

# Mutational and phylogenetic status of west siberian strains of BLV

Natalia Blazhko<sup>1</sup>, Kirill Shatokhin<sup>1\*</sup>, Yuriy Khripko<sup>2,3</sup>, and Charity Ngirande<sup>1</sup>, Nikolai Kochnev<sup>1</sup>

<sup>1</sup>Novosibirsk State Agrarian University, 630039Novosibirsk, Russia

<sup>2</sup>State Research Center of Virology and Biotechnology “Vector”, Koltsovo630559, Novosibirsk Region, Russia

<sup>3</sup>Research Institute of Experimental and Clinical Medicine, 630117Novosibirsk, Russia

**Abstract.** The study is devoted of full-genome BLV sequences circulating in cattle populations of the Novosibirsk region, Russia. The phylogenetic tree shows that the West Siberian isolates are quite closely related to such previously isolated strains as *AF399704* (Brazil), *AP018007*, *AP018016*, *AP018019*, *LC007988*, *LC007991* (Japan) and *EF065638* (Belgium) we calculations show that the number of mutations that could independently occur in parallel evolving BLV strains significantly exceeds the expected number based on the probability of corresponding substitutions. It was also found that the studied isolates have some mutations, the presence of which, at first glance, is possible only with their divergent development in different independently evolving branches. However, calculations show that the probability of an independent origin of an identical mutation is extremely small, which indicates the possibility of exchanging RNA sites between isolates circulating in West Siberian cattle populations.

## 1 Introduction

Viruses are the most genetically diverse category of organisms, whose total number of genes exceeds their total number in living cell forms [1]. Bovine leukaemia, a very common disease of cattle, causes millions of euros worth of damage to the global economy [2]. Causative agent of bovine leukemia, bovine leukemia virus (*BLV*) is an RNA virus of the *Retroviridae* family [3, 4]. The molecular structure of the virus has been thoroughly studied by many researchers [5,6], and the description of specific genes and the nucleotide composition of individual strains are deposited in various publicly available databases [7]. Although *BLV* has a relatively conserved genome [8], speciation within the global population still takes place, which results in a rather difficult phylogenetic classification. There is an opinion that the global *BLV* gene pool can be sufficiently divided into three or four sets[9]. On the other hand, at least 10 *BLV* genotypes have been described, within which separate strains have been isolated [10, 11, 12], even though, their study is irregular.

---

\*Corresponding author: true\_genetic@mail.ru

In particular, the molecular structure and classification of North-and South American isolates have been fairly well described in scientific literature [3, 13]. However, descriptions of the molecular structure of Eastern European and Russian isolates, and especially those occurring in Siberia, has been mostly based on the lengths of restriction fragment meta-polymorphism analysis of individual [5, 14]. A full-genomic study of Russian *BLV* isolates has not yet been carried out due to the high cost of the analysis [15].

Taking into account the regular import of breeding young animals to Western Siberia there is a possibility of the spread of *BLV* strains that were not previously found in this region. The combination of factors: periodic updating of the gene pool of Siberian cattle, its adaptation to the conditions of keeping and feeding, artificial selection for milk productivity, enrichment of the *BLV* gene pool due to the import of new strains and continuous mutagenesis create quite sufficient conditions for the formation of a new previously undescribed strain in Western Siberia. The purpose of our research is the phylogenetic analysis of the complete genomes of the virus isolated from the biological material of cattle bred in the Novosibirsk region.

## 2 Materials and methods

The object of research was the cattle of black-and-white holstein breed obtained from farms around the Novosibirsk region (n=3954). Blood samples were drawn from the subcutaneous vein of animals in sterile disposable IV tubes using EDTA a preservative. The blood (15-25 ml of each sample) was analyzed using an automatic hematological analyzer (RCE-90 Vetin) in accordance with GOST 25382 [16]. Hematological studies and DNA isolation (using the "DNA-Sorb-B" kit; FSUN Central Research Institute of Epidemiology, Russia) were performed on the day of receipt of the material in the laboratory. DNA was isolated. According to the cumulative results, 780 samples showed the presence of *BLV* proviral DNA in blood samples [14]. Of these 15 samples were selected for sequencing, meeting the following requirements: Sequencing was performed using the Sanger method according to the recommendations [17] on the 3730xl DNA Analyzer (Applied Biosystems, 850 Lincoln Center Drive, Foster City, United States). In addition to the 15 studied *BLV* sequences, the sequences of eight genotypes presented in the DDBJ database were included in the alignment (<http://getentry.ddbj.nig.ac.jp/>). The assessment of the degree of evolutionary divergence was calculated in pairs, as a result of which a matrix of evolutionary distances was constructed. The phylogenetic analysis was carried out using the Bayes method [18] with the GTR substitution model and the neighboring compound method (NJ) and the Tamura-Nei nucleotide substitution model with cluster reliability estimated at loading 1000 repeats (Fig. 1). The average nucleotide distances were estimated by applying the Tamura Nei model in the Ugene program.

