	dN	Z-test	dN/dS ratio	dN/dS ratio*
Whole gp51	Overall	0.0304	0.0045	<0.0001	0.154	0.106
Sequential epitopes	A	0.0147	0.0073	<0.0001	0.501	0.462
	B, B'	0.0359	0.0198	<0.001	0.552	0.078
	D, D'	0.0223	0.0000	<0.0001	0.000	0.184
	E, E'	0.0167	0.0000	<0.0001	0.000	0.052
	F	0.0000	0.0000	>0.999**	0.000	0.000
	G	0.0756	0.0905	<0.001	1.198	2.787
	H	0.0000	0.0673	<0.0001	N/A***	0.218
Neutralization domains	ND1	0.0000	0.0000	>0.9999	0.000	0.137
	ND2	0.0452	0.0190	<0.0001	0.421	0.298
	ND3	0.0277	0.0000	<0.0001	0.00000	0.007
T-cell epitopes	CD4 ⁺	0.0456	0.0000	<0.0001	0.00000	0.036
	CD8 ⁺	0.0273	0.0000	<0.0001	0.00000	0.028
	CD8 ⁺ (N5)	0.0175	0.0000	<0.0001	0.00000	0.124
	CD8 ⁺ (N11)	0.0458	0.0177	<0.0001	0.388	0.194
	CD8 ⁺ (N12)	0.0219	0.0055	<0.0001	0.253	0.064
Zinc-binding peptide	Zbp	0.0239	0.0187	<0.001	0.781	0.217
Transmembrane hydrophobic region	TMHR	0.0919	0.0000	<0.0001	0.00000	0.152

* dS/dN calculated for ten genetic groups of BLV G1-G10

** Null hypothesis that dS=dN

*** N/A- not analyzed

genotypes, G7D and G7E, within which clustering occurred according to the place of origin, and they co-existed in the same herds. Correlation between genetic affiliation of the viruses and their geographical location has been demonstrated for many retroviruses, including HTLV-1 [33], FIV [34], and FeLV [35]. Similarly, the presence of more than one genotype of BLV in certain geographical areas has been reported by Camargos *et al.* [36], Zhao *et al.* [28] and Polat *et al.* [37]. In our previous work, we also reported the presence of well-separated BLV genogroups composed of isolates of similar geographical origin with a close genetic relationship to each other [8]. It seems that a tendency towards diversification of the virus into homogenous genetic groups is a common concept of global BLV diversification that results from virus transmission, possibly associated with genetic drift. Once dispersed, the virus is then assimilated within herd populations, and gradually, these became homogenous. Confirmation of such a scenario was found in the presence of the single amino acid substitution of tyrosine to cysteine at residue 229 that is unique to the G7E sub-genotype and the asparagine-to-histidine change at residue 141 that is typical of the G7D sub-genotype. Once introduced into small populations, these substitutions became fixed in the proviral genomes. Since the dissemination of BLV infection is associated mainly with introduction of infected cattle into a herd, we speculate

that animal trade that took place between countries belonging to CMEA (Council for Mutual Economic Assistance) played a critical role in modulation of the topology of the phylogenetic tree. The role of animal trade in BLV diversification has been emphasized in some studies [30, 32, 38]. Although we demonstrated that most Moldovan strains were of genotype G7, two strains clustered within G4, in close proximity to isolates from France, Mongolia and Belgium. These strains came from two distinct regions in Moldova where dairy cattle have been extensively back-crossed to Holstein-Friesian cattle, which were originally derived from the Dutch Friesian breed [39]. Since G4 represents the isolates that originally came from France and Belgium, this reinforces the concept of an impact of international animal trade on the distribution of particular genotypes and the topology of phylogenetic trees.

