





Article

Phylogenetic Analysis of South African Bovine Leukaemia Virus (BLV) Isolates

Akiko Suzuki ^{1,†}, Rosamund Chapman ^{1,2,*,†}, Nicola Douglass ^{1,2}, Olivia Carulei ¹,
Johan van Rensburg ³ and Anna-Lise Williamson ^{1,2}

¹ Division of Medical Virology, Department of Pathology, Faculty of Health Sciences, University of Cape Town, Cape Town 7925, South Africa; szkaki001@myuct.ac.za (A.S.); niki.douglass@uct.ac.za (N.D.); olivia.carulei@uct.ac.za (O.C.); anna-lise.williamson@uct.ac.za (A.-L.W.)

² Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town 7925, South Africa

³ Afrivet, Johannesburg 0081, South Africa; johan.vanrensborg@afrivet.co.za

* Correspondence: ros.chapman@uct.ac.za

† These authors contributed equally to this work.

Received: 13 July 2020; Accepted: 10 August 2020; Published: 17 August 2020



Abstract: Bovine leukaemia virus (BLV) causes chronic lymphoproliferative disorder and fatal lymphosarcoma in cattle, leading to significant economic losses in the beef and dairy industries. BLV is endemic globally and eleven genotypes have been identified. To date, only Zambian isolates have been genotyped from Africa. Although high BLV prevalence has been reported in South Africa, there has been no molecular characterisation of South African BLV isolates. To characterise BLV isolates in South Africa for the first time, we investigated the phylogenetic relationships and compared the genetic variability of eight South African BLV isolates with BLV isolates representing the eleven known genotypes from different geographical regions worldwide. Phylogenetic analyses based on full-length and partial *env* sequences as well as full-length *gag* sequences revealed that at least two genotypes, genotypes 1 (G1) and 4 (G4), are present in cattle in South Africa, which is consistent with studies from Zambia. However, our analysis revealed that the G1 South African isolate is more similar to other G1 isolates than the G1 Zambian isolates whereas, the G4 South African isolates are more divergent from other G4 isolates but closely related to the G4 Zambian isolate. Lastly, amino acid sequence alignment identified genotype-specific as well as novel amino acid substitutions in the South African isolates. The detection of two genotypes (G1 and G4) in southern Africa highlights the urgent need for disease management and the development of an efficacious vaccine against local strains.

Keywords: bovine leukaemia virus; phylogenetic analysis; BLV envelope; BLV *gag*; South African BLV genotypes

1. Introduction

Bovine leukosis is the most common neoplastic disease in cattle and consists of two types of leukosis: enzootic bovine leukosis (EBL) and sporadic bovine leukosis (SBL). The occurrence of EBL is more frequent than that of SBL [1,2]. SBL is not transmissible and the causative agent of SBL remains unknown. The aetiological agent of EBL is bovine leukaemia virus (BLV), which is a B-lymphotropic oncovirus, belonging to the *Deltaretrovirus* genus of the *Retroviridae* family [3]. BLV naturally infects cattle, water buffalo and zebu [4]. The majority of transmission is mediated by direct transfer of infected body fluids during iatrogenic procedures such as ear tattooing, dehorning, re-use of needles gloves and plastic sleeves for veterinary procedures, including rectal palpation for the detection of

pregnancy and artificial insemination for embryo transfer [5–9], which are common practices in the management of dairy and beef cattle.

EBL is a chronic infection characterised by abnormal proliferation of non-neoplastic B cells during the persistent lymphocytosis (PL) stage culminating in the development of malignant neoplastic lymphosarcoma in various organs [10,11]. The lymphosarcoma stage is accompanied by increased monoclonal populations of neoplastic B lymphocytes and eventually leads to death [12].

The complete BLV RNA genome consists of 8714 bp. As a complex retrovirus, the BLV genome has a conserved modular genomic organisation with a set of structural genes (*gag*, protease (*pro*), polymerase (*pol*) and *env*) at the 5' end of the genome, followed by accessory genes (*G4*, *R3* and microRNAs) and regulatory genes (*tax*, *rex*) at the 3' end. Two identical long terminal repeats (LTRs) are present at the 5' and 3' ends of the genome. The BLV envelope (Env) proteins are glycoproteins presented on the surface of the virion as a trimer of surface (gp51) and transmembrane protein (gp30) heterodimers. The *env* gene is 1548bp in size and is translated first into the Env precursor (pr72), which is subsequently cleaved by subtilisin/kexin-like convertases (e.g., furin) into the gp51 and gp30 glycoproteins [13]. The gp51 surface protein is rich in binding domains and epitopes such as putative receptor binding domain (RBD) [14,15], antibody epitopes (conformational and linear epitopes) [16–21], T-cell epitopes [22,23], neutralising domains [19,24–26] and Zn²⁺-binding site [15]. These domains and epitopes are important in cell fusion and entry during natural infection as well as immunogenicity. The gp30 transmembrane protein contains a fusion peptide at the N-terminus, which is important for cell fusion and entry [27,28], and a cytoplasmic tail at the C-terminus, which is essential for viral entry and immune evasion [29–31]. Unlike gp51 [32,33], gp30 is not rich in domains and epitopes and is poorly immunogenic [34].

Gag is another structural protein essential for BLV virion formation. The *gag* gene is 1182 bp in size and encodes the Gag-Pro precursor (pr66) from the Gag-Pro-Pol mRNA [35,36]. The Gag-Pro precursor is then, proteolytically cleaved by the viral protease into the Gag precursor (pr45) [37]. The Gag pr45 is further processed by the viral protease into the matrix (p15), capsid (p24) and nucleocapsid (p12) proteins [37]. The capsid is another major target of the host immune response but unlike anti-gp51 antibodies, which are persistently detectable in leukemic and aleukemic animals with low proviral loads, the anti-p24 antibodies are undetectable in aleukemic animals with low proviral loads [38].

Currently, BLV is endemic globally and is detected on all continents, except Antarctica [39]. However, twenty-two European countries had achieved EBL-free status by 2017 and EBL remains present in only seven countries in Europe (Bulgaria, Croatia, Greece, Hungary, Malta, Portugal and Romania) [40]. As of 2019, BLV isolates have been classified into eleven genotypes (G1 to G11). In 2009, Rodriguez et al. [41] conducted a seminal study by comparing the BLV full-length *env* sequences of 28 Argentine isolates and 46 sequences that had been previously described in various studies but had no standardised classification amongst different studies. Using neighbour-joining, Bayesian, maximum-likelihood and parsimony methods, they re-classified these isolates into seven genotypes (G1 to G7). In 2012, genotype 8 (G8) was identified in Croatia [42] and subsequently in Ukraine and Russia in 2013 [43]. In 2016, using BLV whole-genome sequences of G1, G2, G4 and G6 strains as well as 25 newly identified isolates from Peru, Bolivia and Paraguay, G9 was identified from Bolivian isolates by Polat et al. [44]. Interestingly, in the last several years, new genotypes have been identified in Asia. G10 was identified in Thailand in 2016 [45], Myanmar in 2017 [46] and China in 2018 [47]. Furthermore, G11 was identified in China in 2019 [48]. BLV whole-genome sequences are available from G1, G2, G4, G6, G9 and G10.

In addition to the global distribution of eleven genotypes, multiple genotypes are circulating in Western Europe, Asia and South America. Some countries in these regions with multiple detected genotypes such as America, Bulgaria, Japan and Argentina have a long history of BLV prevalence or high BLV prevalence rates [40,49–55], indicating that not only may the prolonged or delayed eradication process have resulted in the persistence of BLV but also allowed the introduction of new and multiple genotypic strains. This is also a serious concern with respect to the development of an efficacious BLV vaccine.

To date, only Zambian BLV isolates have been genotyped from Africa. Although high BLV prevalence has been reported in South Africa (Afrivet, 2018), there has been no molecular characterisation of South African BLV isolates. Therefore, this study was aimed to sequence BLV *env* and *gag* genes from South African isolates in order to investigate phylogenetic relationships and compare genetic variability of the South African isolates with other BLV isolates representing the eleven known genotypes from different geographical regions worldwide. Our analyses revealed that at least two genotypes, G1 and G4, are currently circulating in South Africa. Pairwise comparison and intragenotype evolutionary distances of full-length Env nucleotide and amino acid sequences showed that the G4 South African isolates are closely related to each other and their sequences are highly conserved. Furthermore, our analyses using full-length Gag nucleotide and amino acid sequences indicated that Gag sequences are generally more variable than Env sequences. Lastly, the alignment of BLV Env and Gag amino acid sequences identified genotype-specific as well as novel substitutions in the South African isolates.

2. Materials and Methods

2.1. Blood Samples and Extraction of Total Bovine Genomic DNA

Whole blood samples were obtained from a single herd of dairy cows in Port Elizabeth in South Africa in 2018. The blood samples were delivered to the University of Cape Town, South Africa for further analyses. Genomic DNA was extracted from whole blood samples ($n = 8$) using New MagNA Pure Compact Nucleic Acid Isolation Kits on the MagNA Pure Compact Instrument (Roche, Switzerland) as per the manufacturer's instructions. The extracted DNA was stored at $-20\text{ }^{\circ}\text{C}$ until needed. The eight blood samples were designated sample IDs: K1170, K1194, M1878, M2746, P591, P2152, P2677 and L3401.

2.2. Amplification of BLV *env* Gene by Nested Polymerase Chain Reaction (nPCR) and Sanger Sequencing of Full-Length Env Gene

A 1782 bp fragment containing the full-length BLV *env* gene was amplified from total bovine DNA extracted from the eight blood samples by nPCR using the following primers (Figure 1): Env OF (Outer Forward) (5'-CCTCCTACCAATTCTAAAGACC-3'), Env OR (Outer Reverse) (5'-CACGCAGAAGCGACAATCTC-3'), Env if (Inner Forward) (5'-GGGCGGAGAAACACCCYAAGG-3') and Env ir (Inner Reverse) (5'-CACTGACTATTCCACTAAGCC-3'). Both rounds of nPCR were performed in a 20 μL reaction mixture containing 1 μL of total genomic DNA or PCR product from the first round of nPCR, 1 μL of 2x KAPA 2G Robust HotStart ReadyMix (KAPA BIOSYSTEMS, Wilmington, MA, USA), 1 μL of each primer (10 μM each) and high-performance liquid chromatography (HPLC) water (Sigma-Aldrich, St Louis, MO, USA). The following thermocycling parameters were used: initial denaturation at $95\text{ }^{\circ}\text{C}$ for 3 min followed by 25 cycles (30 cycles for the second round) of denaturation at $95\text{ }^{\circ}\text{C}$ for 15 s, annealing at $55\text{ }^{\circ}\text{C}$ ($61\text{ }^{\circ}\text{C}$ for the second round) for 15 s, extension at $72\text{ }^{\circ}\text{C}$ for 2 min and final extension at $72\text{ }^{\circ}\text{C}$ for 4 min.

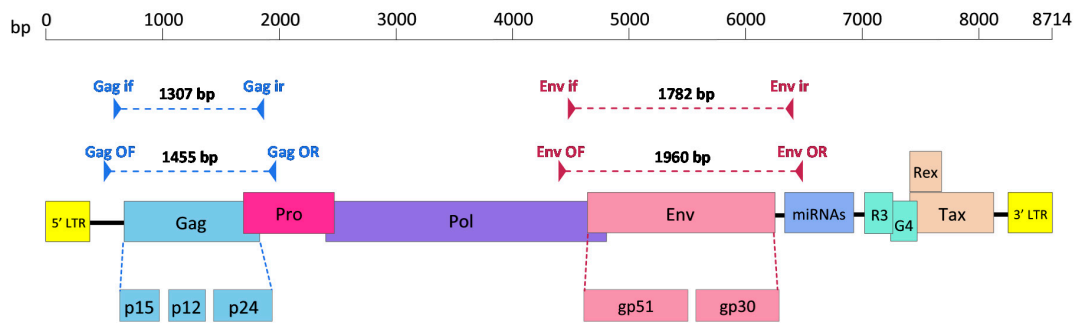


Figure 1. Schematic diagram showing open reading frames of the bovine leukaemia virus (BLV) genome and *gag* and *env* regions amplified using nested PCR (nPCR). Triangles denote primers and dotted lines denote amplified regions. Gag if, Gag inner forward primer; Gag ir, Gag inner reverse primer; Gag OF, Gag outer forward primer; Gag OR, Gag outer reverse primer; Env if, Env inner forward primer; Env ir, Env inner reverse primer; Env OF, Env outer forward primer; Env OR, Env outer reverse primer; Pro, protease; Pol, polymerase; Env, envelope; miRNAs, microRNAs; p15, matrix protein; p12, nucleocapsid protein; p24, capsid protein; gp51, surface protein; gp30, transmembrane protein. Adapted from Polat et al. (2017) [56].