The isolation of strains due to the lack of specially developed algorithms [19] was determined using standard statistical methods. The degree of randomness of the differences was determined by comparing the actual genetic distances, expressed as a percentage, with the critical value for randomness equal to 1% [1] using the Student's criterion [20]. The reliability of pairwise differences in genetic distances was evaluated by the criterion  $\chi^2$  [20]. The isolation factor was determined by analyzing the variances of the genetic distances between the selected samples and between each of the local isolates with the control strain FLK was determined by one-factor analysis of variance using the One-way function of the ANOVA application included in the STATISTICA8 software package.

To further confirm the uniqueness of the sequenced samples the probability of repeating combinations of single-nucleotide substitutions was calculated. Since probability in fact is a

qualitative feature, the error was calculated according to the corresponding formula [20]. In the absence of generally accepted frequency constant mutations [21, 22] the probability was calculated using two models developed directly by the author's team. The first of them is simplified, where the probability of replacement, based on the fact that only four types of nucleotides (A, T, G, C) are included in the DNA is a priori 0.33 [23]. The formula for calculating the simplified probability ( $\omega_s$ ) of repeating mutations of each particular isolate is as follows:

$$\omega_s = \frac{0.33 n}{8401} \times 100 \tag{1}$$

where  $n$  is the number of mutations of a particular isolate relative to the most related strain; 8401 is the length of the nucleotide sequence of the control strain *BLV* (according to NCBI: BioProject: PRJNA485481).

Since, according to some data, the probability of each type of single-nucleotide substitutions is a gene-specific value in the conditions of a particular biological species [23] a more complex model took into account the frequency of each type of single-nucleotide substitutions relative to the total number. The calculation formula looks like this:

$$\omega = \frac{K_i}{8401} \times 100 \tag{2}$$

Where  $K_i$  is the total probability coefficient of nucleotide substitutions, taking into account the probability of each type of single-nucleotide substitutions and their number, defined as:

$$K_i = \begin{cases} k^{A>T}n_i^{A>T} + k^{A>C}n_i^{A>C} + k^{A>G}n_i^{A>G} \\ k^{T>A}n_i^{T>A} + k^{T>C}n_i^{T>C} + k^{T>G}n_i^{T>G} \\ k^{C>A}n_i^{C>A} + k^{C>T}n_i^{C>T} + k^{C>G}n_i^{C>G} \\ k^{G>A}n_i^{G>A} + k^{G>T}n_i^{G>T} + k^{G>C}n_i^{G>C} \end{cases} \tag{3}$$

Where  $k$  is the probability coefficient of each type of nucleotide substitution equal to the ratio of the number of substitutions to the total number of SNPs in the sample;  $n_i$  is the number of this type of SNP in the isolate. The proportion of mutations ( $p_i, \%$ ) with the probability of repetition was calculated by the formula:

$$p_i = \frac{n_i}{N_i} \times 100 \tag{4}$$

Where  $n_i$  is the number of SNPs with the probability of repetition,  $N_i$  is the total number of mutations. The expected number of repeated mutations was calculated by the formula:

$$n_{oc} = n_i \times 0.33^2 \tag{5}$$

Where  $n_i$  is the observed number of SNPs with the probability of repetition; 0.33 is the simplified probability of replacement with a similar nucleotide. Taking into account the different probability of occurrence of transitions and transversions [1], the expected number of repeated mutations was additionally calculated taking into account each type of substitutions. Repeated mutations were taken into account if they were detected in different strains that did not have a common cleavage from the dendrogram. The calculation formula looks like this:

$$n_o = \sum p_i k_i^2 \tag{6}$$

Reverse mutations were taken into account if they were present in the studied isolate and the control strain of FLK provided that it had a cleavage from the branch of strains where a direct mutation was detected. The proportion of probable reverse substitutions was calculated using the formula:

$$p'_i = \frac{n'_i}{8401} \times 100 \quad (7)$$

Where  $n'_i$  the number of reverse mutations; 8401 is the length of the sequence of the control strain FLK (b.p.).