The *env* gene of BLV exhibits a relatively high degree of conservation among different strains overall, which is a distinctive feature of viruses belonging to the genus *Deltaretrovirus* [12, 33, 35]. Our analysis revealed only eight aa substitutions in Moldovan strains concentrated exclusively within known epitopes of the gp51 glycoprotein. This is in agreement with previous studies of isolates from other countries showing that most aa substitutions occur within epitopes rather than at random locations in the surface subunit [11, 28, 32]. The high degree of sequence

conservation is consistent with the general finding that the *env* gene is subjected to negative selection [28, 32]. However, strong positive selection has been observed for residues in the G and D-D' epitopes [28]. Similarly, we found a high dN/dS ratio for epitope G, indicating stringent positive selection for this conformational epitope. Three aa substitutions associated with this epitope highlight its role in virus evolution and the possible involvement of epitope G in the evolution of escape mutants. We also found one substitution within a second conformational epitope, H, at position 56, which was previously observed to influence epitope-H-specific antibody recognition [40]. Substitutions within epitopes G and H have been reported previously by several authors [12, 8, 28, 32, 41], but further studies are needed to determine whether these changes can affect the three-dimensional structure of the gp51 glycoprotein and to investigate their possible impact on the effectiveness of diagnostic serological tests. We speculate that the existence of BLV variants with mutations within epitopes G and H may explain the continued recurrence of BLV infection in cattle caused by variants that circumvent the immune response [14, 42]. We also report here two amino acid substitutions in the sequence of a CD8⁺ T-cell (N11) epitope that were identical to those recently found to induce T-cell-dependent cytotoxicity in BLV-infected calves [17]. These mutations were found in aa sequences deduced from nucleotide sequences of seropositive cattle; therefore, they did not seem to affect the diagnostic capacity of serological tests. However, their significance in the context of disease progression and host cellular anti-viral immunity remains to be elucidated. Dube *et al.* [43] described a unique mutation, E161G, in the C-terminal part of a CD8⁺ T-cell epitope of an Argentinian isolate that theoretically could alter the stimulation of the anti-CD8⁺ T-cell response. Furthermore, the CD8⁺ T-cell epitope (N11) overlaps with the zinc-binding peptide, which in all deltaretroviruses is indispensable for viral fusion and infectivity *in vivo* [18]. Thus, any aa changes within this region may strongly affect BLV infectivity.

In summary, here, we assessed the phylogenetic clustering of Moldovan BLV strains into three new subgroups of G4 and G7. Similar to what has been described previously for strains from Eastern Europe and Siberia [8], the Moldovan strains showed a tendency to group according to their geographical origin. Molecular characterization of the *env* gene fragment encoding envelope glycoprotein gp51 of these strains showed relatively high nucleotide sequence conservation, as has been reported for other BLV strains [11, 11, 31]. Nevertheless, we found some nucleotide changes in specific regions of the *env* gene and found that these regions were subjected to positive selection. Further studies will be desired to examine the impact of genetic diversity of

BLV on disease progression, virus infectivity, and the capacity of diagnostic tests.

Acknowledgements The authors would like to acknowledge the funding of this work: this research was funded by KNOW (Leading National Research Centre) Scientific Consortium “Healthy Animal - Safe Food”, decision of Ministry of Science and Higher Education of Poland No. 05-1/KNOW2/2015. The authors thank Lorraine Albritton (University of Tennessee Health Science Center) for the critical reading and editing of this manuscript.

