The structure and content of the database on the nucleotide sequences of BLV isolates improved with the use of information technologies

N G Kozyreva
Federal State Budget Scientific Institution “Federal Scientific Centre VIEV” RAS (FSC VIEV), 24, bld. 1, Ryazansky avenue, Moscow, 109428, Russia

E-mail: nk07-73@mail.ru

Abstract. As a result of molecular genetic monitoring of bovine leukemia based on a set of methods (gene diagnostics - polymerase chain reaction (PCR), automatic sequencing; molecular phylogenetics) phylogenetic analysis presents the structure of an improved own replenished database (DB) for the characterized nucleotide sequences of the bovine leukemia virus (BLV), which includes information arrays contained in interconnected tables, which allows accumulating, storing, and combining all data through the formation of structured queries and reports with specified parameters in the database management system. At the same time, the results of the study are presented on the prevalence of BLV in the territories of some regions of Russia and Ukraine among the livestock; comparative analysis of genetic variants of the virus by genome loci: pol with clustering of the studied isolates into clade I; env with the identification of 5 genotypes (GI (4%), GII (6%), GIV (79%), GVII (10%), GVIII (1%)) with the dominance of genotype IV (“European cluster”). In the course of the study, a low level of genetic variability among field isolates of BLV was established for the pol - 1.9% and env - 0.2-1.8% gene loci (depending on the genotype). This information product is registered in the register of databases: certificate of state registration No 2020620530 dated 23.03.2020.

1. Introduction
An urgent stage in the development of research activities, in particular, in the field of molecular genetic diagnosis of infectious diseases, was the creation of personal computers and the subsequent widespread introduction into practice of computer information technologies and, in the future, network information technologies, which significantly changed the tools of the daily work of a researcher. By now, quite large volumes of experimental data have been accumulated, which are stored, among other things, in computer databases. Analysis and processing of these data is fundamentally impossible without the involvement of modern information technologies, effective methods of analysis and modeling of biological systems and processes.

The ability to obtain and analyze a large amount of genetic information quickly in the study of the spread of infectious diseases and the development of measures to prevent and limit them allows solving traditional problems of epizootology at the molecular level using the results of analysis of genetic polymorphism of pathogens.

BLV infects cattle all over the world, inducing enzootic leukemia in cattle; it also causes significant economic damage to the livestock industry. The causative agent belongs to the Retroviridae family, the
Deltaretrovirus genus, along with such representatives as T-cell lymphotropic viruses of primates (PTLV, STLV) and humans (HTLV) [1].

Monitoring of the spread, emergence of new genetic variants of the pathological agent of retroviral BLV infection, which is undoubtedly a reflection of its evolution, is necessary to improve diagnostics, determine the source of the pathogen during export-import operations, as well as to establish the relationship between infection with certain genetic variants of BLV and severity pathological process, development of clinical symptoms or humoral immune response in the host [2-5].

The family of these viruses is distinguished by the presence of a unique stage in the replicative cycle - reverse transcription, mediated by the enzyme reverse transcriptase - RNA-dependent DNA polymerase. The pol gene encodes this enzyme and is the most conserved region in the genome of retroviruses, including BLV [6-7]. It is known that phylogenetically BLV and primate T-lymphotropic viruses (PTLV) have a common ancestor, and when comparing the nucleotide sequence of different strains for the pol gene as a standard, a difference of 42% was revealed [8-10]. Thus, BLV forms a separate phylogenetic branch within the retrovirus family, which also differs from other retroviruses, for example, equine infectious anemia virus (EIAV) (~ 40%) [11]. Within the subgroup, on the one hand, BLV divergences in nucleotide sequences are less than 6% for the pol and env genes, which indicates a high degree of conservatism of virus strains from different geographic regions of the world [9, 12-15]. On the other hand, the env gene is a variable / variable region of the BLV genome - most studies on BLV polymorphism and the classification of its genetic variants are based on determining the primary sequence of the env gene with subsequent phylogenetic analysis, and today at least 11 BLV genotypes with different geographic localization [13, 16-20]. At the same time, some researchers distinguish various subgroups within the G4, G6 and G7 genotypes [14, 21-22].