### 3 Results

Based on the results of phylogenetic analysis (Fig. 1), it can be concluded that there are several clusters within the sample that are related to strains *EF065638* (Belgium), *AF399704* (Brazil), *AP018007*, *AP018016*, *AP018019*, *LC007988* and *LC007991* (Japan). It is interesting that sample No. 2 is located separately from other isolates of the first genotype. To determine the phylogenetic status of *BLV* samples isolated from the blood of cattle of the Novosibirsk region, the genetic distances between individual isolates were evaluated using data on the molecular structure of strains stored in the DDBJ database (<http://getentry.ddbj.nig.ac.jp/>) and the control strain FLK (Table 1, Fig. 1). It is noteworthy that the average distance estimated by the degree of evolutionary divergence between the isolated isolates ( $0.0229 \pm 0.0013$ ) was slightly larger ( $t = 1.25$ ; not valuable) than the average distance between each of the selected samples with the control strain FLK ( $0.0185 \pm 0.0032$ ). Testing of the distance factor (isolate-isolate or isolate-FLK) did not confirm the isolation of the entire array of local isolates relative to the control strain ( $F=1.488$ ;  $p=0.225$ ), which also indicates the heterogeneity of the origin of *BLV* isolates circulating on the cattle of the Novosibirsk region.

In the case of isolated mutagenesis, it is much more likely that the average distance of evolutionary divergence between the isolates of the Novosibirsk region is significantly smaller than between each of them and the control strain of FLK. The criterion  $\chi^2$ , which evaluates the reliability of the differences in the minimum distances between each of the isolates identified in the Novosibirsk region, as well as between the Siberian isolates and strains from the DDBJ database did not exceed 2.7 which indicates the lack of reliability of the differences. The minimum differences between the samples (Table. 1) have no significant differences from the random error (1%), despite the fact that the largest genetic distances indicate the accumulation of non-random mutations. At the same time, despite the apparent proximity, none of the isolates has accumulated a sufficient number of mutations relative to both the samples we isolated and the strains from the DDBJ database (Table. 1), which would be considered an independent entity without correction for random discrepancies. It is curious that the degree of differences between individual samples (Table. 1) from the control strain, FLK still exceeded 1%, despite the fact that 60% (9 out of 15) of the samples did not accumulate a critical (1%) number of nucleotide substitutions.

**Table 1.** Minimum distances of each isolate.

№	With other isolates in the Novosibirsk region		$\chi^2$	With strains from the DDBJ database		$\chi^2$	With the FLK control strain	
	isolate	Dmin, %		strain	Dmin, %		Dmin, %	$\chi^2$
1	9	0.40	0.360	AP018019 AP018016 AP018007	0.60	0.160	0.69	0.0961
2	7, 8	1.20	0.040	LC007988	0.30	0.490	0.94	0.004
3	4	0.70	0.090	AF399704	0.90	0.010	3.12	4.494**
4	3	0.70	0.090	AF399704	1.40	0.160	3.62	6.864*
5	6	0.40	0.360	AP018019	0.80	0.040	0.83	0.029
6	5	0.40	0.360	AP018019	0.60	0.160	0.68	0.102
7	8	0.20	0.640	AP018019	0.80	0.040	0.81	0.036
8	7, 9	0.20	0.640	AP018016	0.60	0.160	0.64	0.130
9	8	0.20	0.640	AP018007	0.60	0.160	0.68	0.102
10	11	0.08	0.846	EF065638	1.18	0.032	3.25	5.063*
11	10	0.08	0.846		1.18	0.032	3.21	4.884*
12	13	0.20	0.640		1.10	0.010	3.1	4.410*
13	12	0.20	0.640		1.10	0.010	3.29	5.244*
14	6	2.10	1.210	LC007991	1.00	0.000	2.19	1.416
15	6	0.50	0.250	AP018019	0.60	0.160	0.69	0.0961

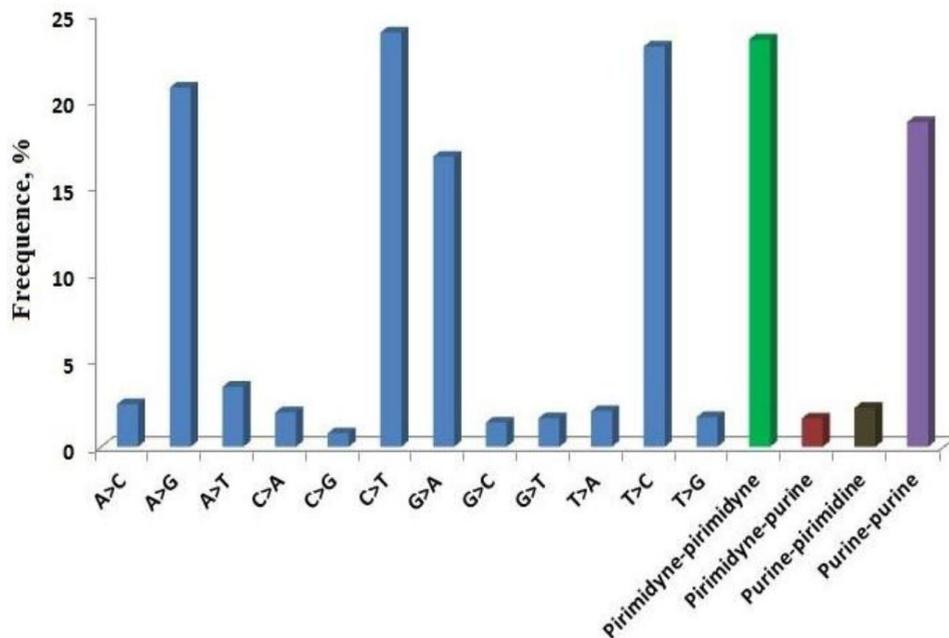
D min – minimum distance,  $\chi^2$  Checking for randomness of differences.

