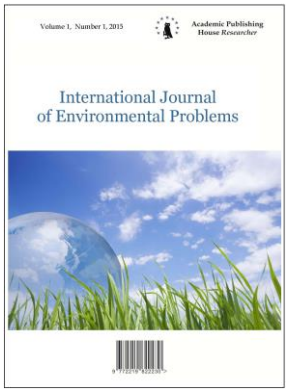
Published in the Russian Federation
International Journal of Environmental Problems
Has been issued since 2015.
ISSN: 2410-9339
E-ISSN: 2413-7561
Vol. 4, Is. 2, pp. 89-98, 2016

DOI: 10.13187/ijep.2016.4.89
www.ejournal33.com



Articles and Statements

UDC 636.2:636.01:575.174.015.3:577.2.08:51-76

Distribution of Potential Non-Canonical DNA Motif in Proviral DNA Genes of the Avian Influenza and Bovine Leukemia Retroviruses

Valery I. Glazko [a, *], Gleb Yu. Kosovsky [b], Tatina T. Glazko [b]

[a] Centre of Experimental Embryology and Reproductive Biotechnologies RAAS, Moscow, Russian Federation
[b] Russian State Agrarian University – MAA named after K.A. Timiriazev, Moscow, Russian Federation

Abstract

A comparative analysis of distributional density of the DNA regions potentially predisposed to quadruplex formation was conducted on proviral DNA encoding hemagglutenin (HA) and neuroaminidase (NA) of the avian influenza virus, and also on the env and pol genes of bovine leukemia virus. We discovered that increased genetic heterogeneity of the HA gene coincides with the high frequency of appearance of nucleotide sequences predisposed to G4 quadruplex formation compared to the NA gene. Similar tendency was observed during the comparison of distribution of potential G4 quadruplexes in the env and pol genes of the bovine leukemia virus. Data obtained allowed to suppose that non-overlapping sequences with relatively heightened predisposition to G-quadruplex formation can reflect, to a large extend, recombination events. And variable sequences can reflect the nucleotide context that contributes to appearance of new mutations. In the present article we discuss possible connection between increased density of non-canonical nucleotide structures in genes, products of which interact with receptors of target cells, and their genetic heterogeneity supported in the host-pathogen system.

Keywords: avian influenza virus, bovine leukemia virus, hemagglutenin, neuroaminidase, reverse transcriptase, RT-PCR, G4 quadruplexes.

1. Introduction

Unification and globalization of farm animals' fodder and genetic material, unstable region borders of new pathogenic isolates leads to a necessity to develop new methods for identification and prognosis of appearance of the new variants of pathogens. These problems gained extreme importance over the past several years (Babii et al., 2015; Cheshko et al., 2015; Glazko, 2014; Glazko et al., 2015; Glazko V., Glazko T., 2015; Glazko V., Glazko T., 2016; Glazko et al., 2016), as well as the concern on the quality of environment (Kalinichenko, 2015). Thus, ILRT studies suggest

* Corresponding author
E-mail address: vigvalery@gmail.com (V.I. Glazko)

that about 2,3 billion people are infected with pathogens carried by animals every year and about 1,7 million of them die (http://www.healthmap.org, 2016).

As a rule, such infectious agents include retroviruses, in particular the avian influenza virus. A distinctive feature of retroviral evolution is the fact that accumulation rate of new genetic variants in host's organism is always slower then that between the hosts (Morelli et al., 2013). It was suggested that the mean frequency of spontaneous mutations in retroviruses can reach values in the range of 10^{-3}-10^{-5} to a single nucleotide to a single transcriptional cycle (Duffy et al., 2008). High rate of the retroviral evolution requires development of prognostic methods to predict appearance of mutations in the regions directly associated with viral patogenicity. One of the promising directions lies within study of the distributional density and the regions of localization of nucleotide sequences potentially predisposed to higher frequency of mutations in retroviral genomes. At present, there are data indicating that the ability to overcome barriers to cross-species transmission and appearance of viral patogenicity of the avian influenza virus in humans is the result of mutations, which lead to amino acid substitutions in a distinctive domain of hemagglutinine. The resulting protein participates in the interaction with lymphocyte receptors (Xiong et al., 2013).

Highly pathogenic avian influenza (HPAI) virus that belongs to A/H5N1 subtype was initially discovered in geese in the Chinese province of Guangdong in 1996 (Choi et al., 2005). Since its' first appearance, outbreaks of H5N1 were discovered among avian populations of the Eastern Asia, Western Africa and even England. H5N1 is primarily an infectious agent of birds. Due to the wide spread of the infection more than 200 million birds died or were destroyed worldwide (Fouchier et al., 2005; World Health Organization, 2008). The origin of the H5N1 expansion in many countries is the population of domestic birds. At present, the virus is often able to overcome cross-species barriers (Choi et al., 2005).