PCR products were gel purified from 0.8% agarose TBE gels using the ZymoClean Gel DNA Recovery Kit (Zymo Research, Irvine, CA, USA) and sent to the Central Analytical Facility at the University of Stellenbosch (South Africa) for Sanger sequencing. The following primers were used: Env fwd1 (5'-AGGCGCTCTC CTGGCTAC-3'), Env fwd2 (5'-CARGTCTCCAGATACAC-3'), Env fwd3 (5'-TGGGGATATGATCCCCTG-3'), Env fwd4 (5'-AATGTTTCTCAAGGCAAC-3'), Env fwd5 (5'-CCCAGAACCGACGGGGGC-3'), Env fwd6 (5'-AATGCTTGACCTCTCGCC-3'), Env rev1 (5'-ATCTGCCCCACATAAGG-3'), Env rev2 (5'-GATGGTTTTGTTATATAC-3'), Env rev3 (5'-AGTCTCTGATGGCTAAGG-3'), Env rev4 (5'-GACTCTTTGCGAGAGAGG-3') and Env rev5 (5'-ACAGCCTGGGGGTGCGTG-3'). The BLV *env* sequences from samples K1170, K1194, L3401, M1878, M2746, P2152, P2677 and P591 were deposited in the GenBank sequence database under the respective accession numbers MN966688, MN966689, MN966690, MN966691, MN966692, MN966693, MN966694, MN966695 (Supplementary Table S1).

2.3. Amplification of BLV Gag Gene by nPCR and Sanger Sequencing of Full-Length Gag Gene

A 1307 bp fragment containing the full-length BLV *gag* gene was amplified from the total bovine DNA extracted from the eight blood samples by nPCR using following primers (Figure 1): Gag OF (5'-GATTGATCACCCCGGAACCC-3'), Gag OR (5'-GTATTTTCAGCCCCGGTGTCC-3'), Gag if (5'-CTCTGGACCCACCCCTCG-3') and Gag ir (5'-CATTCTARTTCGGCCTCACTAAG-3'). Both rounds of nPCR were conducted in a 20 μ L reaction mixture containing 5 μ L of total genomic DNA or 1 μ L of the PCR product from the first round of nPCR, 1 μ L of 2 \times KAPA 2G Robust HotStart ReadyMix (KAPA BIOSYSTEMS, Wilmington, MA, USA), 1 μ L of each primer (10 μ M each) and high-performance liquid chromatography (HPLC) water (Sigma-Aldrich, St Louis, MO, USA). The following thermocycling parameters were used: initial denaturation at 95 $^{\circ}$ C for 3 min followed by 40 cycles (35 cycles for the second round) of denaturation at 95 $^{\circ}$ C for 15 s, annealing at 56 $^{\circ}$ C (61 $^{\circ}$ C for the second round) for 15 s, extension at 72 $^{\circ}$ C for 2 min and final extension at 72 $^{\circ}$ C for 4 min.

PCR products were gel purified and sequenced as described earlier. The following primers were used: Gag fwd1 (5'-CCCTCGGCGGCRTTTTGG-3'), Gag fwd2 (5'-CCTYTGGCTTCAGGCCTG-3'), Gag fwd3 (5'-GACCATGCTATCGATGCC-3'), Gag rev1 (5'-GGGTTGTTCYTCKGGGGC-3'), Gag rev2 (5'-GRTTAAACCCCTGGAGGG-3'), Gag rev3 (5'-YCCCACCGGGGCGGCCAC-3'), Gag rev4 (5'-CCTCACTAAGRGRATCTG-3'). The BLV *gag* sequences from the samples K1170, K1194, L3401, M1878, M2746, P2152, P2677 and P591 were deposited in the GenBank sequence database under the respective accession numbers MN966696, MN966697, MN966698, MN966699, MN966700, MN966701, MN966702 and MN966703 (Supplementary Table S1).

2.4. Phylogenetic Analysis of BLV Full-Length and Partial Env and Full-Length Gag Nucleotide Sequences

The BLV *env* and *gag* nucleotide sequences of the eight South African isolates were assembled and aligned using CLC Main Workbench 7.9.3 (QIAGEN Bioinformatics, Copenhagen, Denmark). To select reference BLV nucleotide sequences for downstream analyses, 203 BLV full-length (1548 bp) *env*, 564 partial (444 bp) *env* and 113 full-length (1182 bp) *gag* nucleotide sequences with known genotypes were obtained from the Basic Local Alignment Search Tool (BLAST) database [57] using eight South African *env* and *gag* nucleotide sequences as query sequences; identical and highly similar sequences were removed using 98.5% sequence similarity as a threshold value in VSEARCH [58]. The selected sequences were further adjusted so that they represented isolates from a wide range of countries (Supplementary Table S2). The final sequences were aligned with the eight South African isolates using MUSCLE [59]. Model testing was performed to select the best evolutionary model based on the Bayesian information criterion (BIC) [60].

Neighbour-joining (NJ) and Maximum-likelihood (ML) trees based on the BLV full-length and 444 bp partial *env* nucleotide sequences were constructed using the Tamura-Nei substitution model with the Gamma distribution (TN83 + G) [61] and the Kimura-2-parameter model with the Gamma distribution (K2P + G) [62], respectively. NJ and ML trees based on the BLV full-length *gag* nucleotide sequences were constructed using the TN83 + G [61] model with the Gamma distribution (HKY + G) [63]. Nonparametric bootstrap analysis with 1000 iterations [64] was used to evaluate the robustness of evolutionary relationships. Alignment, model testing and tree building were performed using MEGA 7 [59].

2.5. Pairwise Comparison of BLV Env and Gag Nucleotide and Amino Acid Sequence

Pairwise evolutionary distances were estimated using the TN83 + G model for the BLV full-length *env* and *gag* nucleotide sequences and the K2P + G model for the partial *env* nucleotide sequences using MEGA7 [59]. Nonparametric bootstrap analysis with 1000 iterations [64] was used to evaluate the robustness of evolutionary relationships. For the full-length Env and partial Env sequences, the analysis was conducted only between South African isolates and selected G1 and G4 isolates from other geographic regions worldwide.

To evaluate the nucleotide and amino acid sequence percent identity and differences, pairwise comparison of the BLV full-length Env, partial Env and full-length Gag sequences between the South African isolates and selected isolates from other geographic regions (46, 20 and 29 sequences, respectively) was performed using CLC Main Workbench 7.9.3 (QIAGEN Bioinformatics, Copenhagen, Denmark). For the partial Env sequences, sequence comparison was made only between South African isolates and selected G1 and G4 isolates from other geographic regions worldwide.

2.6. Intragenotype and Intergenotype Evolutionary Distances of BLV Env Nucleotide and Amino Acid Sequences

Two hundred and three full-length *env* nucleotide sequences of BLV isolates within nine known genotypes that were obtained from the BLAST database [57] in Section 2.4 were used for this analysis. Translated amino acid sequences were aligned using MUSCLE in MEGA7 [59]. Mean intragenotype and intergenotype evolutionary distances of BLV Env nucleotide and amino acid sequences were estimated using the TN83 + G model [45] and the Jones–Taylor–Thornton model with the Gamma distribution (JTT + G) [49], respectively, using MEGA7 [59]. Nonparametric bootstrap analysis with 1000 iterations [64] was used to evaluate the robustness of evolutionary relationships.

2.7. Alignment of BLV Env and Gag Amino Acid Sequences

To identify genotype-specific amino acid substitutions as well as those unique to the South African isolates, the full-length Env and Gag amino acid sequences of the South African isolates were aligned with 851 Env amino acid sequences of BLV isolates with eleven known genotypes and with 113 Gag amino acid sequences with six known genotypes (G1, G2, G4, G6, G9 and G10) obtained from the GenBank sequence database using MUSCLE in MEGA7 [59].

The NJ and ML trees revealed that seven of the South African isolates (K1170, K1194, M1878, P591, P2152, P2677 and L3401) are closely related to each other with 99.8% bootstrap value and all belong to G4. One South African isolate (M2746) did not cluster with the seven other South African isolates and grouped with G1. This phylogenetic analysis also indicated that the G4 South African isolates are more divergent from other G4 isolates, whereas the G1 South African isolate is less divergent from other G1 isolates. The G4 South African isolates appeared to be moderately related to a Russian isolate (JN695878.1) with 74.5% (NJ) and 75.1% (ML) bootstrap values and their estimated evolutionary divergences were 0.0132 (P2677 and P591), 0.0139 (K1170, K1194, M1878 and P2152) and 0.0153 (L3401) (Supplementary Figure S2). In contrast, the G1 South African isolate clustered with Saint Kitts and Nevis isolates (KX674367.1 and KX674370.1) with 90.5% (NJ) and 88% (ML) bootstrap values and their estimated evolutionary divergence was 0.0046 (KX674367.1) and 0.0039 (KX674370.1) (Supplementary Figure S2).

3.2. Phylogenetic Analysis of BLV Partial Env Nucleotide Sequences (444 bp)

The only African BLV sequences that have been sequenced are partial (444 bp) sequences of the *gp51* gene from Zambian isolates [65,66]. Furthermore, since the partial *env* sequences have been used for BLV genotyping in various studies [45,67–72], they are available from 10 genotypes from various geographic regions. To confirm the results obtained from the phylogenetic analyses based on the full-length *env* nucleotide sequences and to further investigate the phylogenetic relationships between the South African isolates with more diverse BLV isolates from different geographic regions, NJ (Supplementary Figure S3) and ML (Figure 3) trees based on the alignment of partial *env* nucleotide sequences were constructed using 112 BLV isolates from 32 countries.

Despite the low to moderate bootstrap values, this second analysis was consistent with the first analysis using the full-length *env* nucleotide sequences, showing that seven of the South African isolates belong to G4 and one (M2746) belongs to G1. G2 sequences in the NJ tree clustered differently from the ML tree due to the difference in tree-building methods (distance-based vs. character-based methods) but the ML tree based on the partial *env* nucleotide sequences was congruent with the phylogenetic tree based on the full-length *env* nucleotide sequences. The seven G4 South African isolates clustered within the G4 group with 44% (NJ) and 44.2% (ML) bootstrap values whilst the South African isolate (M2746) clustered with other G1 isolates with 70.6% (NJ) and 66% (ML) bootstrap value. The G1 South African isolate did not cluster with three selected G1 Zambian isolates (LC440653.1, LC440663.1 and LC440666.1), which form a distinct monophyletic group as in Phiri's study [66]. In contrast, despite low bootstrap values (32.5% for NJ and 36.4% for ML), the G4 South African isolates are more closely related to a G4 Zambian isolate (LC193462.1) than other G4 isolates. The estimated evolutionary distance between the seven G4 South African isolates and the G4 Zambian isolate was 0.0023 (K1170, K1194, P2677 and P591) and 0.0045 (L3401, M1878 and P2152) (Supplementary Figure S4).

3.3. Phylogenetic Analysis of BLV Full-Length Gag Nucleotide Sequences

Although the BLV *gag* sequence has rarely been used for phylogenetic analyses, this was included here to add more insight into the genetic variabilities of the BLV isolates and to compare the results with the phylogenetic analyses of the BLV *env* nucleotide sequences. NJ (Supplementary Figure S5) and ML (Figure 4) trees based on the alignment of BLV full-length *gag* nucleotide sequences were constructed using 40 BLV isolates representing G1, G2, G4, G6, G9 and G10 from 11 countries.