**Table 2.** Probability ( $\omega$ ) of repeating the combination of nucleotides of west Siberian *BLV* isolates using the FLK strain as a control.

No	Number of SNP	With the probability of replacing the nucleotide with any other =0.33		Taking into account the probability of each type SNP	
		$\omega \pm \bar{S}_\omega$ , %	Reliability of differences with 1% expressed by the Student's criterion	$\omega \pm \bar{S}_\omega$ , %	Reliability of differences with 1% expressed by the Student's criterion.
1	155	0.6089±0.6248	0.626	0.3470±0.4723	1.383
2	176	0.6913±0.6246	0.494	0.3618±0.4526	1.410
3	359	1.4102±0.6223	0.659	0.8059±0.4719	0.411
4	401	1.5752±0.6218	0.925	0.8603±0.4612	0.303
5	167	0.6560±0.6247	0.551	0.3664±0.4676	1.355
6	158	0.6206±0.6248	0.607	0.3544±0.4728	1.366
7	169	0.6638±0.6247	0.538	0.3695±0.4667	1.351
8	153	0.6010±0.6249	0.639	0.3492±0.4769	1.365
9	165	0.6481±0.6247	0.563	0.3562±0.4638	1.388
10	364	1.4298±0.6222	0.691	0.8123±0.4705	0.399
11	359	1.4102±0.6223	0.659	0.8016±0.4706	0.422
12	166	0.6521±0.6247	0.557	0.3272±0.4433	1.518
13	367	1.4416±0.6222	0.710	0.8228±0.4715	0.376
14	274	1.0763±0.6234	0.122	0.5975±0.4656	0.864
15	150	0.5892±0.6249	0.657	0.3405±0.4756	1.387



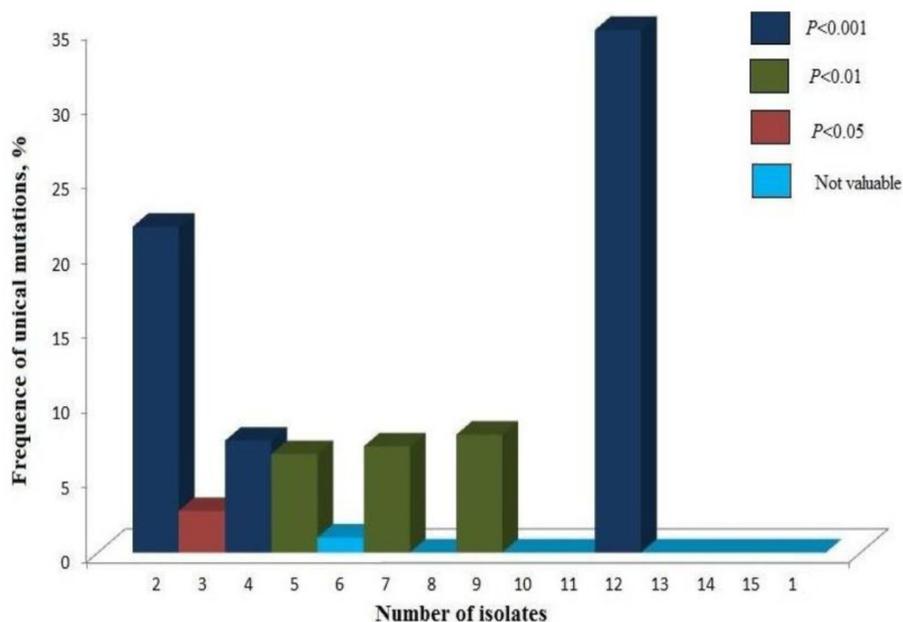
transversions that occurred in our study (Fig. 2), the probability of uniqueness of mutations was expected to be even lower than with random mutagenesis. In both cases, the probability of repeating mutations of any of the isolates relative to the control strain of FLK had no significant differences from 1% (Table. 2), that is, it corresponds to the value of the random error [1]. It would seem that each *BLV* isolate isolated in the Novosibirsk region should represent a unique neoplasm in its own way. However, in practice, some of the samples do not have or practically do not have mutations peculiar only to this sample. Only less than half of the samples had a frequency of unique mutations exceeding the value of the random error (Fig. 3).