The major antigene determinants of the type A avian influenza virus are hemagglutinin (HA) and neuroaminidase (NA). These are transmembrane glycoproteins capable of initiating a full subtypical specific immunological reaction. Vietnam is a country where incidents of the highly pathogenic avian influenza virus H5N1 infections are most common. The virus was identified for the first time in 2001 (Choi et al., 2005). Since December 2003 outbreaks of infection were registered among domestic birds in more than 59 provinces out of 64. The first case of infection in a human was registered in Vietnam in 2004. By August 2012 there were already 123 cases, 61 of which were lethal according to WHO (Wan et al., 2008).

During the last few years, special attention was devoted to investigation of the mechanisms and molecular and genetic basis of the spreading of different retroviral infection (Bovine leukaemia virus, BLV) among farm animals. The relevance of such research is determined not only by economical losses in diary production (about 525 million dollars a year), but also by the fact that in a large number of studies infectivity and genetic heterogeniety of the different regions of the BLV genome is viewed as an experimental model, which allows investigation of genetic conditions of spreading mechanisms of such retroviruses as, for example, HTLV (Human T-cell Lymphotropic Viruses) (Gillet et al., 2007; Hajj et al., 2012).

Bovine leucosis virus (BLV) infectivity and its' ability to neutralize immunological respones are connected with the envelope glycoproten gp51 and the transmembrane glycoprotein gp30 encoded in the sequences of the *env* gene (Rola-Łuszczak et al., 2013). The ability of those coding sequences to accumulate non-synonymous nucleotide substitutions, which affect immunogenecity and infectivity characteristics of BLV, at fast rate is shown in experiments *in vitro* (Johnston et al., 2002).

In Russia identification of the viral carrier state for a number of retroviruses of the farm animals in most cases is conducted using methods of evaluation of presence of the antibodies against envelop proteins of retroviruses in the blood samples. However high speed of mutations in *env* gene, dependence of antibody genesis on the age of the animal and other characteristic of animal immune system, problems with identification of antibodies in sperm products during AI significantly lower reliability of immunological methods for pathogen control. Existing detection methods based on diagnostics of the proviral DNA integration into host's genome are complicated by the high variability of proviral genes, genomic overlode with endogenous retroviruses and their debries in mammals. The complication exists, because those methods are based on the design of

primers for the regions of proviral DNA with understudied mutability and potential predisposition to cross hybridisation with the dispersed repeats of host's genome.

According to literature, there is growing evidence that the higher predesposition to spontaneous mutations of a certain DNA region is the result of increased localization density of non-canonical DNA structures such as G-quadruplexes. G-quadruplexes are four stranded structures, in which guanines form a plane with their transversal bonds and obtain increased heat stability (Collie, Parkinson, 2011). It can be expected that the discovery of such regions in the genes, products of which play a significant role in retroviral pathogenicity can facilitate future development of the methods to predict their variability.

To verify possible connection between spontaneous polymorphism of the retroviral structural genes and their participation in the formation of retroviral infectivity, analysis of distribution of the nucleotide motifs potentially predisposed to G-quadruplex formation (G-quadruplex sequences) was made using cDNA RNA of such genes as hemoagglutenin and neroaminidase of the avian influenza virus and the env and pol genes (reverse transcriptase) of the bovine leucosis virus.

2. Materials and methods

The sequencing results of the genome segments of the hemagglutenin gene and neuroaminidase genes (HA and NA) from 24 isolates of the avian influenza virus obtained from poultry populations in different regions of Vietnam were published earlier (Toan et al., 2013). The segments were compared against sequences of type A H5N1 virus published in GenBank database and accessible for analysis. The analysis included sequences of the HA segment (about 1700 bp) and the NA segment (about 616 bp).

The BLV *env* gene (1548 bp) and *pol* (2538 bp) sequences were obtained from the database presented at GenBank (AF547184.2 and D00647.1). A comparative analysis of non-synonymous to synonymous substitutions ratios (Kn/s ratio) was done for protein-coding gene sequences of the env and pol genes. In the gene env nucleotide substitutions were calculated separately for different functionally significant regions of this gene. These regions encode glycoprotein gp51, which participates in the direct contact of viral particles with receptors of target cells, and glycoprotein gp30, which forms transmembrane domain. Sequences published in the GenBank database were used. Sequence alignment was done manually using BioEdit software and CLUSTAL W software. Amino acid sequences were obtained using software program “Mega”.

Localization of G-quadruplexes in the sequenced fragments was evaluated using QGRS Mapper program (Quadruplex forming G-Rich Sequences). Sequence potentially predisposed to quadruplex formation represents following structure: G2-5NL1, G2-5NL2, G2-5NL3, where NL1-3 are loops of various length and nucleotide composition.

Homology search in archive of repeats was conducted using RepeatMsker and Giri (http://www.repeatmasker.org/; http://www.girinst.org/censor/).

