AGRICULTURAL BIOLOGY,
2017, V. 52, № 4, pp. 658-668
(SEL'SKOKHOZYAISTVENNAYA BIOLOGIYA)

ISSN 2412-0324 (English ed. Online)
ISSN 0131-6397 (Russian ed. Print)
ISSN 2313-4836 (Russian ed. Online)

Genome structure and genetic diversity

UDC 636.2:577.212.3

doi: 10.15389/agrobiology.2017.4.658rus
doi: 10.15389/agrobiology.2017.4.658eng

DOMAIN DISTRIBUTION OF MOBILE GENETIC ELEMENTS IN THE BOVINE GENOME

V.I. GLAZKO[1, 2], O.I. SKOBEL[1], G.Yu. KOSOVSKY[1], T.T. GLAZKO[1, 2]

[1]*Center for Experimental Embryology and Reproductive Biotechnology,* Federal Agency of Scientific Organizations, 12/4, ul. Kostyakova, Moscow, 127422 Russia, e-mail vigvalery@gmail.com (corresponding author), skobelolga@gmail.com, gkosovsky@mail.ru, tglazko@rambler.ru;
[2]*K.A. Timiryazev Russian State Agrarian University—Moscow Agrarian Academy,* 49, ul. Timiryazevskaya, Moscow, 127550 Russia
ORCID:
Glazko V.I. orcid.org/0000-0002-8566-8717
Skobel O.I. orcid.org/0000-0002-0599-9787
Kosovsky G.Yu. orcid.org/0000-0003-3808-3086
Glazko T.T. orcid.org/0000-0002-3879-6935
The authors declare no conflict of interests
Received December 5, 2016

Abstract

Genetic landscape of bovine genome attracts a lot of attention in recent years. This is due to the complexity of genomic selection task solution, i.e. the use of multilocus genotypes in order to simplify and hasten breeding. Accumulated data show that there is high evolutionary speed of different genetic elements and also they have structure functional polymorphism intensity (L. Chen et al., 2017). It was shown that interspersed repeats account for about 50 % of nucleotide sequence of the bovine genome (R.L. Tellam et al., 2009). Also it was found that some of the interspersed repeats cluster into conservative domains along the bovine genome due to joint localization (D.L. Adelson et al., 2009). The characteristics of domain distribution are still not fully studied despite the fact that it is very important to identify conservative and variable domains throughout the bovine genome to solve traditional tasks of their genetic resources management and controlling. In this work domain distribution of mobile genetic elements and their products of recombination in nucleotide sequences of 13,436,028 nucleotides of bovine chromosome 1 were analyzed by means of Repeat Masker mobile genetic elements database and Integrated Genome Browser software. It was revealed that the most prevalent types throughout analyzed region are SINE/tRNA-Core-RTE, LINE/RTE-BovB, LINE/L1 and LTR/ERV. Their joint localization in bovine genome has complicated structure. The most common pairwise clusters are SINE and LINE, SINE/tRNA-Core-RTE and LTR EVR, (LTR/ERVK)/(LINE/RTE-BovB), (LTR/ERVK)/(LINE/L1). Two last pairs are the bases for such triple clusters as (LINE/RTE-BovB)/(BTLTR1)/(LINE/RTE-BovB) and (LINE/L1)/(BTLTR1J)/(LINE/L1). It should be mentioned that there is no such clustering with other retrotransposons. It was revealed that there is some certain bias of these triple clusters high density to the distal end of studied region of chromosome 1. By the means of Integrated Genome Browser software the localization of obtained triple products of recombination between LINE and LTR ERV to structural genes was analyzed. It was found that only 34 clusters are localized in 12 structural genes (other are located in intergenic space). Besides, 10 and 12 copies are located in two genes that are closely connected with the function of central nervous system in mammals, *grik1* and *app*. The fact that 9 copies of triple gene construct (LINE/RTE-BovB)/(BTLTR1)/(LINE/RTE-BovB) are found in each of two genes and (LINE/L1)/(BTLTR1J)/(LINE/L1) had only 1 copy in *grik1* and 3 copies in *app*, suggests that these genes are ancestral targets for such insertions and their conservations. It also should be mentioned that (LINE/L1)/(BTLTR1J)/(LINE/L1) construct was found only in these two genes but not in other 10 genes where (LINE/RTE-BovB)/(BTLTR1)/(LINE/RTE-BovB) is also located. Specific features of distribution of products of recombination between LINE and LTR ERV throughout the studied chromosome 1 area and their localization in structural genes suggest the possible presence of structure functional elements there. Revealing of such elements is the subject of our further study.