**Fig.2.** Frequency of unique SNP in the *BLV* isolates of the Novosibirsk region.

The type of purine-purine, pyrimidine-purine, purine-pyrimidine or pyrimidine-pyrimidine substitutions practically does not affect the occurrence of identical SNPs in independently evolving strains ( $F=0.200$ ;  $p=0.637$ ), which, taking into account the unequal probability of transitions and transversions (Fig. 2), rather indicates not the independent occurrence of the same SNPs, but the "embedding" of the corresponding fragment of the genome. Surprisingly, there was no stable correlation between the total number of mutations and the number of probable repeats ( $r=0.487$ , not valuable). Thus, the calculations indicate either the very fact of the presence of repeated mutations as such, or the presence of recombinations in the selected samples, which is a fairly likely scenario in the conditions of joint livestock keeping and will allow us to say that recombinations, insertions and deletions could make some "corrections" to the phylogenetic analysis. So the mutation T>C in position N.o. 2766 is present in the samples № 1, 2, 3, 4, 7-11, 13, 14 and at the same time it is absent in samples 5, 6, 12, 15. It is interesting that the dendrogram (Fig. 1) assumes an early cleavage of the evolutionary branch of sample No. 2, which has a substitution that is absent in the late split-off samples No. 5, 6 and 15. This means either an independently generated identical mutation in the evolutionary branches of sample No. 2 and samples № 1, 3, 4, 7-11, 13, 14, or the reverse substitution of C>T in the phylogenetic

branch of samples No. 5, 6 and 15. Or, what seems much more plausible, there was an exchange of the corresponding sites of nucleic acids already directly in the livestock population on the territory of Western Siberia. Earlier studies of influenza viruses [25] and *HIV* [26] recorded the exchange of nucleic acid sites between viral particles, often accompanied by the formation of new strains, which indirectly confirms the assumption put forward.



**Fig.3.** Frequency of unique SNP in the BLV isolates of the Novosibirsk region

Reverse nucleotide substitutions apparently also took place in the formation process of *BLV* isolates of the Novosibirsk region, which undoubtedly introduced certain errors in the phylogenetic analysis (Table 3). As an example, we give a replacement T>C at position 3256, which is present in all samples except No. 5, which allows for a reverse replacement in the phylogenetic branch of this isolate, although the probability of a reverse replacement according to a random model is 1/6, this value is still not zero. But it should be understood that such coincidences can be observed in independent evolutionary branches of different strains. However, this is still not enough to allow for the independent random occurrence of several dozen identical mutations in independent evolutionary branches. In general, the number of reverse substitutions in 13 out of 15 samples had serious differences with the theoretically expected, and there were deviations both in the larger and smaller sides (Table 3). Moreover, the number of samples that did not belong to the 1st genotype *BLV* namely 10, 11, 12 and 13, had more than expected. It is noteworthy that the proportion of reverse mutations in sample No. 12 is  $34.2937 \pm 0.5179$  %, which significantly differs from the random error ( $P > 0.001$ ). At the same time, the proportion of expected reverse substitutions in other isolates does not differ significantly from 1%, and therefore falls within the range of random error.

**Table 3.** The number of observed and expected, the reverse substitutions.

No	The number of back substitutions		$\chi^2$	Percentage of replacements, %	t
	Observed	Expected			
1	5	15	6.309*	0.075±0.030	0.855
2	1	15	12.830***	0.153±0.043	0.718
3	28	34	1.167	0.014±0.013	0.972
4	2	37	32.732***	0.390±0.068	0.373
5	1	15	0.0309	0.001±0.002	0.999
6	14	5	18.439***	0.220±0.051	0.609
7	5	15	7.091**	0.084±0.032	0.838
8	4	15	7.717**	0.092±0.033	0.825
9	7	15	4.021*	0.048±0.024	0.907
10	73	35	41.520**	0.494±0.077	0.256
11	76	34	50.201***	0.598±0.084	0.162
12	217	14	2881.014***	34.294±0.518	1108.471*
13	59	35	15.755***	0.188±0.047	0.660
14	3	25	19.566***	0.231±0.053	0.588
15	7	14	3.616***	0.043±0.023	0.916

t – the value of the Student criterion showing the confidence of differences from 1 %.