The env gene encodes the surface gp51 and transmembrane gp30 glycoproteins, which are targets for neutralizing antibodies and play a key role in the life cycle of the virus, being responsible for cell tropism, virus penetration, and are involved in the induction of cell fusion and the formation of syncytia. Violation of replication in vivo and an increase in the rate of development of the disease can be caused by a mutation affecting only one codon of the env gene, which, in turn, may affect the development of means of specific prevention of the disease (serological, molecular diagnostics, vaccine preparations) and the peculiarities of the biology of this pathogen. Thus, based on bioinformatic analysis of 256 nucleotide sequences (>500 bp) of the env gene from the GenBank database, Polish authors identified determinants / cell epitopes of B cells and T cells and compared these sequences, which will undoubtedly be useful. in the development of an effective vaccine against BLV and elucidation of the mechanisms of the phenomenon of "escape" from the immune system of the host organism during a viral strategy [23]. In this analysis, we used, inter alia, 27 Russian isolates obtained at the laboratory of leukemia VIEW under the following registration numbers (GenBank No.): JN695878-82; JQ675756-60; JQ686089-98, JQ686106-107, JQ686111-112, JQ686116-120 from Moscow, Kaluga, Rostov, Novosibirsk, Vologda regions of the Russian Federation [24].

Such topical aspects of the study of BLV have scientific and practical significance. The discovery of closely related variants of the virus in infected individuals indicates that these animals were infected from the same source. Thus, the high variability of viruses makes it possible to establish an epizootic relationship between donors and recipients of the virus based on the results of phylogenetic analysis: the closer the evolutionary relationship between the variants of the virus, the, in general, the closer the epizootic relationship between the individuals from which they were obtained.

Since 2008, on the basis of the VIEW, a sufficient amount of experimental data has been accumulated on the study of BLV polymorphism in some regions of the Russian Federation, stored on Microsoft (MS) Excel platforms in separate tables, bioinformatics programs. Thus, based on this material, modern information technologies (IT) and effective methods of analysis were involved in order to create a new information product - the database "Characterization of nucleotide sequences of genomes of bovine leukemia virus isolates" in MS Access format as an IT tool used in the system of epizootic monitoring in the planning and implementation of anti-leukemic health measures.
2. Materials and research methods

In the course of PCR with electrophoretic detection of amplification products, 438 bp amplicons were obtained (fragment of the pol gene) [25], 341 bp (fragment of the env gene) [26]. The obtained data were processed using the algorithms of the following computer programs.

The work with genetic information was carried out using the BioEdit v.7 software [27], algorithms of the NCBI service (USA) [28]: for comparison, we used the nucleotide sequences of the target genes of some reference BLV isolates from the international primary data bank GenBank, the BLAST program (basic local alignment search tool, Altschul et al., 1990-2000), the deposition of genetic sequences was carried out using the Sequin or BankIt programs [29];

Phylogenetic analysis was performed using the Mega program (molecular evolutionary genetics analysis, Kimura et al., 1993-2012), while: multiple alignment of fragments was performed for the corresponding loci, sizes: pol - about 400 nucleotides, env - about 300 nucleotides;

The construction of dendrograms was carried out by remote methods of minimum evolution (ME, minimum evolution) [30-32], Neighbor joining (NJ) [33] with the definition of p-distances; by the method of analysis of discrete features of maximum likelihood (ML, maximum likelihood) [30, 34-35] (data not shown);

The statistical significance/reliability of the tree topology was assessed using the bootstrap analysis method (Zharkikh and Li, 1992) with 1000 random samples; evolutionary distances (within (intragroup) and between (intergroup) groups) were calculated using the Jukes-Cantor model (Jukes-Cantor model) [36], Kimura (two-parameter) (Kimura model) [37], Tajima and Nei model [38].

BLV genotypes were differentiated according to the classification proposed by the authors of the source [16] with subsequent additions.

When working in MS Access, the following database management system components (DBMS) were used: constructors of tables, screen forms, SQL queries (Structured Query Language), reports, printed out; as well as functions: related queries, links with external tables [39]).

3. Research results and discussion

Evolutionary studies in the case of molecular epizootology include not only experimental data on genetic information, but also information on the characteristics of the epizootic process. As a result of processing, analyzing, and converting such an array of data into one format, a single electronic platform is required for the effective use of material and database management.