Keywords: mobile genetic elements, retrotransposons, DNA transposons, products of recombination, domain distribution, genomic landscape, cattle

Observations of mobile genetic elements (MGE) in the genomes of mammals have a rather long history. The largest part (from 40 to 50 % of the total length) in the mammalian genomes is occupied by interspersed repeats. In most

mammalian species, including bovine cattle, retrotransposons (RTs) dominate in the genomes among mobile genetic elements, which for their movement use the mechanisms of reproduction of exogenous retroviruses [1]. One of the reptile units as *Squamata* could be the source of a number of mobile genetic elements for ruminants [2-4]. It is assumed that the horizontal transfer of RTs between taxonomically removed forms is promoted by the common habitat [5].

Many of the identified RTs were common to all mammals, which probably indicates the ancientry of their origin. Most RTs have lost their activity, but the increased polymorphism of some RTs may reflect their relatively recent origin and involvement in genomic reorganization processes, usually of functional and evolutionary significance [6]. Thus, insertions of some RTs into the promoter regions of structural genes substantially change the expression of the latter, and in the amino acid sequence they lead to the occurring of new proteins [7]. In bovine cattle, a number of structural gene mutations associated with RT insertions lead to lethal effects in the homozygous state [8].

In mammals, the main RT is a Long Interspersed Nuclear Element 1 (LINE1). They carry active and other LINE in the genome, such as the LINE RTE family. Non-autonomous Short Interspersed Nuclear Element (SINE) for transpositions requires LINE. Primates for transposition of SINE Alu require LINE L1. The ancient RT LINE2 (L2) encodes the proteins necessary for the distribution of the SINE MIR, widely represented in the genomes of mammals. In ruminants and marsupials, LINE RTE encodes proteins required for transpositions of SINE BovA, the (BOV-A2, Bov-tA1, 2, 3)/SINE ART2A or SINE RTE, respectively. RTE LINE contains BovB repeats, suggesting that they are horizontally transferred from reptiles to ruminants and marsupials [4, 9-11]. Ancient clusters of L2/MIR repeats form domains that are conserved in human and bovine genomes, and there are no younger repetitions, such as RTE/ART2A, in such domains. Since the ancient repetitions are clustered into evolutionarily conserved domains, this allows us to assume that there are special mechanisms providing such conservativeness that can be associated with blocks of expressed genes, with the specificity of localization in the interphase nucleus space, differences in the methylation pattern, or features of genomic sites different in protection from re-integration of retrotransposons [2]. A large amount of data has been accumulated, indicating that retrotranspositions in mammals are involved in the occurring of new genes and functional evolution [12], as well as in gene duplications [13]. Endogenous retroviruses (ERV) containing long terminal repeats (LTR) are another variant of RTs which is widespread in mammalian genomes (in particular, in bovine cattle) and characterized by high polymorphism; interbreed differences are described by the presence of some of them [14].

Detection of the conserved and variable domains of RT localization is of significant practical importance for the selection of the most polymorphic genomic elements suitable for use as anchors in genomic scanning (poly-locus genotyping) in controlling the genetic structure and its dynamics at the species and intra-specific differentiation level [15]. These studies acquired particular significance after the revealing zonality of the different RTs distribution and the short fragment copy number variability (CNV) in the cells of the germinal series and somatic cells, identified in human genomes [16].

Despite the importance of studying conserved and variable domains in the genomes of farm animals for solving traditional problems of control and management of genetic resources, such works still remain rare. Moreover, as a rule, these papers consider colocalization of full-size RTs, while in genomes there is a huge number of RT fragments marking the sites of transpositions and recombinations between them [17, 18].

In the present study, the patterns of colocalization and clustering of homology sections to the most frequently encountered retrotransposons of different families and their recombination products (binomial and trinomial associations) in chromosome 1, the longest autosome of bovine cattle, are revealed.

The aim of the work was to analyze the distribution and positioning of mobile genetic elements to clarify the possible patterns of their structural organization in the genome.