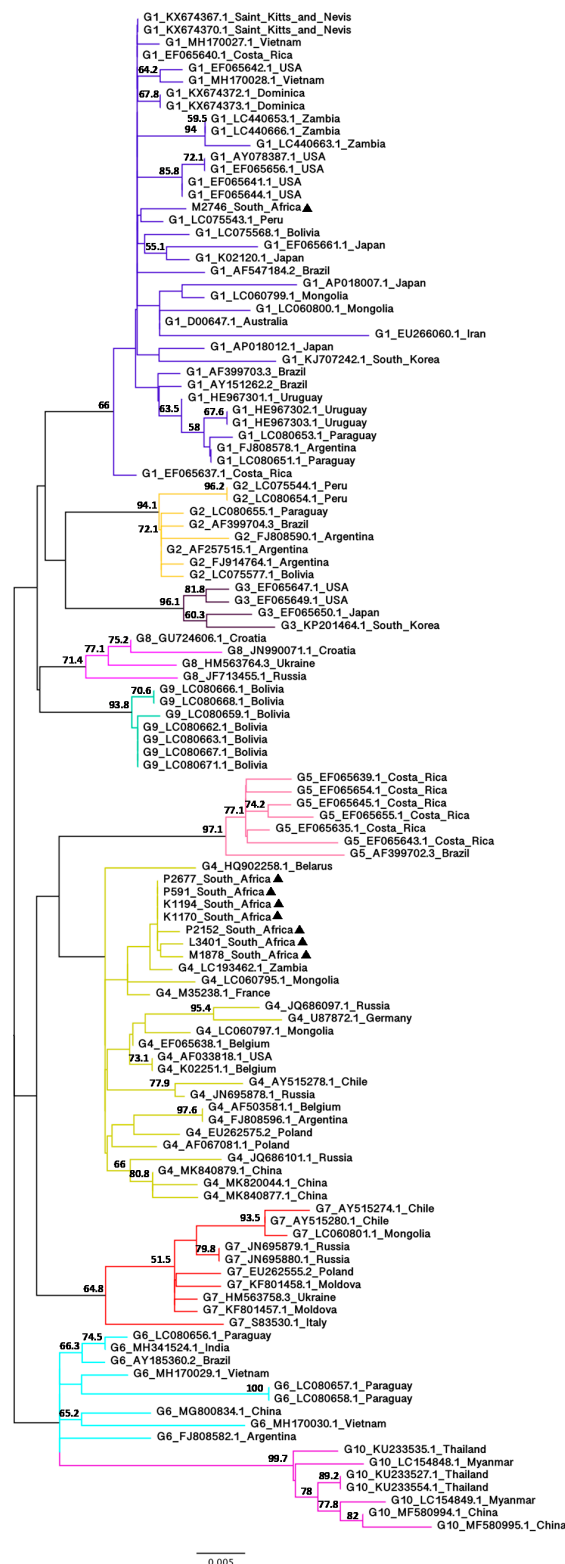


Figure 3. Maximum-likelihood phylogenetic tree based on the alignment of BLV partial *env* nucleotide sequences (444 bp) from South Africa and other geographic regions worldwide. The South African BLV strains identified in this study are indicated by filled triangles (▲) following their sample ID. Other strains are indicated by their genotype, GenBank accession number and country of origin. Numbers at the branches denote bootstrap support (1000 iterations). Bootstrap values of $\geq 50\%$ are shown. The bar at the bottom of the figure denotes genetic distance.

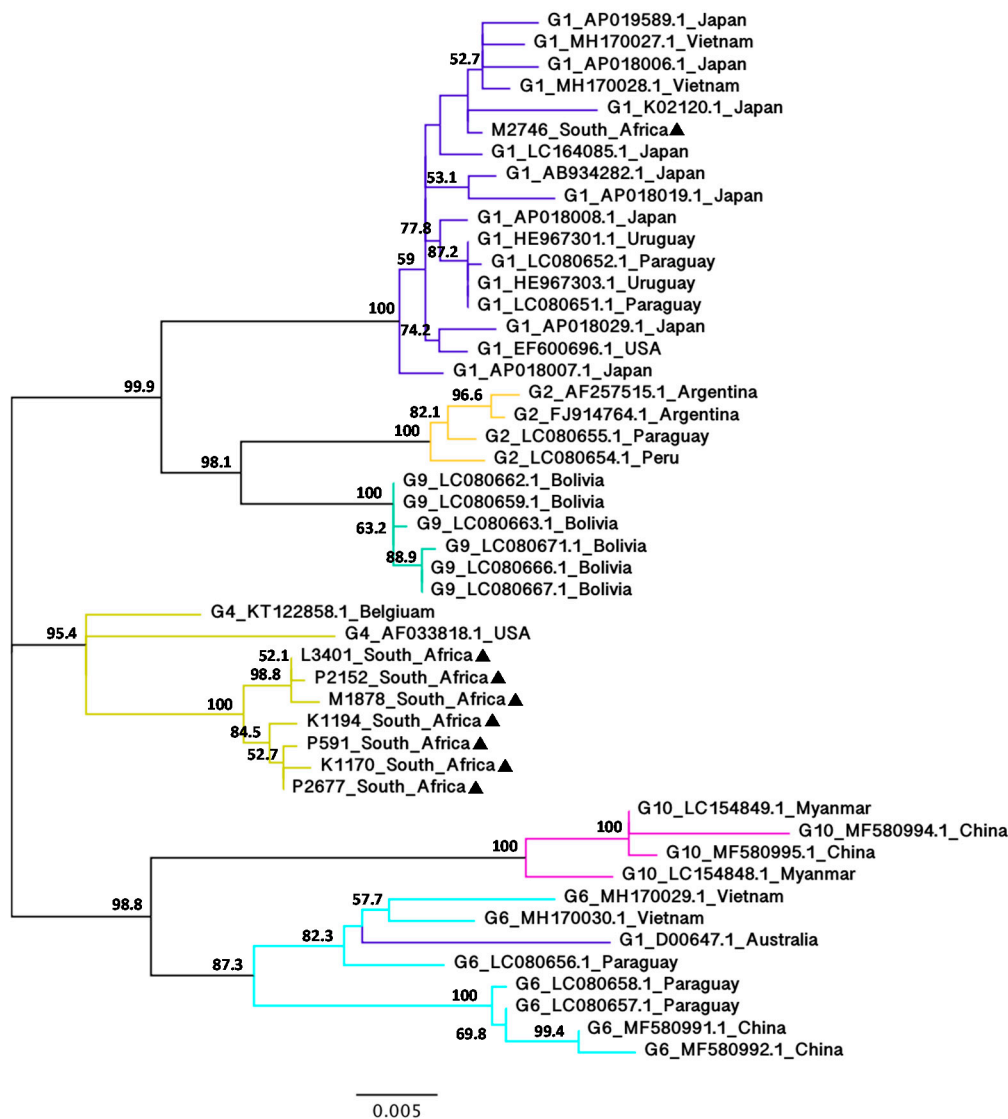


Figure 4. Maximum-likelihood phylogenetic tree based on the alignment of BLV full-length *gag* nucleotide sequences (1182 bp) from South Africa and other geographic regions worldwide. The South African BLV strains identified in this study are indicated by filled triangles (▲) following their sample ID. Other strains are indicated by their genotype, GenBank accession number and country of origin. Numbers at the branches denote bootstrap support (1000 iterations). Bootstrap values of $\geq 50\%$ are shown. The bar at the bottom of the figure denotes genetic distance.

Despite the limited number of full-length *gag* nucleotide sequences currently available on the GenBank sequence database, NJ and ML trees showed congruent topologies with the phylogenetic analyses of the BLV full-length *env* nucleotide sequences supported by high bootstrap values. Similar to the BLV *env* phylogenetic trees, the BLV *gag* trees showed that seven of the South African isolates belong to G4 with 90.4% (NJ) and 95.4% (ML) bootstrap values, whereas one South African isolate (M2746) belongs to G1 with 100% bootstrap value (NJ and ML). Interestingly, there appeared to be more genetic variability in the BLV *gag* nucleotide sequences amongst different genotypes as evident from the tree topology, as well as the pairwise evolutionary distance (Supplementary Figure S6).

3.4. Pairwise Comparison of the BLV Full-Length *Env* Sequences

Pairwise comparison of the BLV full-length *Env* nucleotide and amino acid sequences between the South African isolates and 46 selected isolates (Supplementary Figures S7 and S8) were performed.

Figure 5 shows pairwise percent identity of the BLV full-length Env sequences between the South African isolates and 17 selected G1 and G4 isolates. The analysis revealed that the G4 South African isolates shared significantly high sequence identity with each other: $\geq 99.74\%$ nucleotide sequence identity (0–4 differences in 1548 nucleotides) and $\geq 99.81\%$ amino acid sequence identity (0–2 differences in 515 amino acids) (Figure 5 and Supplementary Figure S8). In particular, the nucleotide sequences of the P2677 and P591 isolates are identical. Furthermore, this analysis also showed that despite some nucleotide sequence variabilities (99.74–99.94% identity or 1–4 differences in 1548 nucleotides), five South African isolates K1170, P591, P2152, P2677 and L3401 shared 100% amino acid sequence identity, indicating that their nucleotide substitutions were synonymous mutations.

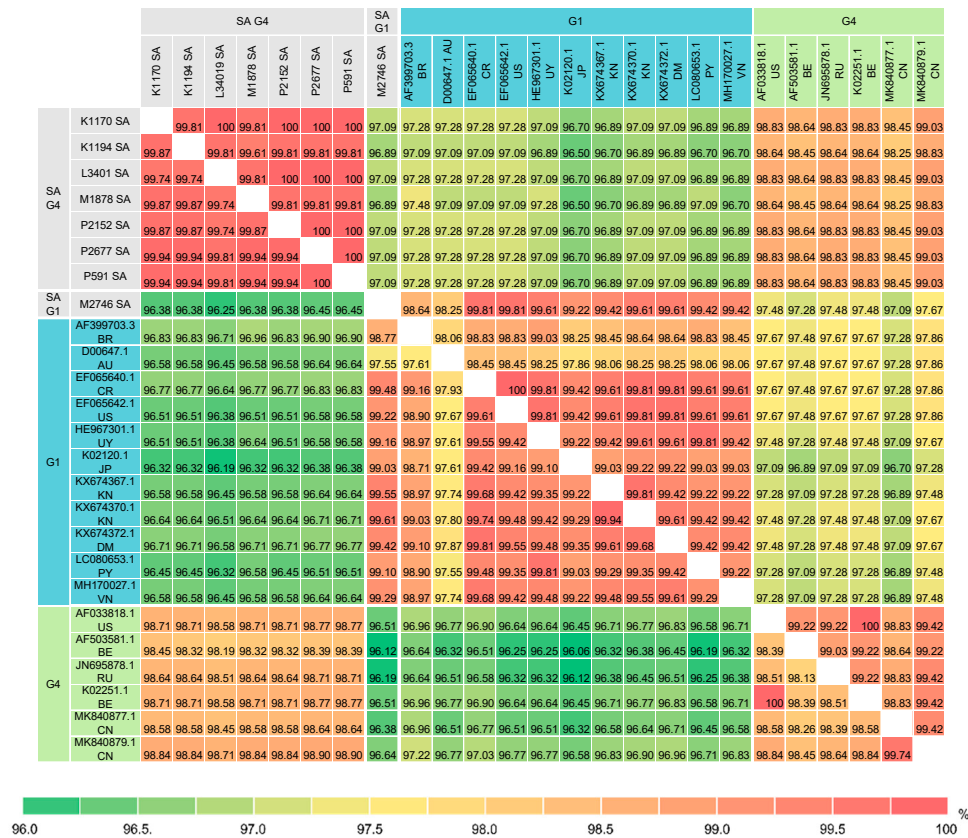


Figure 5. Pairwise percent identity of BLV full-length Env nucleotide (1548 bp) and amino acid (515 amino acids) sequences between G1 and G4 isolates from South Africa and other geographic regions worldwide. The lower matrix shows the percent identity of nucleotide sequences and the upper matrix shows that of amino acid sequences. The country of origin is indicated by 2-letter country codes (see Supplementary Table S2). SA, South Africa.

Consistent with the phylogenetic analyses based on the NJ and ML trees, the full-length Env nucleotide and amino acid sequences of the G1 South African isolate showed high sequence identity with the majority of the G1 isolates (Figure 5 and Supplementary Figure S8). The G1 South African full-length Env sequence shared 98.77–99.61% nucleotide sequence identity (6–19 differences in 1548 nucleotides) and 98.64–99.81% amino acid sequence identity (1–7 differences in 515 amino acids) with all G1 isolates except the Australian isolate (D00647.1). In contrast, there is slightly more sequence variability between the G4 South African isolates and other G4 isolates. The G4 South African isolates shared 98.19–98.90% nucleotide sequence identity (18–28 differences in 1548 nucleotides) and 98.25–99.03% amino acid sequence identity (5–9 differences in 515 amino acids) with all other G4 isolates.