The developed database included information arrays based on: the results of sequencing of the BLV pol and env gene loci (directly decoded nucleotide sequences - sequences), on animals from which isolates were isolated, depositing genetic information in an international database, developing own or using the recommended oligonucleotide primers.

In the process of work, the data array was formed from the initial moment, the logical compilation and filling of the main tables, the compilation of the structural-logical diagram, the analysis of the data described in the subject area and the creation of a database on the MS Access platform within the framework of the relational model [40].

During the implementation of logical design, a structural-logical matrix was developed, in the process of which the selection of tables that were to be included in the database, the columns belonging to each table, the relationship between tables and columns was made. The physical design created the table structures including field names, data types, field sizes, etc. Then the corresponding information was added, which was ordered in a certain way within a given structure and is represented by two blocks of tables: "Nucleotide sequences of the BLV proviral genome" and "Characteristics of oligonucleotide primers", interconnected by foreign keys - unique codes ("ID") in the form of a block -schemes (figure 1). Each table has a unique name. Moreover, if the data changes in one table, then they are automatically changed in all tables associated with the first. Also, if there are key fields in the tables, duplicate records are excluded. Data integrity was ensured by the algorithms "Cascade update" of related fields and "Cascade delete" of related records.
Figure 1. Block diagram: structural-logical matrix and diagram of links/interconnections of structural blocks (framed) of database tables.

To work with information stored in tables: sorting, filtering, performing calculations, grouping and modifying data, we used queries created in the "Query Designer" mode in the standard information access language - SQL, which allows building sentences in different constructions.

Also, when exporting the created SQL queries to MS Excel, statistical data processing was performed.

When implementing an SQL query to combine records from different tables, for example, "Characterization of oligonucleotide primers" and "Characterization of isolate sequence", data on the primers used to determine the primary nucleotide sequence of a particular isolate or group of BLV isolates were obtained.

The database array is represented by 81 and 192 nucleotide sequences of the pol and env gene loci, respectively, with a unique name assigned to each isolate; 20 sequences of oligonucleotide primers used to obtain these fragments of the BLV genome. In addition, the information contained in the columns is included with the following names: source of isolation, year of collection of the biomaterial, isolation of the isolate, geographic affiliation.

In separate tables "Popularization of information", the presence of data on each isolate in the publication and reporting materials was recorded; "Information on the method of application" included data on the method used, the right to use a specific technique, links to sources; "Depositing into international databases" was entered information on 116 (to date) registered in the international GenBank data bank of the characterized nucleotide sequences of BLV isolates, for which individual passports for sequencing (IPS) were also developed.

When implementing a SQL query to combine records from tables, for example, "Characteristics of oligonucleotide primers" and "Characteristics of an isolate sequence", data were obtained on primers, for example, used either only when determining the primary nucleotide sequence of a particular isolate or group of BLV isolates, or only when carrying out a reaction amplification or for both procedures.

To navigate between records, a special form was developed - a customizable dialog box - as a database object that allows you to add, edit, and display stored data. As a basis for this main form (Figure 2), the corresponding base tables "Nucleotide sequences of the BLV proviral genome" and "Characteristics of oligonucleotide primers" were used. After creating the form directly in the "Design" mode, then the "Button" tool was used to implement and activate direct transitions through the records.
Figure 2. Main navigation form.

By establishing a link between the tables "Geographic localization of isolate" and "Adm2", data on the distribution of BLV genotypes at the zoning level under the control of MS Access DBMS were included at the level of a certain layer in the geographic information system - GIS cattle leukemia on the ArcGIS software platform when compiling the corresponding thematic GIS map further. In this case, the layer is based on actual information from the database and will not be displayed on the map if there is no access to the data source. In turn, using the ArcGIS format, it is possible to productively use geographic information as an informative resource when conducting analysis, adapting the data obtained in the process of making effective decisions.

In total, the geography of distribution of the BLV included 6 federal districts of the Russian Federation: Central, Volga, Southern, Northern, Northwestern, Far Eastern, Southern, including 14 regions: Ryazan, Tula, Moscow, Kaluga, Yaroslavl, Smolensk, Ulyanovsk, Nizhny Novgorod, Penza, Rostov, Vologda regions, the Republic of Kalmykia, Altai, Kamchatka regions; and 5 regions of Ukraine, including Donetsk, Volyn, Kiev, Zaporozyhe, Rivne regions.