3. Results and discussion

Sequencing of the HA gene segment from the avian influenza virus isolates allowed us to discover high variability in this segment compared to the sequence of the HA gene A/goose/Vietnam/3/05 from the H5N1 strain published earlier. In general, polymorphism was observed in the regions between nucleotides number 406 and 476, number 639 and 720, number 935 and 996, number 1385 and 1442. All of those regions encode the HA protein (Toan et al., 2013). A total of 74 amino acid substitutions were discovered. Nucleotide substitutions were also found in segments of the NA gene during comparison of the sequenced DNA regions with the published sequence of the NA gene from A/HongKong/483/1997 (H5N1) strain. However such substitutions were found only between nucleotides number 445 and 756 in the NA gene. And there were significantly fewer substitutions (30) compared to the HA gene. The data obtained is the evidence of relatively increased genetic heterogeniety of sequences regions of the HA compared to the NA gene (Toan et al., 2013).

A comparative analysis of the nucleotide sequences potentially predispose to G-quadruplex formation was conducted in order to validate a possible connection between observed differences and distinctive features of the structural organization of these genes. Higher density of such sequences is observed in the DNA double-strand break “hotspots” (DSB). Localization of G-

quadruplex structures in sequenced fragments of the genome was evaluated using QGRS (Quadruplex forming G-Rich Sequences) Mapper.

The following data were obtained as the result of our study. Seven non-overlapping potential G-quadruplexe structures (Fig. 1) and 138 overlapping sequences potentially predisposed to G-quadruplex formation were observed in the hemagglutenin gene fragment 1779 bp in length. In general these seqences were concentrated in two regions of the gene. The first segment was 42 bp in length (nucleotides number 1076 to 1117), where 117 potential G-quadruplex sequences were discovered. The second segment was 13 bp in length (nucleotides 1258 to 1270), where eleven potential G-quadruplex seqences were discovered (Fig. 1, 2). It is important to point out that this segment in particular contains nucleotide triplets coding for amino acids 342-346 of the hemagglutenin. These amino acids form domain of the hemagglutenin, which is directly connected with the virulence of the avian flu virus type A (Horimoto, Kawaoka, 1994).

There were significantly fewer sequences of that kind in the examined fragment of the NA gene. Only four non-overlapping sequences potentially predisposed to G-quadruplex formation and only eight overlapping sequences were discovered in the region 1401 bp in length. Three out of eight of the overlapping sequences were located between nucleotides number 543 and 564 (Fig. 3).

No homologous sequences with repeats found in RepeatMasker and Giri archives were discovered in the examined regions of the genes HA and NA of the avian flu virus.

Then we conducted a distribution density analysis of the sequences potentially predisposed to G-quadruplex formation for the BLV genes *env* and *pol*, because these genes also differ in terms of genetic heterogeneity (Rola-Łuszczak et al., 2013).

As it turned out, four non-overlapping sequences potentially capable of G-quadruplex formation and 105 overlapping sequences were discovered in the region of the *env* gene 1548 bp in length. The major part of the sequences was localized in three segments of this gene: between nucleotides number 272 and 296 (34 variants), between nucleotides number 1080 and 1109 (19 variants) and between nucleotides number 1080 and 1272 (69 variants).

The nucleotide sequence *env* codes for a signal peptide that consists of 33 amino acids (nucleotides number 1 to 99), for the envelope glycoprotein gp51, which consists of 301 amino acids and starts with tryptophan (nucleotides number 100 to 901), and the transmembrane glycoproten gp30, which contains amino acids 302 to 515 and starts with serine (nucleotides number 904 to 1545).

According to data obtained by Moratorio et al. (2013), who were the first to describe in significant detail all amino acid sequences of the BLV proteins, it is possible to see that 34 overlapping variants of potential G-quadruplexes located between nucleotides number 272 and 296 (amino acids 91 to 99) coincide with the region of the conformational epitope.

It is known that the antibodies against epitopes F,G and H of the gp51 glycoprotein possess neutralizing activity. No BLV strain was found, in which antigenes F, G and H were absent at the same time. This implies that these epitopes are connected with viral infectivity (Moratorio et al., 2013). Rabbit antibodies against peptides 9–48, 78–92, 144–157 and 177–192 neutralize VSV/BLV pseudotypes *in vitro*, which is also a sign of their connection with viral infectivity (Johnston et al., 2002).

According to the literature (Johnston et al., 2002), amino acids number 144 to 157 correspond to a region of the HTLV-1 envelope glycopeptide that also participates in antibody neutralization. Cell fusions into syncytium are inhibited by antibodies against peptides 64-73, 98-117 and 177-192 (Johnston et al., 2002). The last sequence (especially amino acids P177 and D178) is an epitope of a T-helper and stimulates proliferation of lymphocytes obtained from infected cows. CD8-dependent cytotoxic activity is associated with peptides 121-140, 131-150, or 24-31. Modeling of protein folding shows that the receptor binding to the target cells is formed by the most effective neutralizing epitopes (Johnston et al., 2002).