3.5. Pairwise Comparison of BLV Partial Env Sequences

Pairwise comparison of the BLV partial Env nucleotide and amino acid sequences between the South African isolates and 20 selected G1 and G4 isolates (Supplementary Figures S9 and S10). Furthermore, in order to compare South African and Zambian isolates, all fourteen G1 Zambian sequences currently available on the GenBank sequence database were used in this analysis (Supplementary Tables S3 and S4). Supplementary Figure S11 shows alignment of the deduced amino acid sequences of the 148 amino acid region in the Env gp51 protein between the South African isolates and G1 and G4 isolates.

As opposed to the analyses based on the full-length Env sequences, sequences of the 444 bp region in the *gp51* gene appear to be slightly more variable as this region contains highly polymorphic epitope sequences and neutralising domains. The G1 South African sequence shared moderate sequence identity with the G1 isolates 97.30–99.55% nucleotide sequence identity (2–12 differences in 444 nucleotides) and 95.95–99.32% amino acid sequence identity (1–6 differences in 148 amino acids) (Supplementary Figures S9 and S10). The G4 South African isolates and G4 isolates showed 97.75–99.55% nucleotide sequence identity (2–10 differences in 444 nucleotides) and their amino acid sequences showed 95.95–99.32% (1–6 differences in 148 amino acids).

Comparison of the South African isolates with the Zambian isolates indicated that the G1 South African isolate is more similar to G1 isolates from other parts of the world than to the G1 Zambian isolates. The G1 South African nucleotide sequence is only moderately similar to the fourteen G1 Zambian isolates (Supplementary Table S3), showing 98.20–98.87% sequence identity or 5–8 differences in 444 nucleotides. However, their amino acid sequences shared relatively low sequence identity (96.62–98.65% identity or 2–5 differences in 148 amino acids). In contrast, the G4 South African isolates are more divergent from the other G4 isolates but highly similar to the G4 Zambian isolate. Three G4 South African isolates (M1878, P2677 and L3401) and four other G4 South African isolates (K1170, K1194, P591 and P2152) shared 99.55% (two nucleotide differences) and 99.77% nucleotide sequence identity (one nucleotide differences) with the Zambian isolate (LC193462), respectively (Supplementary Table S4). Furthermore, all G4 South African isolates except the M1878 isolate shared 100% amino acid sequence identity with the G4 Zambian isolate. The M1878 isolate shared 99.32% amino acid sequence identity with the Zambian isolate, which is only one amino acid difference between these isolates.

3.6. Pairwise Comparison of BLV Gag Sequences

Pairwise comparison of the BLV full-length Gag nucleotide and amino acid sequences between the South African isolates and 29 selected isolates from other geographic regions (Figure 6 and Supplementary Figures S12 and S13) showed that Gag sequences from different genotypes appeared to be more variable than Env sequences. Pairwise sequence identities of the full-length *env* nucleotide sequences were $\geq 94.44\%$ (Figure 5 and Supplementary Figure S7) whereas those of the full-length *gag* nucleotide sequences were $\geq 92.58\%$ (Figure 6 and Supplementary Figure S12).

Similarly, the Gag sequences amongst the G4 South African isolates appeared to be slightly more variable than their Env sequences. Whilst the full-length Env sequences of the G4 South African isolates shared high nucleotide and amino acid sequence identity ($\geq 99.74\%$ and $\geq 99.61\%$, respectively) with each other and Env amino acid sequences of the five G4 South African isolates are identical, their Gag sequences shared $\geq 99.24\%$ nucleotide sequence identity (1–9 differences in 1182 nucleotides) and $\geq 98.98\%$ amino acid sequence identity (0–4 differences in 393 amino acid) with each other (Supplementary Figures S12 and S13). The K1170 and P2677 isolates shared 100% Gag amino acid sequence identity, as did P2152 and L3401 isolates.



Figure 6. Pairwise percent identity of BLV full-length Gag nucleotide (1182 bp) and amino acid (393 amino acids) sequences between G1 and G4 isolates from South Africa and other geographic regions worldwide. The lower matrix shows the percent identity of nucleotide sequences and the upper matrix shows that of amino acid sequences. The country of origin is indicated by 2-letter country codes (see Supplementary Table S2). SA, South Africa.

Only two G4 *gag* sequences have been published and these are from Belgium and America. Consistent with the phylogenetic analyses, this analysis showed that the G4 South African isolates are more divergent from the G4 Belgian (KT122858.1) and American (AF033818) isolates. The G4 South African isolates showed only 97.29–98.14% nucleotide identity and 97.71–98.47% amino acid identity with these two G4 isolates, which corresponds to 22–32 nucleotide differences and 6–9 amino acid differences.

In contrast to the G4 isolates, the G1 South African isolate shared relatively high sequence identity, 99.24–99.66% nucleotide sequence identity (4–9 differences in 1182 nucleotides) and 99.24–99.75% amino acid sequence identity (1–4 differences in 393 amino acids), with all G1 isolates except an Australian isolate (D000647) and a Japanese isolate (K02120.1). It should be noted that the Australian and Japanese isolates have a three-nucleotide deletion, equivalent to one amino acid deletion (which appears to result in an in-frame deletion) and they are more divergent from other G1 isolates.

3.7. Genetic Variabilities of BLV Env Sequences

Using 203 full-length Env sequences with nine known genotypes obtained from the GenBank sequence database and the eight South African Env sequences, mean intragenotype and intergenotype nucleotide and amino acid distances were estimated (Supplementary Table S5). The analysis confirmed that the seven G4 South African isolates and one G1 South African isolate belong to G4 and G1, respectively, showing the lowest mean intergenotype distances. The mean nucleotide and amino acid distances between the G1 South Africa isolate and the G1 isolates were 0.0082 and 0.0063, respectively and those between the G4 South African isolates and the G4 isolates were 0.0138 and 0.0141, respectively. Furthermore, comparison of these mean intergenotype evolutionary distances indicated that the G1 South African isolate is less divergent from other G1 isolates whereas the G4 South African isolates are slightly more divergent from other G4 isolates, which also confirmed the results of the phylogenetic analyses based on the full-length *env* nucleotide sequences.

This analysis was also consistent with the results obtained from pairwise comparison of the full-length Env sequences, showing that the Env sequences of the G4 South African isolates themselves are significantly conserved. The mean intragenotype distances of the G4 South African Env sequences were 0.0133 and 0.0011 for nucleotide and amino acid sequences, respectively.

Furthermore, this analysis revealed that the Env sequences of G6 and G10 isolates are more variable than other genotypes. The G6 and G10 Env sequences showed mean intragenotype nucleotide distances of 0.0267 and 0.0273, respectively, and mean intragenotype amino acid distances of 0.0249 and 0.0320, respectively.

3.8. Alignment of BLV Env Amino Acid Sequences

To assess whether the nucleotide substitutions altered amino acid sequences in the South African isolates and to identify the genotype-specific amino acid substitutions previously described [41,44,45,73] as well as amino acid substitutions unique to the South African isolates, the full-length Env and Gag deduced amino acid sequences of the South African isolates were aligned with a total of the 851 Env amino sequences and 113 Gag amino acid sequences available from the GenBank sequence database. Table 1 shows the alignment of Env amino acid sequences of the South African isolates with those of the 43 selected isolates representing eleven genotypes.

Table 1. Amino acid substitutions in the BLV Env protein of the South African sequences and 43 selected sequences representing eleven genotypes.

Epitopes and Motifs	gp51											gp30						
	Leader Sequence	G Epitope	H Epitope	G Epitope				ND2			D Epitope		PXXP Motif					
	Amino acid position	29	48	56	59	73	74	82	121	132	134	144	153	254	476	479	480	504
References	d	a,d	a		a,b	a,d	a,d	a,c	a	a,d	a,d		a,d		d	d	d	
SA G1	M2746	R	A	S	I	A	K	S	R	Q	D	I	Q	S	E	L	T	V
G1	AF399703.3 BR	Q	A	S	I	A	K	F	R	Q	N	I	H	S	E	S	T	V
	AP018021.1 JP	R	A	S	I	A	K	S	R	Q	D	I	H	S	E	L	T	V
	EF065641.1 US	R	A	S	I	A	K	S	R	Q	D	I	H	S	E	L	T	V
	EF065646.1 JP	Q	A	S	I	A	K	S	R	Q	D	I	H	S	E	L	T	V
	HE967301.1 UY	R	A	S	I	A	K	S	R	Q	N	I	H	S	E	L	T	V
	KX674639.1 KN	R	A	S	I	A	K	S	R	Q	D	I	H	S	E	L	T	V
	LC440653.1 ZM	-	-	-	-	-	-	-	R	Q	D	I	H	-	-	-	-	-
	LC075558.1 PY	-	-	-	-	-	-	-	R	Q	D	I	H	-	-	-	-	-
G2	AF257515.1 AR	Q	A	S	I	A	K	F	R	Q	D	I	H	L	E	L	A	T
	LC080654.1 PE	Q	V	S	I	A	K	F	R	Q	D	I	H	L	E	F	A	T
	LC080655.1 PY	Q	A	S	I	A	K	F	R	Q	D	I	H	L	E	F	A	T
G3	EF065647.1 US	Q	A	S	I	A	K	F	R	Q	D	I	H	L	E	F	A	T
	EF065650.1 JP	Q	A	S	I	A	K	F	R	Q	D	I	H	L	E	F	A	T
	JQ686091.1 RU	Q	T	S	I	P	R	F	H	Q	D	I	H	L	E	F	P	T
SA G4	K1170	Q	T	F	I	P	R	F	H	R	D	T	H	L	D	F	P	T
	K1194	Q	T	F	L	P	R	F	H	R	D	T	H	L	D	F	P	T
	L3401	Q	T	F	I	A	R	F	H	R	D	T	H	L	D	F	P	T
	M1878	Q	T	F	I	P	R	F	H	R	N	T	H	L	D	F	P	T
	P2152	Q	T	F	I	P	R	F	H	R	D	T	H	L	D	F	P	T
	P2677	Q	T	F	I	P	R	F	H	R	D	T	H	L	D	F	P	T
	P591	Q	T	F	I	P	R	F	H	R	D	T	H	L	D	F	P	T
	AF033818.1 US	Q	T	S	I	P	R	F	H	Q	D	I	H	L	E	F	P	T
G4	AF067081.1 PL	-	-	-	-	-	-	-	H	Q	D	I	H	-	-	-	-	
	AY515279.1 CL	Q	T	S	I	P	R	F	H	Q	D	I	H	L	E	F	P	S
	JN695878.1 RU	Q	T	S	I	P	R	F	H	Q	D	I	H	L	E	F	P	T
	KF801460.2 MD	Q	T	S	I	P	R	F	H	Q	D	I	H	L	E	F	P	T
	LC193462.1 ZM	-	-	-	-	-	-	-	H	R	D	T	H	-	-	-	-	
	M35238.1 FR	Q	T	F	I	P	R	F	H	Q	D	T	H	L	E	F	P	T
	M35240.1 BE	Q	T	S	I	P	R	F	H	Q	D	I	H	L	E	F	P	T
	MK820044.1 CN	Q	T	S	I	P	R	F	H	Q	D	I	H	L	E	F	P	T

Table 1. Cont.