Compliance with ethical standards

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

Ethical approval This article does not contain any studies with animals performed by any of the authors.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Gillet N, Florins A, Boxus M, Burteau C, Nigro A, Vandermeers F, Balon H, Bouzar AB, Defoiche J, Burny A, Reichert M, Kettmann R, Willems L (2007) Mechanisms of leukemogenesis induced by bovine leukemia virus: prospects for novel anti-retroviral therapies in human. *Retrovirology* 4:18
2. Lee E, Kim EJ, Ratthanophart J, Vitoonpong R, Kim BH, Cho IS, Song JY, Lee KK, Shin YK (2016) Molecular epidemiological and serological studies of bovine leukemia virus (BLV) infection in Thailand cattle. *Infect Genet Evol* 41:245–254
3. Nekouei O, VanLeeuwen J, Sanchez J, Kelton D, Tiwari A, Keefe G (2015) Herd-level risk factors for infection with bovine leukemia virus in Canadian dairy herds. *Prev Vet Med* 119(3–4):105–113. doi:10.1016/j.prevetmed.2015.02.025
4. Morovati H, Shirvani E, Noaman V, Lotfi M, Kamalzadeh M, Hatami A, Bahreyari M, Shahramyar Z, Morovati MH, Azimi M, Sakhaei D (2012) Seroprevalence of bovine leukemia virus (BLV) infection in dairy cattle in Isfahan Province, Iran. *Trop Anim Health Prod* 44(6):1127–1129. doi:10.1007/s11250-011-0062-4
5. Murakami K, Kobayashi S, Konishi M, Kameyama K, Tsutsui T (2013) Nationwide survey of bovine leukemia virus infection among dairy and beef breeding cattle in Japan from 2009–2011. *J Vet Med Sci* 75(8):1123–1126
6. Balic D, Lojic I, Periskic M, Bedekovic T, Jungic A, Lemo N, Roic B, Cac Z, Barbic L, Madic J (2012) Identification of a new genotype of bovine leukemia virus. *Arch Virol* 157:1281–1290. doi:10.1007/s00705-012-1300-4
7. 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* 90:2788–2797. doi:10.1099/vir.0.011791-0
8. Rola-Łuszczak M, Pluta A, Olech M, Donnik I, Petropavlovskiy M, Gerilovych A, Vinogradova I, Choudhury B, Kuźmak J (2013) The molecular characterization of bovine leukemia virus