The remaining two segments with high density of overlapping sequences potentially predisposed to G-quadurplex formation are localized in the transmembrane glycoprotein gp30 (amino acids 360 to 424), which coincides with important domains of this protien (from (Johnston et al., 2002)). The transmembrane (TM) protein gp30 plays a key role during cell fusion, since it destabilizes cytoplasmic membrane of the target cells due to it's ability to be inserted into the lipid bilayer. Two domains of the *env* protein 19-27 and 39-103 are also necessary for effective cell fusion (Johnston et al., 2002).

Region 104-123 interacts with zink and significantly influences infectability of the virus *in vivo* (Johnston et al., 2002). Besides having a role in cell fusion, TM protein also participates in the signaling pathway through a motif located on the cytoplasmic end of tyrosine-based immunoreceptor (immunoreceptor tyrosine-based activation, ITAM). The key sequence in ITAM is YXXL motif, where X represents variable residue (Johnston et al., 2002).

Ten non-overlapping sequences potentially predisposed to G-quadruplex formation and 49 overlapping sequences were found in the *pol* gene 2538 bp in length (Fig. 4). This is significantly lower than the density of distribution of these elements in the *env* gene.

We also conducted comparative analysis of the distribution of single nucleotide substitutions (polymorphism) for different regions of the BLV *env* and *pol* genes in the porviral DNA and (comparative) analysis of a ratio of non-synonymous to synonymous substitution frequencies (Kn/s ratio) for coding sequences of those genes. Following data were obtained.

In the poviral DNA of the *env* gene the frequency of nucleotide substitutions was 0.1766 in sequences encoding gp51 (Kn/s was 0.5955) and 0,1636 in the sequences encoding gp30 (Kn/s was 0,3636). Calculations were based on the analysis of 52 sequences presented in GenBank database.

The data obtained signify that the frequency of non-synonymous substitutions in the genomic sequence encoding gp51 is significantly higher compared to the gp30 sequence (0.5955 vs 0.3636), even though there is a similarity between mean frequencies of nucleotide substitutions in these sequences summarily. This difference suggests that there exists more intense pressure of the "purifying" selection on the gp30 sequence compared to the gp51 sequence. It can be expected that this higher pressure is caused by the fact of the gp30 glycoprotein belonging to a transmembrane protein family.

Ratios of nucleotide substitutions to a single nucleotide in three regions of the *pol* gene between nucleotides number 2325 and 2818 of the proviral genome (29 sequences), between nucleotides 3986-4218 (35 sequences), and between nucleotides 4499-4631 (33 sequences) were close to each other: 0.1417, 0.1552, and 0.1654, respectively. The only exception was the fragment between nucleotides number 4257 and 4393 (38 sequences), where the mean level of substitution was relatively high (0.2774). The lowest value of Kn/s was observed in 3986-4218 regions (0.2414). Ratios close to the one discovered in the gp30 gene were found in two regions of the *pol* gene between nucleotides 2325 and 2818 and nucleotides 4499 and 4631. Kn/s were equal to 0.4000 and 0.4667, respectively.

The highest values of Kn/s (0.9000) were observed in the *pol* gene between nucleotides number 4257 and 4393. There was also the highest ratio of the mean frequency of nucleotide substitutions to a single nucleotide. A close to one Kn/s ratio in this region can indicate absence of intensive influence of the"purifying" selection, therefore the probability of mistakes becomes higher, when this region is used for diagnostics of proviral DNA integration into host genome.

A relatively high frequency of nucleotide substitutions to a single nucleotide (0.6667) with relatively high value of Kn/s (0.7143) is observed in regions with increased density of potential G-quadruplex structures. Two non-overlapping sequences potentially predisposed to G4-quadruplex formation were localized in the *pol* gene segment between nucleotides number 4256 and 42780.

The data obtained can serve as evidence for existence of significant differences in ratios of non-synonymous to synonymous substitutions in the examined genes of porviral DNA, which apparently reflects different intensities of the "purifying" selection. A certain association of relatively high frequency of non-synonymous nucleotide substitutions and increased localization density of nucleotide sequences predisposed to G4-quadruplex formation (for example, fragments between nuclotides 272 and 296, amino acids 91-99 of the *env* gene) were also discovered in separate regions of the examined genes.

In addition, a search for homology to the BLF *env* and *pol* genes was conducted using Repeat Masker and Giri software (http://www.repeatmasker.org/, http://www.girinst.org/censor/).

It was discovered that there were no homology regions in the BLV *env* gene as well, as in the HA and NA genes of the avian flu virus. However, such regions were present in the BLV *pol* gene in order that can be seen in table 1.