## 4 Discussion

The similarity of the isolates that were studied to different strains geographically related to Belgium, Japan, the USA, Brazil and Argentina (Fig. 3) is of considerable interest, since this feature is not typical for most viruses. For example, rabbit hemorrhagic disease virus (RHDV) strains tend to cluster geographically [27], which was not observed in this case. These results, as well as others from a number of previous studies [13], in fact highlight the circulation of several strains in the same geographical location. A fairly logical explanation for this fact is the constant movement of livestock associated with the selection necessity, in particular with the purpose of improving the productivity of animals. However, a single introduction of *BLV* strains into Western Siberia in the process of improving local black-and-white cattle by gene pool introduction of Holstein Friesian cattle may well have caused the heterogeneity of circulating *BLV* strains. However, the question of direct introduction of Brazilian, Argentine and Japanese strains remains open. A previous study [14] showed a low probability of direct introduction of *BLV* strains from South America to Western Siberia due to the lack of both economic and zootechnical expediency of transporting livestock along such a long route. Also, there was no information about the import of cattle or bull semen from Japan in publicly available sources. Therefore, the similarity of Siberian isolates to Asian and South American strains can be explained by two hypotheses. Some assume that the similarity of viral strains can be explained by the accidental coincidence of stochastic mutations that independently occurred at the same loci in geographically isolated populations of viruses, resulting in the presence of the same nucleotides in similar positions [19]. R. Dawkins [28] believes that the probability of similarity of two nucleotide sequences due to random mutations is inversely dependent on the length of the compared fragments. However, others believe that the identity of nucleotide sequences can be considered a consequence of the coincidence of random mutations only if the possibility of direct gene flow is excluded by 100% [1, 24]. Studies of

satellite vertebrate DNA [29] have shown that allopatric isolation almost always leads to the accumulation of unique mutations in isolated populations. Despite the phenotypic similarity due to the convergence effect caused by the similarity of the orientation of selection vectors, the observed similarity is often the result of mutations at completely different loci [24]. In this case, it is well known that the gene pool of local livestock was repeatedly enriched, and therefore, the probability of isolation of West Siberian *BLV* strains by definition cannot be 100%. Given that any gene flow invariably creates an error in phylogenetic analysis [30], similarity of some of the *BLV* isolates from Novosibirsk region to Japanese and South American strains can be explained by divergence within the global *BLV* gene pool before their geographical separation with subsequent independent evolution. The identified similarities can be explained by independent penetration of ancestral forms of *AF399704* (Brazil) and *LC007988* (Japan) into Novosibirsk and corresponding regions, and the differences found are a consequence of the accumulation of random mutations [14].

This observation gives rise to an interesting hypothesis. It is known that any viral genome can be classified into conservative and variable components. The latter can also be divided into conditionally and truly variable parts. Presumably, the accumulation of differences between the Novosibirsk isolate (sample 2) and *AF399704* (Brazil) and *LC007988* (Japan) was due to mutagenesis of the truth of a variable fragment of the *BLV* genome. Conditionally, the variable part in both cases has been preserved in the form of the most ancient 'root' and least virulent sequences, which is quite logical considering the evolutionary vector of viruses for accelerated reproduction [31]. The assumption looks quite plausible, because it is confirmed by the presence of asymptomatic *BLV* carriers which results in hidden transmissions of the virus (30-70 % of cases) in cattle populations [32]. The more virulent the strain, the more likely is to provoke physiological reactions of the body, and therefore to be detected at the next veterinary blood test and subsequent culling together with the animal carrier. This artificial selection is aimed at eradicating highly virulent strains, and therefore, strains with virulence below a certain critical level are able to have a symbiotic relationship with cattle and probably account for the same 30-70% of cases of latent presence of *BLV* in the animal body.

It is known that bacterial and viral genomes contain so-called mutator genes [32], that aggravate the intensity of mutagenesis. In the case of *BLV*, the function of mutator genes is attributed to the LTR –region [33]. Apparently, mutations of the *LTR* region occurred in the evolution of *BLV*, which led to a split into two fundamentally different vectors, as indicated by some earlier studies [31]. The first of them is conservative, characteristic of the 1st genotype, which has preserved the maximum similarity with the original strain. The second branch is mutable and includes all other genotypes with the accumulation of corresponding mutations in the LTR-region. To confirm the hypothesis based on table 5 data, the average number of accumulated mutations for isolates of the 1st genotype samples № 2, 5-9, 14, 15, 16), as well as the 2nd and 4th genotypes (samples # 3, 4, 10-13) was calculate. Next, the average number of mutations in samples of the first genotype (174.11±13.57) was compared to that of isolates that do not belong to the first genotype (336.00±37.91). The obtained value of the student's criterion (4.02) at  $k=13$  corresponds to the reliability of differences  $P<0.01$ . The number of reverse mutations in samples 3 and 4 as well as 10-13 differed from 1% by a wide margin, whereas in the samples of the 1st genotype it were less or not significant from the random error. It turns out that the number of mutations in general and the reverse ones in particular in the evolutionary line of the 1st genotype were lower than in the others. Apparently, in the split-off evolutionary branch that gave rise to all other genotypes, active mutagenesis of SNPS took place SNP, which led to variability of the conditionally conservative part of the *BLV* genome, since most of it still remains practically unchanged.