Figure 3. Distribution structure of BLV genotypes circulating in some regions of Russia (A) and Ukraine (B), (%).

The homogeneity was assessed at the pol gene locus: all the characterized BLV isolates belong to clade 1 (Figure 4), which includes strains from Argentina, Australia, Japan, USA, Brazil and differ from international strains from the USA (M10987, BLV GAGA), Northern Italy (S83529), assigned to clade 2, as well as the USA (NC001414, AF033818) - clade 3; and heterogeneity at the locus of the env gene: 5 - GI, GII, GIV (dominant), GVII, GVIII - were identified among the studied population of BLV out of 11 genotypes currently existing all over the world (Figure 5).

The study of BLV polymorphism based on phylogenetic analysis at the env gene locus revealed the heterogeneity of the pathogen population. The dominance of the GIV genotype (the so-called "European" subgroup) of the virus was demonstrated, which confirms the long-term circulation of the BLV variant imported from European countries with infected cattle after World War II as a result of
international trade operations carried out with various breeds of cattle in including within the framework of the Council for Mutual Economic Assistance during the Soviet Union.

At the same time, it was found that the BLV populations in the surveyed territories of 11 regions of the Russian Federation and 4 regions of Ukraine are not homogeneous. So, in the territories of the subjects of the Russian Federation: Penza (20.4% isolates), Ulyanovsk (18.2% isolates) - GIV, GVII; Ryazan (20.4% isolates) - GII, GIV; Rostov (13.3% isolates) - GI, GII, GIV; Novosibirsk (7.2% isolates) - GI, GIV, GVII; Moscow (7.2% isolates) - GI, GIV, GVII, GVIII - regions circulate more than one (from 2 to 4) pathogen genotypes, in contrast to other regions in whose territories separate genetic variants of the virus are found: Kamchatka Territory (0.6% isolates) - GVII, regions - Smolensk (5.5% isolates) - GIV, Yaroslavl (2.2% isolates) - GI, Kaluga (2.2% isolates) - GIV, Vologda (2.8% isolates) - GVII. On the territories of some regions of Ukraine, monopopulations of VLBV were revealed: in Donetsk (25% isolates) - GI, Volyn (25% isolates), Kiev (12.5% isolates) - GIV, Zaporozhye (37.5% isolates) - GII (figure 6).

![Figure 4](image-url)  
**Figure 4.** Phylogenetic comparison of the pol gene regions of representatives of BLV proviruses, including some representatives of BLV from the VIEV DB.
Where: A - main tree (built using the remote method of joining neighbors, p-distances, bootstrap analysis with 1000 random samples; the data array consisted of 92 sequences); B - a branch of the tree with representatives of proviruses of clade 1. Symbols indicate ▲ some of the characterized isolates of the VIEV DB, ▀ KK isolate FLK-BLV (K +).

Figure 5. Phylogenetic comparison of the env gene regions of representatives of BLV proviruses, including some representatives from the VIEV database.
Where: The tree was built using the remote minimum evolution method, evolutionary distances were calculated using the p-distance method, with bootstrap analysis for 1000 random samples; the data array was 206 sequences. Symbols denote: ▲, Δ some of the characterized isolates of VIEV DB, KK isolate FLK-BLV (K+).

The presence of a low level of genetic variability among the field isolates of BLV BD by intragroup distances for genes pol - 1.9% and env - 0.2-1.8% (depending on the genotype) was shown, which correlates with the results of world studies in this direction [13, 15, 21]. Genetic differences between the BLV variants circulating in different populations, present in the analyzed target sites of the pathogen genome as genetic markers of this infection, can have a potential effect on the results of diagnostic tests, which undoubtedly must be taken into account when developing diagnostic and prophylactic drugs.