Homology regions are distributed within the *pol* gene in the following order: between nucleotides 20–486 – ERV3-2_CJa-I (66 % of homology, transposable element endogenous retroviruses from the common marmose) (Jurka, 2008); between nucleotides 1693 – 1755 DIRS-14_DR (70 % of homology, DIRS-type LTR retrotransposon from zebrafish) (Jurka, 2011a);

between nucleotides 1833 – 2156 ERV2-4_CPo-I (64 % of homology, Endogenous retrovirus from guinea pig) (Kojima, Jurka, 2010); between nucleotides 2481–2522 – DNA-2-26_DR (79 % of homology, DNA transposon from zebrafish) (Jurka, 2011b) (Fig. 5).

As it can be seen on the Fig. 5, three out of four homology regions are located in the *pol* gene sequences (nucleotides 1781 to 2381), where non-overlapping sequences potentially predisposed to G-quadruplex formation were found.

Our findings suggest that non-overlapping sequences with relatively heightened predisposition to G-quadruplex formation can reflect, to a large extend, recombination events. And variable sequences can reflect the nucleotide context that contributes to appearance of new mutations.

It can be expected that discovered differences in the distribution of the overlapping nucleotide sequences predisposed to non-canonical structure formation such as G-quadruplexes have certain connection to spontaneous mutation spectra in genes studied in our work.

The data obtained suggest that a relatively high genetic heterogeneity of the env gene compared to the pol gene in BLV isolates, to a certain extent, is caused by different density of non-canonical DNA structures in the proviral DNA. A higher density of such structures in genes, products of which directly interact with receptors of target cells, can support their positive selection by participation in generation of genetic heterogeneity necessary for competitive relations in the host-pathogen system.

4. Conclusion

A comparative analysis of the distribution of nucleotide sequences predisposed to non-canonical structure formation such as G-quadruplexes was conducted on the hemagglutinin and neuraminidase genes of the avian flu virus and on the *env* and *pol* genes of BLV. These sequences mark unstable genomic regions. Occurrence frequency of such fragments is relatively high in the hemaglutenin gene compared to the neuraminidase gene of the avian flu virus and in the *env* gene compared to the *pol* gene in BLV. It is reasonable to expect that the well-known heterogeneity of retroviral genes, products of which interact directly with receptor cell system of the host, to a certain degree, is the result of a higher density of localization of non-canonical stuctures mentioned above.

Non-canonical DNA structures play a key role in the infectability of retroviruses. Presence of such structures in retroviral genome fragments allows us to consider possibility of new approaches to the detection of retroviruses and reduction of their patogenicity.

Our findings indicate that there are substantial differences in the nucleotide substitution frequencies and Kn/s ratios in the coding sequences of the *pol* and *env* genes. These differences should be considered in the process of selection of DNA fragments for identification of BLV proviral sequences in the bovine genome, thus giving an opportunity to create a universal test-system.

References

Babii et al., 2015 – *Babii A. V., Kovalchuk S.N., Glazko T.T., Kosovskii G.Yu., Glazko V.I.* (2015). Multilocus Genotyping of Cattle Using Highly Polymorphic Genomic Elements (Microsatelites, DNA Transposon Helitron). *Biogeosystem Technique,* Vol. (5), Is. 3, pp. 243–255. doi:10.13187/bgt.2015.5.243

Cheshko et al., 2015 – *Cheshko V.T., Kosova Y.V., Glazko V.I.* (2015). Evolutionary Semantics of Anthropogenesis and Bioethics of Nbic-Technologies. *Biogeosystem Technique,* Vol.(5), Is. 3, pp. 256–266, doi**:** 10.13187/bgt.2015.5.256

Choi et al., 2005 – *Choi Young Ki , Tien Dzung Nguyen, Hiroichi Ozaki, Richard J. Webby, Pilaipan Puthavathana, Chantanee Buranathal, Arunee Chaisingh, Prasert Auewarakul, N.T.H. Hanh, Sia Kit Ma, Pui Yan Hui, Yi Guan, Joseph Sriyal Malik Peiris, Robert G. Webster* (2005). Studies of H5N1 Influenza Virus Infection of Pigs by Using Viruses Isolated in Vietnam and Thailand in 2004 // *J. Virol.* August Vol. 79, No. 16 10821–10825, doi:10.1128/JVI.79.16.10821-10825.2005

Collie, Parkinson, 2011 – *Collie G.W., Parkinson G.N.* (2011). The application of DNA and RNA G-quadruplexes to therapeutic medicines // *Chem Soc Rev*. 2011 Dec;40(12):5867–92, doi:10.1039/c1cs15067g

Duffy et al., 2008 – *Duffy S, Shackelton LA, Holmes EC* (2008) Rates of evolutionary change in viruses: patterns and determinants. // *Nat Rev Genet*. 2008 Apr;9(4):267–76. doi: 10.1038/nrg2323

Fouchier et al., 2005 – *Fouchier R.A., Munster V., Wallensten A., Bestebroer T.M., Herfst S., Smith D., Rimmelzwaan G.F., Olsen B., Osterhaus A.D.* (2005). Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls // *J Virol.* Mar;79(5):2814–22, doi:10.1128/JVI.79.5.2814-2822.2005

Gillet et al., 2007 – *Gillet N., Florins A., Boxus M., Burteau C., Nigro A., Vandermeers F., Balon H., Bouzar A.-B., Defoiche J., Burny A., Reichert M., Kettmann R., Willems L.* (2007). Mechanisms of leukemogenesis induced by bovine leukemia virus: prospects for novel anti-retroviral therapies in human // *Retrovirology*, 4:18 doi:10.1186/1742-4690-4-18.