## 5 Conclusion

The results of the studies showed that the *BLV* isolates isolated in the Novosibirsk region are a single set, with characteristic mutations characteristic of some other strains. Among them are Japanese and South American, despite the fact that the probability of direct introduction is almost zero. Most likely, such similarities are explained by the presence of mutations inherent in the original form of the virus, which were subsequently found in strains from different parts of the world. The hypothesis of randomly occurring identical mutations found in divergently evolving strains seems unlikely, since the probability has no significant differences from the random variable = 1%, despite a fairly small number of unique mutations.

## References

1. B Lewin. *Genes IX*. (Jones & Bartlett Learning, 2008)
2. EFSA AHAW Panel “Scientific opinion on enzootic bovine leukosis”, EFSA Journal **13**, 7: 4188 (2015) <https://doi.org/10.2903/j.efsa.2015.4188>
3. M.A. Juliarena, C.N. Barrios, C.M. Lützelshwab, E.N. Esteban, S.E. Gutiérrez. *Vir.Adapt. Treat.* **9**, 13–26 (2017) <https://doi.org/10.2147/VAAT.S113947>.
4. A. Pluta, M. Rola-Łuszczak, R.N. Douville, J. Kuźmak. *Virol J*, **15**, 165 (2018) <https://doi.org/10.1186/s12985-018-1062-z>.
5. M. Rola-Luszczak, A.Pluta, M. Olech, I. Donnik, M. Petropavlovskiy, A. Gerilovych, I. Vinogradova, B. Choudhury, J. Kuzmak, *PLoS One*, **8**, e58705 (2013) <https://doi.org/10.1371/journal.pone.0058705>
6. Y. Yang, P.J. Kelly, J. Bai, R. Zhang, C. Wang., *PLoS ONE*, **11**, 12, e0168379 (2016) <https://doi.org/10.1371/journal.pone.0168379>
7. Bioinformatics and DDBJ Center <https://www.ddbj.nig.ac.jp/index-e.html>
8. L.M. Mansky, L.C. Gajary, *Virology*, **301**, 272–280 (2002)<https://doi.org/10.1006/viro.2002.1578>
9. M.F. Camargos, A. Pereda, D. Stancek, M.A. Rocha, J.K. Reis, I. Greiser-Wilke, R.C. Leite, *Virus Genes*, **34**, 3, 343–350 (2007) <https://doi.org/10.1007/s11262-006-0011-x>
10. S. Gautam, N. Mishra, S. Kalaiyarasu, S.K. Jhade, R. Sood, *Transbound Emerg Dis*, **6**, 1968–1978 (2018) <https://doi.org/10.1111/tbed.12979>
11. M.A. Marawan, H. Mekata, T. Hayashi, S. Sekiguchi, Y. Kirino, A-M. Honri, M. Moustafa, F.K. Arnaout, E.S.M. Galila, J. Norimine, *Japan. J Vet Med Sci*, **79**(5), 912–916 (2017)<https://doi.org/10.1292/jvms.17-0055>
12. M. Polat, S. Takeshima, Y.Aida, *Virol J*, **14**, 209 (2017) <https://doi.org/10.1186/s12985-017-0876-4>.
13. S.M. Rodriguez, M.D. Golemba, R.H. Campos, K. Trono, L.R. Jones, *J Gen Virol*, **90**, 2788–2797 (2009) <https://doi.org/10.1099/vir.0.0111791-0>
14. N. Blazhko, S. Vyshegurov, A. Donchenko, K. Shatokhin, V. Ryabinina, K. Plotnikov, A. Khodakova, S. Pashkovskiy, *BMC Genet.*, **22**, 21(Suppl 1), 70 (2020) doi: 10.1186/s12863-020-00874-y.
15. M.N. Ruzina, B.V. Andrianov, T.M. Suprovich, G.E. Sulimova, *Genetika*, **49**, 8, 975–980 (2013)<https://doi.org/10.7868/s0016675813080134>