Techniques. Information on the genomic location of mobile genetic elements with the nucleotide coordinates in bovine cattle given in the RepeatMasker software [1] was used as the initial data; the archive bosTau7.fa.out.gz [20] created using the RepeatMasker version open-4.0.5 software in October 2011. Information on the distribution of mobile genetic elements within the primary sequence of chromosome 1 (161,428,367 bp), obtained from the archive, was analyzed with respect to the number and frequency of different mobile elements, using the Microsoft Office Word program for bio-informative purposes. For the subsequent study of the most common mobile genetic elements, the 13436028 bp nucleotide sequence of chromosome 1 was chosen. In this sequence, the number of domains and the frequency of their occurrence were determined, identifying the nearest neighborhoods for homology plots between different families of mobile genetic elements. Based on the obtained data on the number of mobile genetic elements that most frequently are found in the isolated site and the domains formed by them, a table was constructed to evaluate the patterns of distribution of mobile genetic elements in alternative DNA strands. Binomial domains with a frequency of more than 60 %, containing the same mobile genetic element in different chains, were studied in order to find more interesting regularities of distribution and structural organization of RTs in the genome. The functional characteristics of the detected clusters were judged on the basis of their positioning within the structural genes using the Integrated Genome Browser software [21]. Information on selected structural genes was obtained from the international GenBank database [22].

Results. The approach that we used made it possible to identify not only the pairwise colocalization of different elements, but also clusters consisting of three mobile elements. In the 161428367 bp fragment of bovine cattle chromosome 1 analyzed, among all interspersed repeats, SINE (38.27 7%) and LINE (34.002 %) were the most frequent (Table 1). It should be noted that the retrotransposon LINE1 is widely distributed in all genomes of eukaryotes, including mammals. It is known that there are full-size variants of RTs of this family in bovine cattle, which retain retrotransposition activity and participate in genomic transformations [23]. The number of microsatellites (simple repeat) and DNA transposons was 9.549 % and 6.105 %, respectively, and the number of microsatellites with low complexity did not exceed 1.452 %. The main families of SINE were tRNA-Core-RTE (20.989 %), Core-RTE (6.979%) and MIR (6.822 %). In the LINE family, L1 (15.380 %), RTE-BovB (11.631 %) and L2 (5.828 %) were most fully represented, ERVL-MaLR (3,133 %), ERV1 (2,218 %), ERVL (2,091 %), ERVK (2,350 %) and Gypsy (0,212 %) were more common in LTR ERV, and more than half of the DNA transposons were numbered in hAT-Charlie (3.361 %).

In the primary 13,436,028 bp sequence, the domains formation with pairwise localization was studied for the indicated families (Table 2, the percentage see in annexes 2.1 and 2.2 in the electronic version of the article at http://www.agrobiology.ru). The comparison showed that the difference in the frequency of occurrence of domains in both chains did not exceed 2.35 % for

1. Distribution of mobile genetic elements along 161,428,367 bp fragment of bovine cattle chromosome 1

Family	Number	Frequency, %
SINE	122589	38.277
/tRNA-Core-RTE	67222	20.989
/Core-RTE	22351	6.979
/MIR	21850	6.822
/tRNA	10781	3.366
/tRNA-RTE	221	0.069
/5S-Deu-L2	126	0.039
/tRNA-Deu	36	0.011
SINE?/tRNA	2	0.001
LINE	108898	34.002
/L1	49,258	15.380
/RTE-BovB	37251	11.631
/L2	18665	5.828
/CR1	2954	0.922
/RTE-X	669	0.209
/Penelope	61	0.019
/Dong-R4	27	0.008
/Jockey	9	0.003
/L1-Tx1	4	0.001
LTR ERV	33222	10.373
/ERVL-MaLR	10035	3.133
/ERV1	7105	2.218
/ERVL	6698	2.091
/ERVK	7526	2.350
/Gypsy	680	0.212
LTR?	336	0.105
LTR	341	0.106
/Gypsy?	302	0.094
/ERVL?	154	0.048
/ERV1?	45	0.014
Simple_repeat	30581	9.549
DNA	19553	6.105
/hAT-Charlie	10763	3.361
/TcMar-Tigger	3214	1.004
/hAT-Tip100	2114	0.660
/TcMar-Mariner	634	0.198
/hAT-Blackjack	881	0.275
/hAT	377	0.118
/hAT-Ac	256	0.080
/hAT-Tip100?	79	0.025
DNA	283	0.088
/TcMar-Tc2	315	0.098
DNA?	184	0.057
/hAT?	101	0.032
/TcMar	80	0.025
/hAT-Tag1	136	0.042
/PIF-Harbinger	16	0.005
DNA?/hAT-Tip100?	34	0.011
/PiggyBac	31	0.010
/TcMar-Tc1	25	0.008
DNA?/PiggyBac?	11	0.003
/TcMar?	6	0.002
/Kolobok	10	0.003
/TcMar-Pogo	3	0.001
Low_complexity	4650	1.452
Other	773	0.241
Unknown	315	0.098
snRNA	100	0.031
tRNA	116	0.036
RC/Helitron	95	0.030
rRNA	66	0.021
Satellite/centr	33	0.010
RNA	29	0.009
srpRNA	2	0.001
scRNA	1	0.000
RC?/Helitron?	16	0.005
Total	320266	