Glazko, 2014 – *Glazko V.I.* (2014). Genomics and Geobiosystems // *Biogeosystem Technique,* Vol. (2), No 2, pp. 125–132. doi: 10.13187/bgt.2014.2.125

Glazko et al., 2015 – *Glazko V.I., Elkina M.A., Glazko T.T.* (2015). Genomic Scanning Using the Retrotransposon Fragments as "Anchors", in Animals and Plants // *Biogeosystem Technique,* Vol.(6), Is. 4, pp. 363–373, doi: 10.13187/bgt.2015.6.363

Glazko V., Glazko T., 2015 – *Glazko V.I.,. Glazko T.T* (2015). Reproductive “Success” and Population-Genetic Consequences of Environmental Stress on the Example of Accidents in Chernobyl and Fukushima // *Biogeosystem Technique,* Vol.(6), Is. 4, pp. 316–326, doi: 10.13187/bgt.2015.6.316

Glazko V., Glazko T., 2016 – *Glazko V.I., Glazko T.T.* (2016). Biosocial Consequences of Industrial Accidents and Social Crises (Chernobyl, Fukushima) // *Biogeosystem Technique,* Vol. (7), Is. 1, pp. 4–16. doi:10.13187/bgt.2016.7.4

Glazko et al., 2016 – *Glazko V.I., Zybaylov B.L., Glazko T.T.* (2016). Improving ionizing radiation, "horizontal" and "vertical" biological and biosocial consequences (the case of the Chernobyl accident and Fukushima-1 nuclear power plant) // *Agricultural Biology*, Vol. 51, № 2, pp. 141–155. doi: 10.15389/agrobiology.2016.2.141rus

Hajj et al., 2012 – *Hajj E.I., Nasr R., Kfoury Y., Dassouki Z., Nasser R., Kchour G., Hermine O., de Thii H., Bazarbachi A.* (2012). Animal models on HTLV-land related viruses: what did we learn? // *Frontiers in Microbiology. Virology*, Vol. 3, doi:10.3389/fmicb.2012.00333.

Horimoto, Kawaoka, 1994 – Horimoto T., Kawaoka Y. (1994). Reverse genetics provides direct evidence for a correlation of hemagglutinin cleavability and virulence of an avian influenza A virus // *J Virol,* May;68(5):3120-8, pp. 3120–3128, PMCID:PMC236802

Mapping of poverty and likely zoonoses hotspots Zoonoses, 2016 – *Mapping of poverty and likely zoonoses hotspots Zoonoses* (2016). International Livestock Research Institute (ILRI)Project 4 Report to Department for International Development, UK / URL: http://r4d.dfid.gov.uk/pdf/outputs/livestock/ZooMapDFIDreport18June2012FINALsm.pdf http://www.healthmap.org/ru/

Johnston et al., 2002 – *Johnston E.R., Albritton L.M., Radke K.* (2002). Envelope Proteins Containing Single Amino Acid Substitutions Support a Structural Model of the Receptor-Binding Domain of Bovine Leukemia Virus Surface Protein // *J. Virology*, Nov;76(21):10861–72, PMCID: PMC136609.

Jurka, 2008 – *Jurka J.* (2008). DNA transposons from zebrafish. // *Repbase Reports*, Vol. 8, No 12, pp. 2135–2135.

Jurka, 2011a – *Jurka J.* (2011). Endogenous retroviruses from guinea pig // *Repbase Reports,* Vol. 11, No 5. pp. 1495–1495.

Jurka, 2011b – *Jurka J.* (2011). Endogenous retroviruses from the common marmoset.//*Repbase Reports*, Vol. 11, No 2, pp. 692–692.

Kalinichenko, 2015 – *Kalinichenko V.P.* (2015). Biogeosystem Technique as a Paradigm of Non-waste Technology in the Biosphere // *Biogeosystem Technique,* Vol. (3), Is. 1, pp. 4–28. DOI: 10.13187/bgt.2015.3.4

Kojima, Jurka, 2010 – *Kojima K.K., Jurka J.* (2010). DIRS-type LTR retrotransposons from zebrafish // *Repbase Reports,* Vol. 10, No 11. pp. 1842–1842.