Epitopes and Motifs	gp51											gp30					
	Leader Sequence	G Epitope	H Epitope	G Epitope				ND2			D Epitope			PXXP Motif			
Amino acid position	29	48	56	59	73	74	82	121	132	134	144	153	254	476	479	480	504
References	d	a,d	a		a,b	a,d	a,d	a,c	a	a,d	a,d		a,d		d	d	d
G5	AF399702.3 BR	Q	T	S	I	A	R	F	R	R	D	I	H	L	-	-	-
	EF065635.1 CR	R	T	S	I	A	R	F	R	R	D	I	H	L	E	F	T
	EF065636.1 CR	R	T	S	I	A	R	F	R	R	D	I	H	L	E	L	T
G6	AY185360.2 BR	Q	T	S	I	A	R	F	R	Q	D	T	H	L	E	F	T
	LC075576.1 BO	Q	T	S	I	A	R	F	R	Q	D	T	H	L	E	F	T
	LC080656.1 PY	Q	T	S	I	A	R	F	R	Q	D	T	H	L	E	F	T
	MF580991.1 CN	Q	T	S	I	A	R	F	R	Q	D	T	Y	L	E	F	T
	KU233530.1 TH	-	-	-	-	-	-	-	R	Q	D	T	H	-	-	-	-
G7	KF801457.1.MD	Q	I	S	I	A	R	F	R	Q	D	I	H	L	E	F	T
	JN695879.1 RU	Q	T	S	I	A	R	F	R	Q	D	I	H	L	E	F	T
	S83530.1 IT	-	-	-	-	-	-	-	R	Q	D	I	H	-	-	-	-
G8	GU724606.1 HR	-	-	-	-	-	-	-	R	Q	D	I	H	-	-	-	-
	JF713455.1 RU	-	-	-	-	-	-	-	R	Q	D	I	H	-	-	-	-
	LT970927.1 IT	-	-	-	-	-	-	-	R	Q	D	I	H	-	-	-	-
G9	LC080659.1 BO	Q	A	S	I	A	K	L	R	Q	D	I	H	L	E	F	A
	LC080664.1 BO	Q	A	S	I	A	K	F	R	Q	D	I	H	L	E	F	A
G10	KU233527.1 TH	-	-	-	-	-	-	-	H	Q	D	T	H	-	-	-	-
	LC154848.1 MM	Q	T	S	I	A	R	F	R	Q	D	T	H	L	E	F	T
	MF580994.1 CN	Q	T	S	I	A	R	F	H	Q	D	T	H	L	E	F	T
G11	KU764746.1 CN	-	-	-	-	-	-	-	R	Q	D	I	H	-	-	-	-

a, amino acid substitutions described by Rodriguez et al. (2009) [41]; b, genotype-specific amino acid substitution described by Zhao and Buehring (2007) [73]; c, genotype-specific amino acid substitution previously described by Lee et al. (2016) [45]; d, genotype-specific amino acid substitutions described by Polat et al. (2016) [44]. SU, surface protein; TM, transmembrane protein; ND 2, neutralising domain 2; PXXP motif, proline-rich motif for the cell signalling; Dash (-) indicates amino acid residues are not available due to the partial sequences. Amino acid substitutions that were exclusively found in the South African isolates are shown in red font. Each amino acid has been assigned a unique background color. The country of origin is indicated by 2-letter codes (see Supplementary Table S2).

Alignment of the BLV Env amino acid sequences confirmed the presence of the G1- and G4-specific amino acid substitutions in the South African isolates at fourteen residues (at positions 29, 48, 56, 73, 74, 82, 121, 132, 134, 144, 254, 479, 480 and 504) in their Env sequences. As some exceptions to these genotype-specific substitutions, a proline-to-alanine substitution at position 73 (P73A) in the L3401 isolate and a D134N substitution in the M1878 isolate were found. Both of these amino acid substitutions appeared to be G1-like and could be random mutations.

Furthermore, three unique amino acid substitutions in the Env sequences (at positions 59, 153 and 476) were found in some South African sequences. An isoleucine-to-leucine substitution at position 59 (I59L) and a histidine-to-glutamine substitution at position 153 (H153Q) were only present in the K1194 and M2746 isolates, respectively and thus, this could be a random mutation. The E476D substitution was consistently found in the seven G4 South African isolates. This substitution was also found in G1 Mexican sequences (KY548770.1, KY548786.1 and KY548787.1; data not shown) and a G9 Bolivian sequence (LC080675.1; data not shown). However, none of the (non-South African) G4 isolates had this substitution, suggesting that this amino acid substitution is unique to the G4 South African isolates.

3.9. Alignment of BLV Gag Amino Acid Sequences

Alignment of the Gag amino acid sequences of the South African isolates with 113 Gag amino acid sequences available on the GenBank sequence data also confirmed the presence of G1- and G4-specific amino acid substitutions in the South African isolates. Table 2 shows the alignment of Gag amino acid sequences of the South African isolates with those of the 21 selected isolates representing six genotypes. The analysis showed that the South African isolates contained genotype specific amino acid substitutions at four residues (at positions 63, 69, 88 and 365) in their Gag sequences. Furthermore, seven amino acid substitutions (positions 29, 107, 108, 278, 318, 341 and 343) in the Gag sequences that had not been previously described were detected in the South African isolates and these were mainly found in the G4 South African isolates.

Table 2. Amino acid substitutions in the BLV Gag protein of the South African sequences and 21 selected sequences representing six genotypes.

Amino Acid Position	p15 MA				p24 CA				p12 NC		
	29	63	69	88	107	108	278	318	341	343	365
References		a	a	a							a
SA G1 M2746	N	T	K	G	A	V	I	M	K	P	A
AP018021.1 JP	N	T	R	G	A	V	I	M	K	P	A
D00647.1 AU	N	A	K	E	A	V	I	T	K	P	T
AB934282.1 JP	N	T	R	G	A	I	I	M	K	P	A
LC005615.1 JP	N	T	R	G	A	V	I	M	K	P	A
G1 LC080651.1 PY	N	T	R	G	A	V	I	M	K	P	A
MH170027.1 VN	N	T	R	G	A	V	I	M	K	P	A
MH170028.1 VN	N	T	R	G	A	V	I	M	K	P	A
EF600696.1 US	N	T	R	G	A	V	I	V	K	P	A
HE967301.1 UY	N	T	R	G	A	V	I	M	K	P	A
G2 FJ914764.1 AR	N	T	R	G	A	V	I	I	K	P	T
LC080654.1 PE	N	T	R	G	A	V	I	I	K	P	T
LC080655.1 PY	N	T	R	G	A	V	I	I	K	P	T
SA G4 K1170	D	A	K	E	A	V	V	I	K	S	T
K1194	D	A	K	E	V	V	V	I	K	S	T
L3401	D	A	K	G	S	V	V	I	K	S	T
M1878	D	A	R	G	S	I	V	I	K	S	T
P2152	D	A	K	G	S	V	V	I	K	S	T
P2677	D	A	K	E	A	V	V	I	K	S	T
P591	D	A	K	E	A	V	V	I	Q	S	T

Table 2. Cont.

Amino Acid Position	p15 MA						p24 CA		p12 NC			
	29	63	69	88	107	108	278	318	341	343	365	
References		a	a	a							a	
G4	KT122858.1 BE	N	A	K	E	A	V	I	I	K	P	T
	AF033818.1 US	N	A	K	E	A	I	I	V	K	P	T
G6	MF580991.1 CN	N	V	K	E	V	V	I	T	K	P	T
	LC080656.1 PY	N	V	K	E	A	V	I	T	K	P	T
	MH170030.1 VN	N	V	K	E	A	I	I	T	K	P	T
G9	LC080659.1 BO	N	T	R	G	A	V	I	I	K	P	T
	LC080664.1 BO	N	T	R	G	A	V	I	I	K	P	T
G10	MF580994.1 CN	N	V	K	E	A	D	I	V	K	P	A
	LC154848.1 MM	N	V	K	E	A	V	I	V	K	P	A

a, genotype-specific amino acid substitutions previously described by Polat et al. (2016) [56]. Amino acid substitutions that are previously not described are highlighted in blue. G1 and G4 sequences are highlighted in yellow for comparison. MA, matrix; CA, capsid; NC, nucleocapsid. The country of origin is indicated by 2-letter codes Dash (-) indicates amino acid residues are not available due to the partial sequences. Amino acid substitutions that were exclusively found in the South African isolates are shown in red font. Each amino acid has been assigned a unique background color. The country of origin is indicated by 2-letter codes (see Supplementary Table S2).

Four of these seven amino acid substitutions were particularly unique to the South African sequences. The amino acid alignment showed that N29D, I278V and P343S substitutions were only found in the G4 South African sequences, though the I278V substitution was also detected in an Iranian isolate (LC193727.1; data not shown), which has not been genotyped. Although previously not described, our analysis indicates that the presence of methionine at position 318 in the M2746 sequence and isoleucine at the same position in the seven G4 South African sequences could be genotype-specific.

A K341Q substitution was only found in the South African P591 isolate amongst 113 isolates studied. In addition, a G1-like K69R substitution was found in the G4 South African M1878 isolate and a V108I substitution was only found in the M1878 isolate but not in other six G4 South African isolates. It is uncertain whether these amino acid substitutions are spontaneous mutations.

A genotype-specific glycine-to-glutamic acid substitution at position 88 (G88E) was previously described [44] and the presence of the glycine residue at this position was identified as G1-specific and the presence of the glutamic acid residue as G4-specific. However, in our study, the G4 South African isolates M1878, P2152 and L3401 had the glycine residue at this position. The A107V substitution in the K1194 isolate and an A107S substitution in the M1878, P2152 and L3401 isolates could be random mutations as these mutations did not appear to be genotype-specific.

4. Discussion

This study is the first known phylogenetic analysis of BLV isolates from South Africa. Full-length *env* and *gag* genes from eight South African BLV isolates were sequenced. Phylogenetic analyses based on these genes revealed that at least two genotypes, genotype 1 and genotype 4, are present in South Africa, which is consistent with the Zambian studies [65,66]. Phylogenetic analyses and pairwise comparison of Gag sequences further provided two findings. Firstly, the Env and Gag sequences of the G4 South African isolates are highly conserved and some of these G4 South African isolates encode identical amino acid sequences. Secondly, Env sequences are highly conserved within as well as between genotypes whereas Gag sequences are more variable between genotypes. Lastly, alignment of Env and Gag amino acid sequences identified genotype-specific as well as novel amino acid substitutions in the South African isolates.

In 2016, Pandey et al. identified a G4 isolate from a dairy cow in Zambia by phylogenetic analysis based on the 444 bp region of the *gp51* gene [65]. In 2018, using the same 444 bp region, they identified fourteen G1 isolates from beef cattle [66]. These Zambian studies enabled comparison of the South African and Zambian BLV partial Env sequences. This comparison revealed that the G1

South African isolate is more closely related to the majority of the G1 isolates (not from Zambia) than to the G1 Zambian isolates. This result was not surprising as Phiri et al. found in their 2018 study that the G1 Zambian isolates form a distinct group within G1 [66]. In contrast, the G4 South African isolates appeared to be highly similar to the G4 Zambian isolate. The G4 South African isolates differed from the G4 Zambian isolate only at one or two nucleotide positions; the amino acid sequence identity was 100% for six of the seven G4 South African isolates (K1170, K1194, P591, P2152, P2677 and L3401) and 99.32% for the M1878 isolate. However, we cannot entirely exclude the possibility that there are G1 South African isolates that are closely related to the G1 Zambian isolate as this was a very small study with samples only taken from a single herd. Similarly, only one G4 isolate was sequenced from Zambia and thus, there might be G4 Zambian isolates that are less similar to South African isolates. Furthermore, if future studies could sequence full-length *env* (1548 bp) and *gag* (1182 bp) genes from Zambian isolates as well as South African isolates, it would enable a more comprehensive analysis of the G1 and G4 isolates circulating in these two countries, as these comparisons are only based on a 444 bp region of *env*.

It should be noted that the collection dates are unknown for the majority of the available BLV sequences. Therefore, we could not perform the analysis and interpretation of the BLV evolutionary history for eleven genotypes as well as the G4 South African isolates.

Although *gag* sequences are rarely used for phylogenetic analysis and the number of *gag* sequences available on the GenBank sequence database is limited, we sequenced the full-length *gag* gene to compare the results obtained from phylogenetic analyses as well as sequence comparison and alignment of full-length and partial *env* sequences. Phylogenetic analyses based on the full-length *gag* nucleotide sequences were in agreement with the phylogenetic analyses based on the full-length and partial *env* nucleotide sequences, validating that one South African isolate belongs to G1 and the seven other South African isolates belong to G4. Consistent with the analyses based on the full-length Env sequences, Gag sequences between the G1 South African isolate and G1 isolates from other parts of the world appeared to be more conserved than those between the G4 South African isolates and G4 Belgian and American isolates.

Furthermore, full-length Gag sequences are slightly more variable between different genotypes than full-length Env sequences as evident from the phylogenetic analysis and pairwise sequence comparison. Similarly, whilst five out of the seven G4 South African Env amino acid sequences are identical, only two of their Gag sequences are identical. Since only 113 Gag sequences with known genotypes are available from a limited number of countries, obtaining more Gag sequences is desirable to validate these conclusions and will provide a more comprehensive understanding of evolutionary changes in the Gag sequences.