The uneven distribution of genotypes may be the result of the multiple effects of pathogen selection in the adaptation of the original sequence (wild type). Once in a susceptible population of cattle, a specific epizootic network, a specific variant of the virus will quickly spread in it. Genetic heterogeneity can also be explained by independent infections from different epizootic/unsuccessful foci, for example, during export-import operations, and analysis of the phylogenetic relationships between potential donors and recipients of the virus makes it possible to establish the source of infection. Large-scale dissemination of variants of the virus within and between distant geographic regions is possibly initiated by the movement of the population of donor animals associated with both the migration of humans and livestock, as well as viral transmission between individual animals during long close contact [16]. The presence of both homogeneous and heterogeneous populations of FLV is observed in certain geographic regions of the world [41-43].

When using the developed database, the choice of the most conservative fragment of the pol gene, the variability of which on average does not exceed 2% when analyzing the sequences of the studied locus, made it possible to design a PCR test system with various formats of detection of amplification products and use it for diagnosing BLV infection in large and small ruminants of cattle [44-47]; Analysis of the genetic diversity of BLV made it possible to find a correlation between the results of phylogenetic analysis based on different genetic targets - pol, env and to make an assumption about the source of infection in a number of cases of export-import transactions, which entails the identification of one of the biological risks, namely, that the import livestock from disadvantaged regions leads to a worsening of the epizootic situation for leukemia [48-49].

4. Conclusion
The current database "Characterization of nucleotide sequences of genomes of bovine leukemia virus isolates" has been improved on the MS Access database platform. This information product is registered in the register of databases - Certificate of state registration No. 2020620530 dated 23.03.2020. An improved database is an independent object for storing and analyzing a large amount of information.

The characterized molecular genetic data are used to study the genetic polymorphism of BLV within the molecular epizootology of the pathogen.
As an electronic repository in the structure of a thematic geoinformation system, the database acts as a tool used in antiepizootic/recreational activities within the framework of this retroviral infection in order to increase the effectiveness of monitoring biological risks.

Acknowledgements
The work was carried out within the framework of a state assignment in accordance with the approved research plan of the FSC VIEW RAS for 2019-2021.