Moratorio et al., 2013 – *Moratorio G., Fischer S., Bianchi S., Tomŭ L., Rama G., Obal G., Carriyn F., Pritsch O., Cristina J.* (2013). A detailed molecular analysis of complete Bovine Leukemia Virus genomes isolated from B-cell lymphosarcomas.// *Vet Res,* Vol. 44, No 1:19. doi:10.1186/1297-9716-44-19

Morelli et al., 2013 – *Morelli M.J., Wright C.F., Knowles N.J., Juleff N., Paton D.J., King D.P., Haydon D.T.* (2013) Evolution of foot-and-mouth disease virus intra-sample sequence diversity during serial transmission in bovine hosts // Vet Res., Mar 1;44:12, doi:10.1186/1297-9716-44-12

Rola-Łuszczak et al., 2013 – *Rola-Łuszczak M., Pluta A., Olech M, Donnik I., Petropavlovskiy M., Gerilovych A., et al.* (2013) The Molecular Characterization of Bovine Leukaemia Virus Isolates from Eastern Europe and Siberia and Its Impact on Phylogeny // PLoS ONE 8(3): e58705. doi:10.1371/journal.pone.0058705

World Health Organization, 2008 – *World Health Organization* (2008). World Organisation for Animal Health / Food and Agriculture Organization H5N1 Evolution Working Group. Towards a unified nomenclature system for highly pathogenic avian influenza virus (H5N1) [conference summary] / URL: http://www.cdc.gov/EID/content/14/7/e1.htm.

Xiong, 2013 – *Xiong X., Coombs P.J., Martin S.R., Liu J., Xiao H., McCauley J.W., Locher K., Walker P.A., Collins P.J., Kawaoka Y., Skehel J.J., Gamblin S.J.* (2013) Receptor binding by a ferret-transmissible H5 avian influenza virus // *Nature,* Vol. 497, No 7449. pp. 392–396. doi:10.1038/nature12144.

Wan et al., 2008 – *Wan Xiu-Feng, Tung Nguyen, C. Todd Davis, Catherine B. Smith, Zi-Ming Zhao* (2008) Evolution of Highly Pathogenic H5N1 Avian Influenza Viruses in Vietnam between 2001 and 2007 // *PLoS ONE,* Vol. 3, No 10: e3462, doi:10.1371/journal.pone.0003462

Position	Length	Potential regions of G4 quadruplex appearing	G-Score
235	30	**GG**ATTGTAGTGTAGCT**GG**AT**GG**CTCCTC**GG**	9
298	26	**GG**AAT**GG**TCTTACATAGT**GG**AGAA**GG**	13
738	29	**GG**CAAAGT**GG**AAGAAT**GG**AGTTCTTCT**GG**	18
1099	14	**GG**GA**GG**AT**GG**CA**GG**	21
1117	18	**GG**TAGAT**GG**TT**GG**TAT**GG**	18
1258	29	**GG**CCGTT**GG**AAG**GG**AATTTAATAACTT**GG**	11
1447	27	**GG**ATAATGCAAA**GG**AGCT**GG**GTAAT**GG**	15

Fig. 1. Non-overlapping genomic regions of the hemagglutinin gene potentially predisposed to the formation of G4-quadruplexes

000001 AGCAAAAGCA GGGGTATGAT CTGTCAAAAT GGAGAAAATA GTGCTTCTTC
TTGCAACAGT CAGTCTTGTT AAAAGTGACC AGATTTGCAT TGGTTACCAT
000101 GCAAACAACT CGACAGAGCA GGTTGACACA ATAATGGAAA AGAATGTTAC
TGTTACACAT GCCCAAGACA TACTGGAAAG GACACACAAC GGGAAGCTCT
000201 GCGATCTAAA TGGAGTGAAG CCTCTGATTT TGAG**GG**ATTG TAGTGTAGCT
GGAT**GG**CTCC TC**GG**AAACCC TATGTGTGAC GAATTCATCA ATGTGCC**GG**A
000301 AT**GG**TCTTAC ATAGT**GG**AGA A**GG**CCAGTCC AGCCAATGAC CTCTGTTATC
CAGGGAATTT CAACGACTAT GAAGAACTGA AACACCTATT GAGCAGAATA
000401 AACCATTTTG AGAAAATTCA GATCATCCCC AAAAGTTCTT GGTCCAATCA
TGATGCCTCA TCAGGGGTAA GCTCAGCATG TCCATACCTT GGGAGGTCCT
000501 CCTTTTTCAG AAATGTGGTA TGGCTTATCA AAAAGAACAG TACATACCCA
ACAATAAAGA GGAGCTACAA TAATACCAAC CAAGAAGATC TTTTGGTACT
000601 GTGGGGGATT CACCATCCTA ATGATGCGGC AGAGCAGACA AAGCTCTATC
AAAACCCAAC CACCTACATT TCCGTTGGAA CATCAACACT GAACCAGAGA