Amino acid sequence alignment between the South African isolates and global isolates representing all eleven genotypes detected seventeen and eleven amino acid substitutions in the South African Env and Gag sequences, respectively. These substitutions include genotype-specific substitutions as well as those unique to the South African isolates. The amino acid substitution at position E476D in the Env sequences as well as the amino acid substitutions at positions N29D, I278V and P343S in the Gag sequences were consistently detected in all G4 South African isolates. These substitutions are absent in other G4 isolates from other countries as well as any isolates from other genotypes. Importantly, these amino acid substitutions which are unique to the G4 South African isolates appear to be fairly conservative and thus, they are unlikely to cause functional changes to the proteins. No insertions and deletions were detected in the South African sequences.

The genotype-specific amino acid substitutions in the Env sequences detected in this study include those at position 29 in the signal sequence, positions 48, 73, 74 and 82 in the G epitope, position 56 in the H epitope, positions 134 and 144 in the second neutralising domain (ND2), position 254 in the D epitope and position 504 in the PXXP motif. The Env mutations in the South African isolates appeared to be concentrated in the conformational epitopes (F, G and H) and the ND2. These results are consistent with previous studies showing that the G epitopes, particularly residues 48, 74 and 82, are under

positive selection and are the most polymorphic sites [51,73–75]. Homology modelling of the BLV Env protein has suggested that the conformational epitopes and the neutralising domains are located on the surface of the virus [14,73]. Therefore, it is not surprising that these epitopes and domains are targets of neutralising antibodies (NAbs) [16,18,19,24] and under strong selection pressure [73], leading to the antigenic variations in these regions [19].

There are three types of conformational epitopes (F, G and H) [17,76,77] on the Env gp51 surface protein which are important for cell fusion and syncytium formation during viral dissemination [16]. In our study, four point mutations in the G epitopes (positions 48, 73, 74 and 82) and one point mutation in the H epitope (position 56) were detected from the South African isolates. Since functional studies on the Env protein and Env mutants are scarce, functional consequences of the Env mutations found in our studies are unknown. However, non-conservative mutations in the conformational epitopes and the neutralising domain 2 (ND2) would affect the recognition of NAbs to these epitopes. A previous study which identified the S56F substitution in the H epitope showed that this substitution altered the recognition of a monoclonal antibody to this epitope owing to the drastic change from the small serine residue to the large hydrophobic phenylalanine residue [78]. This effect would also apply to the S82F substitution in the G epitope. Mutations in the conformational epitopes appear to benefit viral persistence through antigenic changes and immune escape. Previous studies demonstrated that BLV natural variants lacking one or two of the conformational epitopes were viable and appear to have evolved these mutations to evade host immune responses by altered antibody recognition [19,78]. However, BLV mutants lacking all conformational epitopes have not been detected and these deletion mutants retained at least one of the three conformational epitopes [19,78]. These observations indicate that the simultaneous loss of these epitopes would be deleterious and possibly cause loss of infectivity to the virus [16].

The BLV Env protein contains three neutralising domains (ND1 to ND3) [24] and these domains also play a role in viral infectivity (cell fusion and syncytium) as well as immunogenicity [15,24,79]. Previous studies demonstrated that the ND2 is particularly prone to multiple point mutations [41,46] but it was also shown that ND2 is under purifying selection [73,75], reflecting its functional significance. Rodríguez et al. [41] identified eleven point mutations within the ND2 from BLV strains from a variety of geographic regions. In contrast, the ND1 and ND3 appear to be less prone to mutations [41,80]. Using homology modelling, Moratorio et al. [79] demonstrated that the D134N substitution in the ND2 changed net charge in a loop of the Env protein, possibly disturbing the immunogenicity and/or fusogenic properties of the ND2. One G4 South African isolate (M1878) had the G1-like D134N substitution in this ND2. It is unclear whether this is a spontaneous mutation.

Eleven amino acid substitutions were found in the South African Gag sequences and the majority of the substitutions (six out of eleven substitutions) occurred in the p15 matrix (MA) protein (positions 26, 63, 69, 88, 107 and 108) whereas only two amino acid substitutions were found in the p24 capsid (CA) protein (positions 278 and 318). Since the CA protein is the second major target for the NAbs next to the gp51 surface protein [32], this protein could be under selection pressure. Yet, it is possible that mutations in the CA could render a fitness cost to the virus and cause deleterious effects to viral replication. A genetic footprinting analysis on the Moloney murine leukaemia virus demonstrated that the majority of the MA regions were tolerant of insertions whereas the N-terminal region of the CA was susceptible to insertions [81]. They further demonstrated that although Gag mutants harbouring mutations in the MA and N-terminal region of the CA all produced virions, CA mutants produced an abundance of immature virions [81]. Furthermore, whilst the MA mutants showed reverse transcriptase (RT) and nuclear transport activities and appeared infectious, the RT and nuclear transport activities were not detected in the CA mutants, suggesting the mutations affect uncoating or viral entry [81].

Three amino acid substitutions were detected in the p12 nucleocapsid (NC) protein (positions 341, 343 and 365). The retroviral NC protein contains conserved zinc-finger motifs [82–85] and basic residues [86–88] for RNA packaging and virion assembly [89,90]. The K341Q substitution in the NC

protein was only detected from the South African P591 isolate, not from any other South African or global isolates. This mutation might be a spontaneous mutation. However, this amino acid change may cause functional changes in the virion assembly. Wang et al. [91] performed alanine-scanning mutagenesis study for the basic residues in the BLV NC and demonstrated that although the K341A substitution in the NC did not affect RNA packaging, the K341A mutant had two- to three-fold reduction in virion production. Although lysine can be substituted with other polar amino acids including glutamine, it is unknown whether the loss of charge may also affect virion production.

This study served as the first phylogenetic analysis of the BLV full-length *env* and *gag* sequences from eight South African isolates. Since the sample population was small, and was only collected from a single herd, and there was only one G1 isolate in our study, future studies will be required to obtain more samples from different herds and geographic regions to investigate whether other BLV genotypes are circulating in cattle in South Africa. Needless to say, if more BLV sequences from Africa become available, it would provide greater insight into BLV genotypes currently circulating in Africa. Lastly, the detection of two genotypes (G1 and G4) in southern Africa highlights the urgent need for disease management and the development of an efficacious vaccine against local strains.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1999-4915/12/8/898/s1>, Figure S1: Neighbour-joining phylogenetic tree based on the alignment of BLV full-length *env* nucleotide sequences (1548 bp) from South Africa and other geographic regions worldwide. Figure S2: Pairwise evolutionary distance of BLV full-length *env* nucleotide sequences (1548 bp) from the South African isolates and other G1 and G4 isolates from other geographic regions worldwide. Figure S3: Neighbour-joining phylogenetic tree based on the alignment of BLV partial *env* nucleotide sequences (444 bp) from South Africa and other geographic regions worldwide. Figure S4: Pairwise evolutionary distance of BLV partial *env* nucleotide sequences (444 bp) from the South African isolates and other G1 and G4 isolates from other geographic regions worldwide. Figure S5: Neighbour-joining phylogenetic tree based on the alignment of BLV full-length *gag* nucleotide sequences (1182 bp) from South Africa and other geographic regions worldwide. Figure S6: Pairwise evolutionary distance of BLV full-length *gag* nucleotide sequences (1182 bp) from the South African isolates and 29 isolates from other geographic regions worldwide. Figure S7: Pairwise percent identity of BLV full-length Env nucleotide (1548 bp) and amino acid sequences (515 amino acids) between the South African isolates and 46 isolates from other geographic regions worldwide. Figure S8: Pairwise comparison of BLV full-length Env nucleotide (1548 bp) and amino acid (515 amino acids) differences between the South Africa isolates and 46 isolates from other geographic regions worldwide. Figure S9: Pairwise percent identity of the BLV partial Env nucleotide (444 bp) and amino acid (148 amino acids) sequences between the South African isolates and G1 and G4 isolates from other geographic regions worldwide. Figure S10: Pairwise comparison of BLV partial Env nucleotide (444 bp) and amino acid (148 amino acids) differences between the South Africa isolates and G1 and G4 isolates from other geographic regions worldwide. Figure S11: Comparison of 148 amino acid region in the Env gp51 protein between the G1 and G4 South African and Zambian isolates. Figure S12: Pairwise percent identity of BLV full-length Gag nucleotide (1182 bp) and amino acid (393 amino acids) sequences between the South African isolates and 29 isolates from other geographic regions worldwide. Figure S13: Pairwise comparison of BLV full-length Gag nucleotide (1182 bp) and amino acid (393 amino acids) differences between the South Africa isolates and 29 isolates from other geographic regions worldwide. Table S1: GenBank accession numbers, sample IDs and genotypes of the BLV full-length *env* and *gag* nucleotide sequences of the South African isolates used in this study. Table S2: GenBank accession numbers, the country of origin and genotypes of the BLV full-length *env*, partial *env* and full length *gag* nucleotide sequences used in this study. Table S3: Pairwise comparison of BLV partial Env nucleotide and amino acid sequences between G1 South African and Zambian sequences. Table S4: Pairwise comparison of BLV partial Env nucleotide and amino acid sequences between G4 South African and Zambian sequences. Table S5: Mean nucleotide and amino acid distances in the BLV full-length Env sequences within (intra-genotype) and between (inter-genotype) BLV strains from South Africa and other geographic regions worldwide.

Author Contributions: Conceptualization, A.-L.W. and R.C.; sample collection, J.v.R.; methodology, A.S. A.-L.W., R.C., N.D. and O.C.; validation, A.S., R.C. and O.C.; writing—original draft preparation, A.S.; writing—review and editing, A.-L.W., R.C., N.D. and O.C.; funding acquisition, A.-L.W. and R.C. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by South African Research Chairs Initiative (SARChI) of the Department of Science and Technology and National Research Foundation (Grant: 64815).