References
[1] Polat M, Takeshima S and Aida Y 2017 Epidemiology and genetic diversity of bovine leukemia virus J. Virol. 14 209
[2] Inoue E, Matsumura K, Maekawa K, Nagatsu K, Nobuta M, Hirata M, Minagawa A, Osawa Y and Okazaki K 2011 Genetic heterogeneity among bovine leukemia viruses in Japan and their relationship to leukemogenicity Arch. Virol. 156 1137-41
[3] Matsumura K, Inoue E, Osawa Y and Okazaki K 2011 Molecular epidemiology of bovine leukemia virus associated with enzootic bovine leukosis in Japan Virus Res. 155 343-8
[4] Panei C J, Takeshima S N, Omori T, Nunoya T, Davis W C, Ishizaki H, Matoba K and Aida Y 2013 Estimation of bovine leukemia virus (BLV) proviral load harbored by lymphocyte subpopulations in BLV-infected cattle at the subclinical stage of enzootic bovine leukosis using BLV-CoCoMo-qPCR BMC Vet. Res. 9(95)
[5] Aida Y, Murakami H, Takahashi M and Takeshima S N 2013 Mechanisms of pathogenesis induced by bovine leukemia virus as a model for human T-cell leukemia virus Front. Microbiol 4 328
[6] Gerald G and Grangenet D 1980 Retrovirus reverse transcriptase Molecular Biology of RNA tumor viruses ed J Stephenson (New York: Academic Press) 346-94
[7] Kettmann R and Burny A 1994 Bovine Leukemia Virus The Retroviridae 3 (New York: Plenum Press) 39-81
[8] Dube S et al 2000 The complete genomic sequence of a BLV strain from a Holstein cow from Argentina Virology 277 379-86
[9] Gillet N et al. 2007 Mechanisms of leukemogenesis induced by bovine leukemia virus: prospects for novel anti-retroviral therapies in human Retrovirology 4 18
[10] Dube S et al. 2009 The complete genomic sequence of an in vivo low replicating BLV strain J. Virol 6 11
[11] Willems L et al. 2000 Genetic determinants of bovine leukemia virus pathogenesis AIDS Res Hum Retroviruses 16(16) 1787-95
[12] Kozyreva N G, Gulyukin M I, Ivanova L A, Malogolovkin A S, Klimenko A I and Razumovskaya V V 2011 Phylogenetic analysis of a section of the provirus pol gene of bovine leukemia virus isolates from animals in various regions of the Russian Federation Russ. Agricult. Sci. 37 520-3
[13] Polat M et al. 2016 A new genotype of bovine leukemia virus in South America identified by NGS-based whole genome sequencing and molecular evolutionary genetic analysis Retrovirology 13(4) 23
[14] Pluta A, Rola-Łuszczak M, Kubis P, Balov S, Moskalik R, Choudhury B and Kuz´mak J 2017 Molecular characterization of bovine leukemia virus from Moldovan dairy cattle Arch. Virol. 162(6) 1563-76
[15] Bazzucchi M, Iscaro C, Cascieri C, Giammarioli M and Feliziani F 2019 Molecular characterization of Italian bovine leukemia virus isolates reveals the presence of distinct phylogenetic clusters Arch Virol. 164 1697-703
[16] Rodriguez S M, Golemba M D, Campos R H, Trono K and Jones L R 2009 Bovine leukemia virus can be classified into seven genotypes: evidence for the existence of two novel clades J. Gen. Virol. 90 2788-97
[17] Balic D et al. 2012 Identification of a new genotype of bovine leukemia virus Arch Virol. 157 1281-90
[18] Shaeva A Yu, Garayaeva Z R, Vafin R R, Khazipov N Z and Alimov A M 2012 Identification of a new genotype BLV Uchenye zapiski KGAVM 211 192-7
[19] Lee E, Kim E J, Rathnanohart J, Vitoonpong R, Kim B H, Cho I S, Song J Y, Lee K K and Shin Y K 2016 Molecular epidemiological and serological studies of bovine leukemia virus (BLV) infection in Thailand cattle Infect Genet. Evol. 41 245-54
[20] Yu C, Wang X, Zhou Y, Wang Y, Zhang X and Zheng Y 2019 Genotyping bovine leukemia virus in dairy cattle of Heilongjiang, northeastern China BMC Vet Res. 15(1) 179
[21] Rola-Luszczak M, Pluta A, Olech M, Donnik I, Petropavlovskiy M, Gerilovych A, Vinogradova I, Choudhury B and Kuzmaz J 2013 The molecular characterization of bovine leukaemia virus isolates from Eastern Europe and Siberia and its impact on phylogeny PLoS One 8(3)
[22] Gautam S, Mishra N, Kalaiyarasu S, Jhade S K and Sood R 2018 Molecular characterization of bovine leukaemia virus (BLV) strains reveals existence of genotype 6 in cattle in India with evidence of a new subgenotype Transbound Emerging Diseases 65(6) 1968-78
[23] Pluta A, Albritton L M, Rola-Luszczak M and Kuzmak J 2018 Computational analysis of envelope glycoproteins from diverse geographical isolates of bovine leukaemia virus identifies highly conserved peptide motifs Retrovirology 15(1) 2
[24] Lomakina N F, Gulyukin M I, Ivanova L A, Klimenko A I and Kovalenko A V 2010 Phylogenetic analysis of bovine leukaemia viruses circulating in the Moscow Kaluga and Rostov regions Int. Scientific and practical conf Modern problems of diagnosis, treatment and prevention of infectious diseases of animals and birds: collection of articles. scientific. tr. leading scientists of Russia and abroad 3 (Yekaterinburg: Ural Publ.) p 161
[25] Gulyukin M I, Kozyreva N G, Ivanova L A, Kolbasov D V, Tsybanov S Zh, Kalabekov I M and Malogolovkin A S 2011 Methodical guidelines for the detection of DNA of bovine leukaemia provirus by PCR method and sample preparation of the obtained genetic material for the purpose of further determination of the nucleotide sequence by sequencing Modern means and methods of ensuring veterinary welfare in infectious and protozoal pathology of animals, fish and bees ed Gulyukin M I et al. (Moscow: Agency for Creative Technologies Publ.) p 356
[26] Belak S and Ballagi–Pordány A 1993 Experiences on the application of the polymerase chain reaction in a diagnostic laboratory Mol. Cell. Probes 7 241-48
[27] Hall T A 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT Nucleic Acids Symposium Series 41 95-8
[28] National Center for Biotechnology Information (NCBI) [Electronic resource]. U.S. National Library of Medicine 8600 Rockville Pike Bethesda MD 20894 1998
[29] Benson D A, Karsch-Mizrachi I, Lipman D J, Ostell J, Rapp B A and Wheeler D L 2000 GenBank Nucleic Acids Res. 28(1) 15-8
[30] Cavalli-Sforza L L and Edwards A W F 1967 Phylogenetic analysis: models and estimation procedures Evolution international journal of organic evolution 21(3) 550-70
[31] Saitou N and Imanishi T 1989 Relative efficiencies of the Fitch-Margoliash, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree Mol. Biol. Evol. 6(5) 514-25
[32] Rzhetsky A and Nei M 1993 Theoretical foundation of the minimum-evolution method of phylogenetic inference Mol. Biol. Evol. 10(5) 1073-95
[33] Saitou N and Nei M 1987 The neighbor-joining method: a new method for reconstruction of phylogenetic trees Mol. Biol. Evol. 4(4) 406-25
[34] Felsenstein J 1973 Maximum-likelihood estimation of evolutionary trees from continuous characters Am. J. Hum. Genet. 25(5) 471-92
[35] Felsenstein J 1981 Evolutionary trees from DNA sequences, a maximum likelihood approach J. of Molecular Evolution 17 368-76
[36] Jukes T H and Cantor C R 1969 Evolution of protein molecules *Mammalian Protein Metabolism* ed H N Munro (New York, USA: Academic Press) p 21-132