the indicated SINE families (377 MIR/tRNA-Core-RTE и 346 tRNA-Core-RTE/MIR per 1320 elements of MIR), 1.35 % for LINE (280 L2/tRNA-Core-RTE and 296 tRNA-Core-RTE/L2 per 1183 L2), 4.76 % for LTR ERV (for example, 3 Gypsy/Core-RTE and 0 Core-RTE/Gypsy per 63 Gypsy elements, although in the case of Gypsy/L1 and L1/Gypsy the difference was 7.94 %, i.e. 14.29 versus 6.35 %), and 2.58 % for DNA transposons. So further we considered pairs in the same strand.

Colocalization analysis of SINE with other mobile genetic elements showed that SINE, with the highest frequency, forms domains with elements of its own family. The tRNA-Core-RTE family is particularly active: the frequency of core-RTE/tRNA-Core-RTE domains for Core-RTE is 27.33 % (568 pairs per 2078 elements), of tRNA/tRNA-Core-RTE domains for tRNA is 21.48 % (290 pairs per 1350 elements), and MIR is adjacent to tRNA-Core-RTE in 28.56 % of cases (377 pairs per 1320 elements). In this case, the tRNA-Core-RTE family itself most actively interacts with RTE-BovB (21.06 %, or 1611 pairs per 7651 elements).

Core-RTE is clustered with RTE-BovB and L1 at a frequencies of 16.46 % (342 domains per 2078 Core-RTE) and 15.21 % (316 domains per 2078 Core-RTE), respectively. The tRNA family in 16.74 % of cases is adjacent to L1 (226 domains per 1350 tRNA). MIR/L1 domains occur twice as often as MIR/L2, i.e. 154 (11.67 %) and 75 (5.68 %), respectively, per 1320 MIR. Obtained data contradict the conclusions of D.L. Adelson et al. [2], who considered this option as forming conserved and most ancient binomial domains. Apparently, such disjunctions may be due to the fact that D.L. Adelson et al. considered combined localization of the full-size genes of MIR/L2 retrotransposons. In our study, we estimated the colocalization of homology sites to RTs, the sizes of which could be less than the total length of the identification genes of mobile elements. The data obtained by us shows that when analyzing the formation of retrotransposon domains in the genomic landscape in case of species-specific “young” mobile elements which are actively involved in transpositions, it is necessary to take into account not only full-size sequences

2. The number of pairs of mobile genetic elements (MGE) most often found along 13,436,028 bp fragment of bovine cattle chromosome 1

Family	SINE				LINE			LTR					DNA	Total MGE
	/tRNA-Core-RTE	/Core-RTE	/tRNA	/MIR	/L1	/RTE-BovB	/L2	/ERVL-MaLR	/ERV1	/ERVL	/ERVK	/Gypsy	/hAT-Charlie	number
SINE														
/tRNA-Core-RTE	1425	568	290	377	1038	1574	280	322	197	201	11	10	211	7651
/Core-RTE	584	107	73	77	307	373	81	64	42	40	6	3	56	2078
/tRNA	299	64	76	89	217	58	82	68	20	46	3	2	63	1350
/MIR	346	90	87	101	149	98	75	31	23	28	3	4	56	1320
LINE														
/L1	1055	316	226	154	925	318	118	129	67	72	127	9	88	4162
/RTE-BovB	1611	342	70	71	350	324	80	82	70	49	422	3	53	3756
/L2	296	78	83	75	125	65	142	46	19	30	5	1	45	1183
LTR														
/ERVL-MaLR	312	72	84	54	132	81	39	106	26	22	1	4	19	1071
/ERV1	189	43	41	26	70	71	20	34	134	26	3	2	15	743
/ERVL	191	45	34	24	78	59	25	32	32	69	2	3	16	709
/ERVK	10	2	1	4	135	416	2	2	4	3	7	0	3	599
/Gypsy	13	0	1	2	4	5	3	2	2	4	0	9	5	63
DNA														
/hAT-Charlie	190	55	67	46	97	41	57	28	15	15	2	2	79	813

but also products of MGE recombinations.