Acknowledgments: We would like to thank the SARChI for the funding support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Onuma, M.; Honma, T.; Mikami, T.; Ichijo, S.; Konishi, T. Studies on the sporadic and enzootic forms of bovine leukosis. *J. Comp. Pathol.* **1979**, *89*, 159–167. [[CrossRef](#)]
2. Theilen, G.H.; Dungworth, D.L. Bovine lymphosarcoma in California. 3. Calf form. *Am. J. Vet. Res.* **1965**, *26*, 696–709. [[PubMed](#)]
3. Van Der Maaten, M.J.; Miller, J.M.; Boothe, A.D. Replicating type-C virus particles in monolayer cell cultures of tissues from cattle with lymphosarcoma. *J. Natl. Cancer Inst.* **1974**, *52*, 491–497. [[CrossRef](#)] [[PubMed](#)]
4. Burny, A.; Bruck, C.; Chantrenne, H.; Cleuter, Y.; Dekegel, D.; Ghysdael, J.; Kettmann, R.; Leclercq, M.; Leunen, J.; Mammerickx, M. Bovine leukemia virus: Molecular biology and epidemiology. In *Viral Oncology*; Klein, G., Ed.; Raven Press: New York, NY, USA, 1980.
5. DiGiacomo, R.F.; Darlington, R.L.; Evermann, J.F. Natural transmission of bovine leukemia virus in dairy calves by dehorning. *Can. J. Comp. Med. Rev. Can. De Med. Comp.* **1985**, *49*, 340–342.
6. Hopkins, S.G.; DiGiacomo, R.F. Natural Transmission of Bovine Leukemia Virus in Dairy and Beef Cattle. *Vet. Clin. N. Am. Food Anim. Pract.* **1997**, *13*, 107–128. [[CrossRef](#)]
7. Lucas, M.H.; Roberts, D.H.; Wibberley, G. Ear tattooing as a method of spread of bovine leukosis virus infection. *Br. Vet. J.* **1985**, *141*, 647–649. [[CrossRef](#)]
8. Kohara, J.; Konnai, S.; Onuma, M. Experimental transmission of Bovine leukemia virus in cattle via rectal palpation. *Jpn. J. Vet. Sci.* **2006**, *54*, 25–30.
9. Divers, T.J.; Bartholomew, R.C.; Galligan, D.; Littel, C. Evidence for transmission of bovine leukemia virus by rectal palpation in a commercial dairy herd. *Prev. Vet. Med.* **1995**, *23*, 133–141. [[CrossRef](#)]
10. Schwartz, I.; Levy, D. Pathobiology of bovine leukemia virus. *Vet. Res.* **1994**, *25*, 521–536.
11. Yoon, S.S.; Bae, Y.C.; Lee, K.H.; Han, B.; Han, H.R. Characteristics of Bovine Lymphoma Caused by Bovine Leukemia Virus Infection in Holstein-Friesian Dairy Cattle in Korea. *Asian-Australas J. Ani. Sci.* **2005**, *18*, 728–733. [[CrossRef](#)]
12. Burny, A.; Bruck, C.; Cleuter, Y.; Couez, D.; Deschamps, J.; Gregoire, D.; Ghysdael, J.; Kettmann, R.; Mammerickx, M.; Marbaix, G.; et al. Bovine leukaemia virus and enzootic bovine leukosis. *Onderstepoort J. Vet. Res.* **1985**, *52*, 133–144.
13. Zarkik, S.; Decroly, E.; Wattiez, R.; Seidah Nabil, G.; Burny, A.; Ruyschaert, J.-M. Comparative processing of bovine leukemia virus envelope glycoprotein gp72 by subtilisin/kexin-like mammalian convertases. *FEBS Lett.* **1997**, *406*, 205–210. [[CrossRef](#)]
14. Johnston, E.R.; Albritton, L.M.; Radke, K. Envelope Proteins Containing Single Amino Acid Substitutions Support a Structural Model of the Receptor-Binding Domain of Bovine Leukemia Virus Surface Protein. *J. Virol.* **2002**, *76*, 10861–10872. [[CrossRef](#)] [[PubMed](#)]
15. Gatot, J.S.; Callebaut, I.; Van Lint, C.; Demonte, D.; Kerkhofs, P.; Portetelle, D.; Burny, A.; Willems, L.; 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*, 7956–7967. [[CrossRef](#)] [[PubMed](#)]
16. Bruck, C.; Portetelle, D.; Burny, A.; Zavada, J. Topographical analysis by monoclonal antibodies of BLV-gp51 epitopes involved in viral functions. *Virology* **1982**, *122*, 353–362. [[CrossRef](#)]
17. Bruck, C.; Mathot, S.; Portetelle, D.; Berte, C.; Franssen, J.-D.; Herion, P.; Burny, A. Monoclonal antibodies define eight independent antigenic regions on the bovine leukemia virus (BLV) envelope glycoprotein gp51. *Virology* **1982**, *122*, 342–352. [[CrossRef](#)]
18. Bruck, C.; Portetelle, D.; Mammerickx, M.; Mathot, S.; Burny, A. Epitopes of bovine leukemia virus glycoprotein gp51 recognized by sera of infected cattle and sheep. *Leuk. Res.* **1984**, *8*, 315–321. [[CrossRef](#)]
19. Bruck, C.; Rensonnet, N.; Portetelle, D.; Cleuter, Y.; Mammerickx, M.; Burny, A.; Mamoun, R.; Guillemain, B.; Van Der Maaten, M.J.; Ghysdael, J. Biologically active epitopes of bovine leukemia virus glycoprotein GP51: Their dependence on protein glycosylation and genetic variability. *Virology* **1984**, *136*, 20–31. [[CrossRef](#)]
20. Ban, J.; Czene, S.; Altaner, C.; Callebaut, I.; Krchnak, V.; Merza, M.; Burny, A.; Kettmann, R.; Portetelle, D. Mapping of sequential epitopes recognized by monoclonal antibodies on the bovine leukaemia virus external glycoproteins expressed in *Escherichia coli* by means of antipeptide antibodies. *J. Gen. Virol.* **1992**, *73 Pt 9*, 2457–2461. [[CrossRef](#)]

21. Callebaut, I.; Burny, A.; Krchnak, V.; Gras-Masse, H.; Wathelet, B.; Portetelle, D. Use of synthetic peptides to map sequential epitopes recognized by monoclonal antibodies on the bovine leukemia virus external glycoprotein. *Virology* **1991**, *185*, 48–55. [[CrossRef](#)]
22. Bai, L.; Takeshima, S.N.; Isogai, E.; Kohara, J.; Aida, Y. Novel CD8(+) cytotoxic T cell epitopes in bovine leukemia virus with cattle. *Vaccine* **2015**, *33*, 7194–7202. [[CrossRef](#)] [[PubMed](#)]
23. De Brogniez, A.; Mast, J.; Willems, L. Determinants of the bovine leukemia virus envelope glycoproteins involved in infectivity, replication and pathogenesis. *Viruses* **2016**, *8*, 88. [[CrossRef](#)] [[PubMed](#)]
24. Callebaut, I.; Voneche, V.; Mager, A.; Fumiere, O.; Krchnak, V.; Merza, M.; Zavada, J.; Mammerickx, M.; Burny, A.; Portetelle, D. Mapping of B-neutralizing and T-helper cell epitopes on the bovine leukemia virus external glycoprotein gp51. *J. Virol.* **1993**, *67*, 5321–5327. [[CrossRef](#)]
25. Dube, S.; Abbott, L.; Dube, D.K.; Dolcini, G.; Gutierrez, S.; Ceriani, C.; Juliarena, M.; Ferrer, J.; Perzova, R.; Poiesz, B.J. The complete genomic sequence of an in vivo low replicating BLV strain. *Virol. J.* **2009**, *6*, 120. [[CrossRef](#)]
26. Moratorio, G.; Fischer, S.; Bianchi, S.; Tome, L.; Rama, G.; Obal, G.; Carrion, F.; Pritsch, O.; Cristina, J. A detailed molecular analysis of complete bovine leukemia virus genomes isolated from B-cell lymphosarcomas. *Vet. Res.* **2013**, *44*, 19. [[CrossRef](#)]
27. Voneche, V.; Portetelle, D.; Kettmann, R.; Willems, L.; Limbach, K.; Paoletti, E.; Ruysschaert, J.M.; Burny, A.; Brasseur, R. Fusogenic segments of bovine leukemia virus and simian immunodeficiency virus are interchangeable and mediate fusion by means of oblique insertion in the lipid bilayer of their target cells. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 3810–3814. [[CrossRef](#)]
28. Willems, L.; Kettmann, R.; Dequiedt, F.; Portetelle, D.; Voneche, V.; Cornil, I.; Kerkhofs, P.; Burny, A.; Mammerickx, M. In vivo infection of sheep by bovine leukemia virus mutants. *J. Virol.* **1993**, *67*, 4078–4085. [[CrossRef](#)]
29. Willems, L.; Gatot, J.S.; Mammerickx, M.; Portetelle, D.; Burny, A.; Kerkhofs, P.; Kettmann, R. The YXXL signalling motifs of the bovine leukemia virus transmembrane protein are required for in vivo infection and maintenance of high viral loads. *J. Virol.* **1995**, *69*, 4137–4141. [[CrossRef](#)]
30. Reichert, M.; Winnicka, A.; Willems, L.; Kettmann, R.; Cantor, G.H. Role of the proline-rich motif of bovine leukemia virus transmembrane protein gp30 in viral load and pathogenicity in sheep. *J. Virol.* **2001**, *75*, 8082–8089. [[CrossRef](#)]
31. Inabe, K.; Nishizawa, M.; Tajima, S.; Ikuta, K.; Aida, Y. The YXXL sequences of a transmembrane protein of bovine leukemia virus are required for viral entry and incorporation of viral envelope protein into virions. *J. Virol.* **1999**, *73*, 1293–1301. [[CrossRef](#)]
32. Portetelle, D.; Limbach, K.; Burny, A.; Mammerickx, M.; Desmettre, P.; Riviere, M.; Zavada, J.; Paoletti, E. Recombinant vaccinia virus expression of the bovine leukaemia virus envelope gene and protection of immunized sheep against infection. *Vaccine* **1991**, *9*, 194–200. [[CrossRef](#)]
33. Kumar, S.; Andrew, M.E.; Boyle, D.B.; Brandon, R.B.; Lavin, M.F.; Daniel, R.C. Expression of bovine leukaemia virus envelope gene by recombinant vaccinia viruses. *Virus Res.* **1990**, *17*, 131–142. [[CrossRef](#)]
34. Gillet, N.; Florins, A.; Boxus, M.; Burteau, C.; Nigro, A.; Vandermeers, F.; Balon, H.; Bouzar, A.B.; Defoiche, J.; Burny, A.; et al. Mechanisms of leukemogenesis induced by bovine leukemia virus: Prospects for novel anti-retroviral therapies in human. *Retrovirology* **2007**, *4*, 18. [[CrossRef](#)]
35. Rice, N.R.; Stephens, R.M.; Burny, A.; Gilden, R.V. The gag and pol genes of bovine leukemia virus: Nucleotide sequence and analysis. *Virology* **1985**, *142*, 357–377. [[CrossRef](#)]
36. Llames, L.; Goyache, J.; Domenech, A.; Montaña, A.V.; Suarez, G.; Gomez-Lucia, E. Cellular distribution of bovine leukemia virus proteins gp51SU, Pr72env, and Pr66gag-pro in persistently infected cells. *Virus Res.* **2001**, *79*, 47–57. [[CrossRef](#)]
37. Yoshinaka, Y.; Katoh, I.; Copeland, T.D.; Smythers, G.W.; Oroszlan, S. Bovine leukemia virus protease: Purification, chemical analysis, and in vitro processing of gag precursor polyproteins. *J. Virol.* **1986**, *57*, 826–832. [[CrossRef](#)]
38. Juliarena, M.A.; Gutierrez, S.E.; Ceriani, C. Determination of proviral load in bovine leukemia virus-infected cattle with and without lymphocytosis. *Am. J. Vet. Res.* **2007**, *68*, 1220–1225. [[CrossRef](#)]