[37] Kimura M 1980 A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences *Journal of Molecular Evolution* 16 111-20

[38] Tajima F and Nei M. 1984 Estimation of evolutionary distance between nucleotide sequences *Molecular Biology and Evolution* 1(3) 269-85

[39] Mak-Donal'd M 2007 *The missing manual* (St. Petersburg: BKhV-Peterburg Publ.) p 784

[40] Codd E F 1970 A Relational Model of Data for Large Shared Data Banks *Communications of the ACM: journal.* 13(6) 377-87

[41] Zhao X and Buehring G C 2007 Natural genetic variations in bovine leukemia virus envelope gene: possible effects of selection and escape *Virology* 366 150-65

[42] Camargos M F, Pereda A, Stancek D, Rocha M A, dos Reis J K, Greiser-Wilke I and Leite R C 2007 Molecular characterization of the env gene from Brazilian field isolates of bovine leukemia virus *Virus Genes* 34 343-50

[43] Polat M, Ohno A, Takeshima S N, Kim J, Kikuya M, Matsumoto Y, Mingala C N, Onuma M and Aida Y 2015 Detection and molecular characterization of bovine leukemia virus in Philippine cattle *Arch Virol.* 160 285-96

[44] Kozyreva N G 2012 To study the sensitivity and specificity of the FRT PCR method for detecting DNA of the bovine leukemia provirus in biological material *Veterinary pathology* 2(40) 123-6

[45] Kozyreva N G, Guljukin M I, Ivanova L A, Kolbasov D V, Tsybanov S Z, Kalabekov I M and Malogolovkin A 2012 *Diagnostic technique for cattle leukaemia by polymerase chain reaction* 2010 RU 2445370 C1, 20.03.2012, application: 2010143998/10, 28.10.2010, effective date for property rights: 28.10.2012, date of publication: 20.03.2012 in Russian

[46] Kozyreva N G, Ivanova L A, Stepanova T V and Guljukin M I 2019 *Method for diagnosis of cattle leukaemia by polymerase chain reaction* 2018 RU 2694617 C1, 16.07.2019, application: 2018116549, 04.05.2018, effective date for property rights: 04.05.2018, date of publication: 16.07.2019 in Russian

[47] Kozyreva N G and Guljukin M I 2017 The use of multiplex PCR-RT in molecular diagnostics of BLV at perinatal infection *Mezhdunarodnyy vestnik veterinarii* 4 28-32

[48] Gulyukin M I, Ivanova L A, Gendzhieva O B and Kozyreva N G 2012 Genotyping of BLV isolates spread at the territory of the republic of Kalmykia *Veterinariya Kubani* 4 4-7

[49] Kozyreva N G and Gulyukin M I 2017 Distribution of cattle leukemia and genetic variants of the pathogen on the territory of livestock farms in the Central Federal District of the Russian Federation *Veterinariya Kubani* 6 4-9