- isolates from Eastern Europe and Siberia and its impact on phylogeny. *PLoS One* 8(3):e58705. doi:[10.1371/journal.pone.0058705](https://doi.org/10.1371/journal.pone.0058705)
9. Vafin RR, Khazipov NZ, Shaeva AY, Zakirova ZR, Zaynullin LI, Tyulkin SV, Abdulina IR, Alimov AM (2013) Finding a new (8th) genotype of BLV in various regions of the world. *Fundam Res* 10:1467–1471
 10. Lee E, Kim EJ, Joung HK, Kim BH, Song JY, Cho IS, Lee KK, Shin YK (2015) Sequencing and phylogenetic analysis of the gp51 gene from Korean bovine leukemia virus isolates. *Viol J* 12:64. doi:[10.1186/s12985-015-0286-4](https://doi.org/10.1186/s12985-015-0286-4)
 11. Polat M, Takeshima SN, Hosomichi K, Kim J, Miyasaka T, Yamada K, Arainga M, Murakami T, Matsumoto Y, Barra Diaz V, Panei CJ, González ET, Kanemaki M, Onuma M, Giovambattista G, Aida Y (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. doi:[10.1186/s12977-016-0239-z](https://doi.org/10.1186/s12977-016-0239-z)
 12. Mamoun RZ, Morisson M, Rebeyrotte N, Busetta B, Couez D (1990) Sequence variability of bovine leukemia virus env gene and its relevance to the structure and antigenicity of the glycoproteins. *J Virol* 64:4180–4188
 13. Rice NR, Stephens RM, Couez D, Deschamps J, Kettmann R, Burny A, Gilden RV (1984) The nucleotide sequence of the env gene and post-env region of bovine leukemia virus. *Virology* 138(1):82–93
 14. Fechner H, Blankenstein P, Looman AC, Elwert J, Geue L, Albrecht C, Kurg A, Beier D, Marquardt O, Ebner D (1997) Provirus variants of the bovine leukemia virus and their relation to the serological status of naturally infected cattle. *Virology* 237(2):261–269
 15. Portetelle D, Dandoy C, Burny A, Zavada J, Siakkou H, Gras-Masse H, Drobecq H, Tartar A (1989) Synthetic peptides approach to identification of epitopes on bovine leukemia virus envelope glycoprotein gp51. *Virology* 169:34–41. doi:[10.1016/0042-6822\(89\)90038-X](https://doi.org/10.1016/0042-6822(89)90038-X)
 16. Callebaut I, Voneche V, Mager A, Fumiere O, Krchnak V, Merza M, Zavada J, Mammerrickx 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* 67:5321–5327
 17. Bai L, Takeshima SN, Isogai E, Kohara J, Aida Y (2015) Novel CD8(+) cytotoxic T cell epitopes in bovine leukemia virus with cattle. *Vaccine* 33(51):7194–7202. doi:[10.1016/j.vaccine.2015.10.128](https://doi.org/10.1016/j.vaccine.2015.10.128)
 18. Gatot JS, 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
 19. Hazipov NZ, Tyurikova RP (2008) Reproduktsiya virusa leykoza krupnogo rogatogo skota i immunnyiy otvet na nee. *UchYonyie zapiski KGAVM* 192:152–160
 20. Moskalik RS (2013) Prostoe reshenie slozhnoy problemyi likvidatsii leykoza KRS. *Selskazyaysmvennyy zhurnal* 5(31):24–25
 21. Moskalik RS, Gangal NI, Balova SZ (2014) Effektivnaya realizatsiya moldavskoy sistemyi iskoreneniya leykoza krupnogo rogatogo skota. *Veterinarna Meditsina* 98:91–95
 22. Markiewicz L, Rulka J, Kaminski S (2003) Detection of BLV provirus in different cells by nested-PCR. *Bull Vet Inst Pulawy* 47:325–331
 23. 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* 28:2731–2739
 24. 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* 28:2731–2739
 25. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452
 26. Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A (2010) Geneious v5.5, biomatters. <http://www.geneious.com>. Accessed 11 Feb 2013
 27. Camargos MF, Stancek D, Rocha MA, Lessa LM, Reis JK (2002) Partial sequencing of env gene of bovine leukaemia virus from Brazilian samples and phylogenetic analysis. *J Vet Med B Infect Dis Vet Public Health* 49:325–331. doi:[10.1046/j.1439-0450.2002.00582.x](https://doi.org/10.1046/j.1439-0450.2002.00582.x)
 28. Zhao X, Buehring GC (2007) Natural genetic variations in bovine leukemia virus envelope gene: possible effects of selection and escape. *Virology* 366:150–165
 29. Dube S, Bachman S, Spicer T, Love J, Choi D, Esteban E, Ferrer JF, Poiesz BJ (1997) Degenerate and specific PCR assays for the detection of bovine leukaemia virus and primate T cell leukaemia/lymphoma virus pol DNA and RNA: phylogenetic comparisons of amplified sequences from cattle and primates from around the world. *J Gen Virol* 78(Pt 6):1389–1398
 30. Alkan F, Oğuzoğlu TÇ, Timurkan MO, Karapınar Z (2011) Characterisation of env and gag gene fragments of bovine leukemia viruses (BLVs) from cattle in Turkey. *Arch Virol* 156(10):1891–1896. doi:[10.1007/s00705-011-1060-6](https://doi.org/10.1007/s00705-011-1060-6)
 31. Felmer R, Muñoz G, Zúñiga J, Recabal M (2005) 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* 108(1–2):39–47
 32. Monti G, Schrijver R, Beier D (2005) Genetic diversity and spread of Bovine leukaemia virus isolates in Argentine dairy cattle. *Arch Virol* 150(3):443–458
 33. Slattey JP, Franchini G, Gessain A (1999) Genomic evolution, patterns of global dissemination, and interspecies transmission of human and simian T-cell leukemia/lymphotropic viruses. *Genome Res* 9(6):525–540
 34. Reggeti F, Bienzle D (2004) Feline immunodeficiency virus subtypes A, B and C and intersubtype recombinants in Ontario, Canada. *J Gen Virol* 85(Pt 7):1843–1852
 35. Watanabe S, Kawamura M, Odahara Y, Anai Y, Ochi H, Nakagawa S, Endo Y, Tsujimoto H, Nishigaki K (2013) Phylogenetic and structural diversity in the feline leukemia virus env gene. *PLoS One* 8(4):e61009. doi:[10.1371/journal.pone.0061009](https://doi.org/10.1371/journal.pone.0061009)
 36. Camargos MF, Pereda A, Stancek D, Rocha MA, dos Reis JK, Greiser-Wilke I, Leite RC (2007) Molecular characterization of the env gene from Brazilian field isolates of Bovine leukemia virus. *Virus Genes* 34(3):343–350
 37. Polat M, Ohno A, Takeshima SN, 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. doi:[10.1007/s00705-014-2280-3](https://doi.org/10.1007/s00705-014-2280-3)
 38. Licursi M, Inoshima Y, Wu D, Yokoyama T, González ET, Sentsui H (2003) Provirus variants of bovine leukemia virus in naturally infected cattle from Argentina and Japan. *Vet Microbiol* 96(1):17–23
 39. Chilimar S, Focsa V (2012) New type bovine of Black and White breed. Translated from Universitatea de Științe Agricole și Medicină Veterinară Iași, Lucrări Științifice 52:74–76, Seria Zootehnie; I. Parasca and Chilimar, Quality of milk from cows of the Moldavian type of Black & White and Red Estonian breeds. *Lucrări Științifice - Seria Zootehnie* 58:131–136