39. World Organisation for Animal Health. WAHIS Interface Animal Health Situation: Disease Distribution Maps. Available online: https://www.oie.int/wahis_2/public/wahid.php/Diseaseinformation/Diseasedistributionmap?disease_type_hidden=&disease_id_hidden=&selected_disease_name_hidden=&disease_type=0&disease_id_terrestrial=35&species_t=0&disease_id_aquatic=-999&species_a=0&sta_method=semesterly&selected_start_year=2016&selected_report_period=1&selected_start_month=12&date_submit=OK (accessed on 4 January 2020).
40. European Commission. *Bovine and Swine Diseases Situation 2017*; European Commission: Brussels, Belgium, 2017; pp. 1–35.
41. Rodriguez, S.M.; Golemba, M.D.; Campos, R.H.; Trono, K.; Jones, L.R. Bovine leukemia virus can be classified into seven genotypes: Evidence for the existence of two novel clades. *J. Gen. Virol.* **2009**, *90*, 2788–2797. [[CrossRef](#)]
42. Balic, D.; Lojkic, I.; Periskic, M.; Bedekovic, T.; Jungic, A.; Lemo, N.; Roic, B.; Cac, Z.; Barbic, L.; Madic, J. Identification of a new genotype of bovine leukemia virus. *Arch. Virol.* **2012**, *157*, 1281–1290. [[CrossRef](#)]
43. Rola-Luszczak, M.; Pluta, A.; Olech, M.; Donnik, I.; Petropavlovskiy, M.; Gerilovych, A.; Vinogradova, I.; Choudhury, B.; Kuzmak, J. The molecular characterization of bovine leukaemia virus isolates from Eastern Europe and Siberia and its impact on phylogeny. *PLoS ONE* **2013**, *8*, e58705. [[CrossRef](#)]
44. Polat, M.; Takeshima, S.N.; Hosomichi, K.; Kim, J.; Miyasaka, T.; Yamada, K.; Arainga, M.; Murakami, T.; Matsumoto, Y.; de la Barra Diaz, V.; et al. 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*, 4. [[CrossRef](#)]
45. Lee, E.; Kim, E.J.; Ratthanophart, J.; Vitoonpong, R.; Kim, B.H.; Cho, I.S.; Song, J.Y.; Lee, K.K.; Shin, Y.K. Molecular epidemiological and serological studies of bovine leukemia virus (BLV) infection in Thailand cattle. *Infect Genet Evol.* **2016**, *41*, 245–254. [[CrossRef](#)]
46. Polat, M.; Moe, H.H.; Shimogiri, T.; Moe, K.K.; Takeshima, S.N.; Aida, Y. The molecular epidemiological study of bovine leukemia virus infection in Myanmar cattle. *Arch. Virol.* **2017**, *162*, 425–437. [[CrossRef](#)]
47. Wang, M.; Wang, Y.; Baloch, A.R.; Pan, Y.; Xu, F.; Tian, L.; Zeng, Q. Molecular epidemiology and characterization of bovine leukemia virus in domestic yaks (*Bos grunniens*) on the Qinghai-Tibet Plateau, China. *Arch. Virol.* **2018**, *163*, 659–670. [[CrossRef](#)]
48. Yu, C.; Wang, X.; Zhou, Y.; Wang, Y.; Zhang, X.; Zheng, Y. Genotyping bovine leukemia virus in dairy cattle of Heilongjiang, northeastern China. *BMC Vet. Res.* **2019**, *15*. [[CrossRef](#)]
49. Murakami, K.; Kobayashi, S.; Konishi, M.; Kameyama, K.; Yamamoto, T.; Tsutsui, T. The recent prevalence of bovine leukemia virus (BLV) infection among Japanese cattle. *Vet. Microbiol.* **2011**, *148*, 84–88. [[CrossRef](#)] [[PubMed](#)]
50. Trono, K.G.; Perez-Filgueira, D.M.; Duffy, S.; Borca, M.V.; Carrillo, C. Seroprevalence of bovine leukemia virus in dairy cattle in Argentina: Comparison of sensitivity and specificity of different detection methods. *Vet. Microbiol.* **2001**, *83*, 235–248. [[CrossRef](#)]
51. Monti, G.; Schrijver, R.; Beier, D. Genetic diversity and spread of Bovine leukaemia virus isolates in Argentine dairy cattle. *Arch. Virol.* **2005**, *150*, 443–458. [[CrossRef](#)]
52. Kobayashi, S.; Tsutsui, T.; Yamamoto, T.; Hayama, Y.; Kameyama, K.; Konishi, M.; Murakami, K. Risk factors associated with within-herd transmission of bovine leukemia virus on dairy farms in Japan. *BMC Vet. Res.* **2010**, *6*, 1. [[CrossRef](#)]
53. Sandev, N.; Ilieva, D.; Sizov, I.; Rusenova, N.; Iliev, E. Prevalence of Enzootic Bovine Leukosis in the Republic of Bulgaria in 1997–2004. *Vet. Arh.* **2006**, *76*, 263–268.
54. Bauermann, F.V.; Ridpath, J.F.; Dargatz, D.A. Bovine leukemia virus seroprevalence among cattle presented for slaughter in the United States. *J. Vet. Diagn. Investig. Off. Publ. Am. Assoc. Vet. Lab. Diagn. Inc.* **2017**, *29*, 704–706. [[CrossRef](#)] [[PubMed](#)]
55. LaDronka, R.M.; Ainsworth, S.; Wilkins, M.J.; Norby, B.; Byrem, T.M.; Bartlett, P.C. Prevalence of Bovine Leukemia Virus Antibodies in US Dairy Cattle. *Vet. Med. Int.* **2018**, *2018*, 8. [[CrossRef](#)]
56. Polat, M.; Takeshima, S.-n.; Aida, Y. Epidemiology and genetic diversity of bovine leukemia virus. *Virol. J.* **2017**, *14*. [[CrossRef](#)]
57. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
58. Rognes, T.; Flouri, T.; Nichols, B.; Quince, C.; Mahé, F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* **2016**, *4*, e2584. [[CrossRef](#)]

59. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)]
60. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]
61. Tamura, K.; Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **1993**, *10*, 512–526. [[CrossRef](#)]
62. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **1980**, *16*, 111–120. [[CrossRef](#)]
63. Hasegawa, M.; Kishino, H.; Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **1985**, *22*, 160–174. [[CrossRef](#)]
64. Felsenstein, J. Confidence limits on phylogenies: An approach using the bootstrap. *Evol. Int. J. Org. Evol.* **1985**, *39*, 783–791. [[CrossRef](#)] [[PubMed](#)]
65. Pandey, G.S.; Simulundu, E.; Mwiinga, D.; Samui, K.L.; Mweene, A.S.; Kajihara, M.; Mangani, A.; Mwenda, R.; Ndebe, J.; Konnai, S.; et al. Clinical and subclinical bovine leukemia virus infection in a dairy cattle herd in Zambia. *Arch. Virol.* **2017**, *162*, 1051–1056. [[CrossRef](#)] [[PubMed](#)]
66. Phiri, M.M.; Kaimoyo, E.; Changula, K.; Silwamba, I.; Chambaro, H.M.; Kapila, P.; Kajihara, M.; Simuunza, M.; Muma, J.B.; Pandey, G.S.; et al. Molecular detection and characterization of genotype 1 bovine leukemia virus from beef cattle in the traditional sector in Zambia. *Arch. Virol.* **2019**, *164*, 2531–2536. [[CrossRef](#)]
67. Lee, E.; Kim, E.J.; Joung, H.K.; Kim, B.H.; Song, J.Y.; Cho, I.S.; Lee, K.K.; Shin, Y.K. Sequencing and phylogenetic analysis of the gp51 gene from Korean bovine leukemia virus isolates. *Virol. J.* **2015**, *12*, 64. [[CrossRef](#)]
68. Felmer, R.; Munoz, G.; Zuniga, J.; 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. [[CrossRef](#)]
69. Ababneh, M.M.; Al-Rukibat, R.K.; Hananeh, W.M.; Nasar, A.T.; Al-Zghoul, M.B. Detection and molecular characterization of bovine leukemia viruses from Jordan. *Arch. Virol.* **2012**, *157*, 2343–2348. [[CrossRef](#)]
70. Asfaw, Y.; Tsuduku, S.; Konishi, M.; Murakami, K.; Tsuboi, T.; Wu, D.; Sentsui, H. Distribution and superinfection of bovine leukemia virus genotypes in Japan. *Arch. Virol.* **2005**, *150*, 493–505. [[CrossRef](#)]
71. Bicka, L.; Kuzmak, J.; Rola, M.; Beier, D. Detection of genetic diversity among bovine leukemia virus population by single-strand conformational polymorphism analysis. *Bull. Vet. Inst. Pulawy* **2002**, *46*, 205–212. [[CrossRef](#)]
72. Fechner, H.; Blankenstein, P.; Looman, A.C.; Elwert, J.; Geue, L.; Albrecht, C.; Kurg, A.; Beier, D.; Marquardt, O.; Ebner, D. Provirus variants of the bovine leukemia virus and their relation to the serological status of naturally infected cattle. *Virology* **1997**, *237*, 261–269. [[CrossRef](#)]
73. Zhao, X.; Buehring, G.C. Natural genetic variations in bovine leukemia virus envelope gene: Possible effects of selection and escape. *Virology* **2007**, *366*, 150–165. [[CrossRef](#)]
74. Pluta, A.; Albritton, L.M.; Rola-Luszczak, M.; Kuzmak, J. Computational analysis of envelope glycoproteins from diverse geographical isolates of bovine leukemia virus identifies highly conserved peptide motifs. *Retrovirology* **2018**, *15*, 2. [[CrossRef](#)] [[PubMed](#)]
75. Pluta, A.; Rola-Luszczak, M.; Kubis, P.; Balov, S.; Moskalik, R.; Choudhury, B.; Kuzmak, J. Molecular characterization of bovine leukemia virus from Moldovan dairy cattle. *Arch. Virol.* **2017**, *162*, 1563–1576. [[CrossRef](#)]
76. Portetelle, D.; Couez, D.; Bruck, C.; Kettmann, R.; Mammerickx, M.; Van der Maaten, M.; Brasseur, R.; Burny, A. Antigenic variants of bovine leukemia virus (BLV) are defined by amino acid substitutions in the NH2 part of the envelope glycoprotein gp51. *Virology* **1989**, *169*, 27–33. [[CrossRef](#)]
77. Portetelle, D.; Dandoy, C.; Burny, A.; Zavada, J.; Siakkou, H.; Gras-Masse, H.; Drobecq, H.; Tartar, A. Synthetic peptides approach to identification of epitopes on bovine leukemia virus envelope glycoprotein gp51. *Virology* **1989**, *169*, 34–41. [[CrossRef](#)]
78. Mamoun, R.Z.; Morisson, M.; Rebeyrotte, N.; Busetta, B.; Couez, D.; Kettmann, R.; Hospital, M.; Guillemain, B. Sequence variability of bovine leukemia virus *env* gene and its relevance to the structure and antigenicity of the glycoproteins. *J. Virol.* **1990**, *64*, 4180–4188. [[CrossRef](#)]
79. Moratorio, G.; Obal, G.; Dubra, A.; Correa, A.; Bianchi, S.; Buschiazzo, A.; Cristina, J.; Pritsch, O. Phylogenetic analysis of bovine leukemia viruses isolated in South America reveals diversification in seven distinct genotypes. *Arch. Virol.* **2010**, *155*, 481–489. [[CrossRef](#)]

80. Camargos, M.F.; Pereda, A.; Stancek, D.; Rocha, M.A.; dos Reis, J.K.; Greiser-Wilke, I.; Leite, R.C. Molecular characterization of the *env* gene from Brazilian field isolates of Bovine leukemia virus. *Virus Genes* **2007**, *34*, 343–350. [[CrossRef](#)]
81. Auerbach, M.R.; Shu, C.; Kaplan, A.; Singh, I.R. Functional characterization of a portion of the Moloney murine leukemia virus *gag* gene by genetic footprinting. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 11678. [[CrossRef](#)]
82. Morcock, D.; Kane, B.; Casas-Finet, J. Fluorescence and nucleic acid binding properties of the human T-cell leukemia virus-type 1 nucleocapsid protein. *Biochim. Et. Biophys. Acta* **2000**, *1481*, 381–394. [[CrossRef](#)]
83. Fu, W.; Hu, W.-S. Functional Replacement of Nucleocapsid Flanking Regions by Heterologous Counterparts with Divergent Primary Sequences: Effects of Chimeric Nucleocapsid on the Retroviral Replication Cycle. *J. Virol.* **2003**, *77*, 754–761. [[CrossRef](#)]
84. Gorelick, R.J.; Henderson, L.E.; Hanser, J.P.; Rein, A. Point Mutants of Moloney Murine Leukemia Virus that Fail to Package Viral RNA: Evidence for Specific RNA Recognition by a Zinc FingerLike Protein Sequence. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 8420–8424. [[CrossRef](#)]
85. Schwartz, M.D.; Fiore, D.; Panganiban, A.T. Distinct functions and requirements for the Cys-His boxes of the human immunodeficiency virus type 1 nucleocapsid protein during RNA encapsidation and replication. *J. Virol.* **1997**, *71*, 9295–9305. [[CrossRef](#)]
86. Bowzard, J.B.; Bennett, R.P.; Krishna, N.K.; Ernst, S.M.; Rein, A.; Wills, J.W. Importance of basic residues in the nucleocapsid sequence for retrovirus Gag assembly and complementation rescue. *J. Virol.* **1998**, *72*, 9034–9044. [[CrossRef](#)]
87. Poon, D.T.; Wu, J.; Aldovini, A. Charged amino acid residues of human immunodeficiency virus type 1 nucleocapsid p7 protein involved in RNA packaging and infectivity. *J. Virol.* **1996**, *70*, 6607–6616. [[CrossRef](#)]
88. Schmalzbauer, E.; Strack, B.; Dannull, J.; Guehmann, S.; Moelling, K. Mutations of basic amino acids of NCp7 of human immunodeficiency virus type 1 affect RNA binding in vitro. *J. Virol.* **1996**, *70*, 771–777. [[CrossRef](#)]
89. Dannull, J.; Surovoy, A.; Jung, G.; Moelling, K. Specific binding of HIV-1 nucleocapsid protein to PSI RNA in vitro requires N-terminal zinc finger and flanking basic amino acid residues. *Embo J.* **1994**, *13*, 1525–1533. [[CrossRef](#)]
90. Jewell, N.A.; Mansky, L.M. In the beginning: Genome recognition, RNA encapsidation and the initiation of complex retrovirus assembly. *J. Gen. Virol.* **2000**, *81*, 1889–1899. [[CrossRef](#)]
91. Wang, H.; Norris, K.M.; Mansky, L.M. Involvement of the matrix and nucleocapsid domains of the bovine leukemia virus Gag polyprotein precursor in viral RNA packaging. *J. Virol.* **2003**, *77*, 9431–9438. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).