000701 TTGGTTCCAG AAATAGCTAC TAGACCCAAA GTAAACGGGC AAAGTGGAAG
AATGGAGTTC TTCTGGACAA TTTTAAAGCC GAATGATGCC ATCAATTTCG
000801 AGAGTAATGG AAATTTCATT GCTCCAGAAT ATGCATACAA AATTGTCAAG
AAAGGGGACT CAACAATTAT GAAAAGTGAA TTGGAATATG GTAACTGCAA
000901 CACCAAGTGT CAAACTCCAA TGGGGGCGAT AAACTCTAGT ATGCCATTCC
ACAACATACA CCCCCTCACC ATCGGGGAAT GCCCCAAATA TGTGAAATCA
001001 AACAGATTAG TCCTTGCGAC TGGACTCAGA AATGCCCCTC AAAGAGAGAG
AAGAAGAAAA AAGAGAGGAC TATTTGGAGC TATAGCAGGT TTTATAGAGG
001101 GAGGATGGCA GGGAATGGTA GATGGTTGGT ATGGGTACCA CCATAGCAAT
GAGCAGGGGA GTGGATACGC TGCAGACCAA GAATCCACTC AAAAGGCAAT
001201 AGATGGAGTC ACCAATAAGG TCAACTCGAT CATTGACAAA ATGAACACTC
AGTTTGAGGC CGTTGGAAGG GAATTTAATA ACTTGGAAAG GAGGATAGAG
001301 AATTTAAACA AGAAGATGGA AGACGGATTC CTAGATGTCT GGACTTACAA
TGCCGAACTT CTGGTTCTCA TGGAAAATGA GAGAACTCTA GACTTTCATG
001401 ACTCAAATGT CAAGAACCTT TACGACAAGG TCCGACTACA GCTTAGGGAT
AATGCAAAGG AGCTGGGTAA TGGTTGTTTC GAATTCTATC ACAAATGTGA
001501 TAACGAATGT ATGGAAAGTG TAAAAAACGG AACGTATGAC TACCCGCAGT
ATTCAGAAGA AGCAAGACTA AACAGAGAGG AAATAAGTGG AGTAAAATTG
001601 GAATCAATGG GAACTTACCA AATACTGTCA ATTTATTCAA CAGTGGCGAG
TTCCCTAGCA CTGGCAATCA TGGTAGCTGG TCTATCTTTA TGGATGTGCT
001701 CCAATGGATC GTTACAATGC AGAATTTGCA TTTAAATTTG TGAGTTCAGA
TTGTAGTTAA AAACACCCTT GTTTCTACT

Fig. 2. The nucleotide sequence of the hemagglutinin gene (HA, DQ366314). The large letters indicate the translation initiation triplet. Yellow highlights indicate non-overlapping sequences where G4-quadruplexes are formed

Position	Length	Potential regions of G4 quadruplex appearing	G-Score
543	25	GGAATTTCCGGTCCAGATAATGGGG	11
906	26	GGATATATATGCAGTGGGGTTTTCGG	8
1023	21	GGCAATGGTGTTTGGATCGGG	20
1322	21	GGGTTGGTCTTGGCCAGACGG	18

Fig. 3. Non-overlapping regions of the neuroaminidase gene, potentially predisposed to the formation of G4-quadruplexes

Position	Length	Potential regions of G4 quadruplex appearing	G-Score
52	29	GGCTATATCTCCCCCTGGGACGGGCCAGG	10
99	28	GGTACGGAAACCAAATGGCACCTGGAGG	15
540	20	GGACCTAGGGTTTCAGGTGG	16
690	25	GGCGGTCTTGGGAGACCTCCAGTGG	11
1303	17	GGGGCTACAGGACGAGG	16
1390	26	GGAGAACTAGCAGGTCTCTTGGCGGG	13
1781	23	GGCTCAATCCAACTGGAGGTGGG	10
1826	27	GGTGGGCCCCGAATCATATTTGGCAGG	8
2125	23	GGGTTGGTAGAGCGGACAAATGG	18
2381	21	GGTGGCTAGGACCTCTCCCGG	13

Fig. 4. Non-overlapping regions of the BLV *pol* gene potentially preisposed to the formation of G4-quadruplexes

Fig. 5. Destribution of regions homologous to the zebrafish transposon, the guinea pig transposon and the transposon of common marmoset within the *env* gene of BLV

Note: ERV3-2_CJa-I fragment, positions of the pol gene 209-486 (66 % of homology); DIRS-14_DR fragment, positions of the pol gene 1693-2156 (70 % of homology); ERV2-4_CPo-I fragment, positions of the pol gene 1833-2156 (64 % of homology); DNA-2-26_DR fragment, positions of the pol gene 2481-2522 (70 % of homology).

Table 1. The homology regions found using Giri software in nucleotide sequences of the BLV pol gene

Type of the dispersed repeat	Fragments	Length in b.p.
Transposable elements	4	707
DNA transposon	1	42
Endogenous retroviruses	2	602
ERV2	1	324
ERV3	1	278
LTR retrotransposon	1	63
DIRS	1	63
Total	4	707