40. Johnston ER, Albritton LM, Radke K (2002) Envelope proteins containing single amino acid substitutions support a structural model of the receptor-binding domain of bovine leukemia virus surface protein. *J Virol* 76(21):10861–10872
41. Bruck C, Renzonnet N, Portetelle D, Cleuter Y, Mammerickx M, Burny A, Mamoun R, Guillemain B, van der Maaten MJ, Ghysdael J (1984) Biologically active epitopes of bovine leukemia virus glycoprotein gp51: their dependence on protein glycosylation and genetic variability. *Virology* 136(1):20–31
42. 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(1–2):101–110
43. Dube S, Abbott L, Dube DK, Dolcini G, Gutierrez S, Ceriani C, Juliarena M, Ferrer J, Perzova R, Poiesz BJ (2009) The complete genomic sequence of an in vivo low replicating BLV strain. *Virology* 396(1):115–120. doi:[10.1016/j.virol.2009.06.015](https://doi.org/10.1016/j.virol.2009.06.015)
44. Ochirkhuu N, Konnai S, Odbileg R, Nishimori A, Okagawa T, Murata S, Ohashi K (2015) Detection of bovine leukemia virus and identification of its genotype in Mongolian cattle. *Arch Virol* 161(4):985–991. doi:[10.1007/s00705-015-2676-8](https://doi.org/10.1007/s00705-015-2676-8)
45. Hemmatzadeh F (2007) Sequencing and phylogenetic analysis of gp51 gene of bovine leukaemia virus in Iranian isolates. *Vet Res Commun* 31(6):783–789
46. Coulston J, Naif H, Brandon R, Kumar S, Khan S, Daniel RC, Lavin MF (1990) Molecular cloning and sequencing of an Australian isolate of proviral bovine leukaemia virus DNA: comparison with other isolates. *J Gen Virol* 71:1737–1746
47. Dube S, Dolcini G, Abbott L, Mehta S, Dube D, Gutierrez S et al (2000) The complete genomic sequence of a BLV strain from a Holstein cow from Argentina. *Virology* 277(2):379–386
48. Molteni E, Agresti A, Meneveri R, Marozzi A, Malcovati M, Bonizzi L, Poli G, Ginelli E (1996) Molecular characterization of a variant of proviral bovine leukaemia virus (BLV). *Zentralbl Veterinarmed B* 43(4):201–211
49. Mason IL (1996) A World dictionary of livestock breeds, types and varieties, 4th edn, vol 57. C.A.B International, State of Russian Black Pied cattle breeding in Vologda region. Rozov, N. N. *Trudy vologod moloch Inst* 1970, pp 146–164
50. Willems L, Burny A, Collete D, Dangoisse O, Dequiedt F, Gatot JS, Kerkhofs P, Lefèbvre L, Merezak C, Peremans T, Portetelle D, Twizere JC, Kettmann R (2000) Genetic determinants of bovine leukemia virus pathogenesis. *AIDS Res Hum Retroviruses* 16(16):1787–1795