16. GOST 25382. Cattle, Methods of laboratory diagnostics of leucosis, 18(Moscow, Standart Publ, 1982) (in Russian) <http://docs.cntd.ru/document/gost-25382-82>
17. K Wilson, J Walker. *Principles and Techniques of Biochemistry and Molecular Biology*. (Seventh Sixth edition. Cambridge University Press, Cambridge, 2010). <https://doi.org/10.1002/bmb.20499>
18. A. Gonzalez-Voyer, A. von Hardenberg, *An Introduction to Phylogenetic Path Analysis*, Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology, Springer-Verlag Berlin Heidelberg, 201-229 (2014)[https://doi.org/10.1007/978-3-662-43550-2\\_8](https://doi.org/10.1007/978-3-662-43550-2_8).
19. M. Tanimura, K. Miyamura, N. Takeda. *Jpn. J. Genet.*, **60**, 137–150 (1985) <https://doi.org/10.1266/jjg.60.137>
20. S.A. Glantz. *Primer of Biostatistics, 5th ed.* (McGraw-Hill, New York, US, 2001)
21. T. Abdullah, M. Faiza, P. Pant, M.R. Akhtar, P. Pant., *Bioinformation* **12**, 3, 98–104 (2016)<https://doi.org/10.6026/97320630012098>
22. J.P. Bielawski, K.A. Dunn, Z. Yang, *Genetics*, **156**, 1299–1308 (2000)<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1461304/>
23. F. Rizzato, A. Rodriguez, X. Biarnés, A. Laio, *Genetics* **207**, 2, 643–652 (2017)<https://doi.org/10.1534/genetics.117.300078>
24. A. Markov, E. Naimark. *Evolution. Classical Ideas in the Light of New Discoveries*, (ACT CORPUS Publ., Moscow, Russia, 2015). (in Russian)
25. M.Yu. Shchelkanov, I.M. Kirillov, A.M. Shestopalov, K.E. Litvin, P.G. Deryabin, D.K. Lvov, *Voprosy Virusologii (Problems of Virology, Russian journal)*, **61**(6), 245–256 (2016) (in Russian) <http://dx.doi.org/10.18821/0507-4088-2016-61-6-245-256>
26. G. Fang, B. Weiser, C. Kuiken, S.M. Philpott, S. Rowland-Jones, F. Plummer, J. Kimani, B. Shi, R. Kaul, J. Bwayo J, Anzala O, Burger H, *Recombination following superinfection by HIV-1* **18**(2), 153–159(2004) [https://journals.lww.com/aidsonline/Fulltext/2004/01230/Recombination\\_following\\_superinfection\\_by\\_HIV\\_1.3.aspx](https://journals.lww.com/aidsonline/Fulltext/2004/01230/Recombination_following_superinfection_by_HIV_1.3.aspx)
27. B. Hukowska-Szematowicz, B. Tokarz-Deptuła, W. Deptuła, *J Appl Genet* **54**, 2, 235–248 (2013)<http://dx.doi.org/10.1007/s13353-013-0140-6>
28. R. Dawkins, *The Selfish Gene*, (Oxford University Press., 1978)
29. D. Gomez-Uchida, J.E. Seeb, C. Habicht, L.W. Seeb, *Can. J. Fish. Aquat. Sci.*, **69**, 916–929 (2012)<http://dx.doi.org/10.1139/F2012-029>
30. A. Som, *Briefings in Bioinformatics* **16**, 3, 536–548 (2015)<https://doi.org/10.1093/bib/bbu015>
31. N.V. Blazhko, S.Kh. Vyshegurov, A.S. Donchenko, K.S. Shatokhin, T.I. Krytsyna, V.A. Ryabinina, *Vavilov Journal of Genetics and Breeding*, **23**, 3, 262–269 (2019)<https://doi.org/10.18699/VJ19.491>
32. L.K. Ernst, G.E. Sulimova, A.R. Orlova, *Russ. J. Genet*, **33**, 1, 73–80 (1997)
33. V.K. Korboukh, C.A. Lee, A. Acevedo, M. Vignuzzi, Y. Xiao, J.J. Arnold, S. Hemperly, J.D. Graci, A. August, R. Andino, C.E. Cameron, *J. Biol. Chem* **289**, 43, 29531–29544 (2014) <https://doi.org/10.1074/jbc.M114.592303>
34. C. Merezak, C. Pierreux, E. Adam, F. Lemaigre, G.G. Rousseau, C. Calomnie, C. Van Lint, D. Christophe, P. Kerkhofs, A. Burny, R. Kettmann, L. Willems, *J. Virol*, **75**, 6977–6988 (2001) <https://doi.org/10.1128/JVI.75.15.6977-6988.2001>