The LTR ERV and SINE families formed domains, in particular 84 tRNA/ERVL-MaLR domains, in no more than 6.22 % of cases. Nevertheless, the SINE families were closely associated with all families of LTR ERV, mostly tRNA-Core-RTE. Thus, the ERVL-MaLR/tRNA-Core-RTE domain occurred 322 times (with a frequency of 30.07 %), ERV1/tRNA-Core-RTE — 197 times (26.51 %), ERVL/tRNA-Core-RTE — 201 times (28.35 %), Gypsy/tRNA-Core-RTE — 10 times (15.87 %). Similarly, for DNA-transposons, tRNA was adjacent to hAT-Charlie 67 times with a frequency of 4.96 %, the rest formed binomial domains in less than 4 % of cases. At the same time hAT-Charlie was colocalized with tRNA 63 times with a frequency of 7.75 %, and with tRNA-Core-RTE 211 times (25.95 %). The LINE families were also adjacent to SINE families most often, namely to tRNA-Core-RTE. In particular, the RTE-BovB/tRNA-Core-RTE (1574 domains) accounted for 41.91 % of the sequences, L/tRNA-Core-RTE (1038 domains) for 24.94 %, and L2/tRNA-Core-RTE (280 domains) for 23.67 %. The L1/L1 pair was also common (925 cases, which were 22.22 %). In associations with the LTR ERV family, the RTE-BovB/ERVK domain was most often present (416 cases, or 11.08 %). The frequency of the adjacency of remaining LINE families with LTR ERV families did not exceed 3.30 %. Colocalization with DNA transposons occurred less frequently than in 4.82 % of cases, which was noted for L2/hAT-Charlie (57 domains per 1183 L2), and hAT-Charlie formed domains with the LINE family at a frequency of no more than 10.82 % which was revealed for hAT-Charlie/L1 (88 domains). DNA transposons also rarely coexisted with the LTR ERV family, the frequency of no more than 2.34 % was noted for hAT-Charlie/ERVL-MaLR (19 domains), and of no more than 3.17 % for Gypsy/hAT-Charlie (2 domains).

The most interesting is the colocalization of LTR/ERVK with LINE families. Thus, the number and frequency of domains were 422 cases and 70.45 %, respectively, for ERVK/RTE-BovB type, 127 cases and 21.20 % for ERVK/L1, and 5 cases and 0.83 % for ERVK/L2. In the alternative strand, the frequency for RTE-BovB/ERVK was 69.45 % (416 domains), for L1/ERVK was 22.54 % (135 domains), and for L2/ERVK was 0.33 % (5 domains).

The combined prevalence of the ERVK family with LINE was 92.49 % and 92.32 % in the forward and reverse chains, while the rest of the MGEs did not exceed 4 %. This fact assumes the presence of a trinomial domain of the type LINE/ERVK/LINE.

A closer analysis showed that the ERVK family actually existed in the LINE/ERVK/LINE domain in 85.31 % of cases (511 out of 599). Among the LINE/ERVK/LINE domains, there were two ones noted, the RTE-BovB/BTLTR1/RTE-BovB at a frequency of 74.74 % (382 out of 511), and L1/BTLTR1J/L1 at a frequency of 21.51 % (110 out of 511) (Table 3). Besides, 12 overlapping trinomial clusters were observed, which indicates the high variability of their localization.

Localization analysis of trinomial clusters showed that they cover 6.86 % of the 13,436,028 bp primary sequence of chromosome 1 and are unevenly distributed (Table 4). Moreover, an increase in the density of such clusters was observed closer to the distal end of the fragment. The uneven distribution of families of mobile genetic elements inside and between the chromosomes of bovine cattle was reported [24]. At the same time, analysis of such data is difficult because it is difficult to distinguish between the effects of new insertions and transpositions and the results of their deletion or natural selection against adverse variants (cleansing selection).

3. Types of LINE/ERVK/LINE mobile genetic element domains located along 13,436,028 bp fragment of bovine cattle chromosome 1

Domain type	Domains	
	number	frequency, %
RTE-BovB/BTLTR1/RTE-BovB	382	74.74
RTE-BovB/BTLTR1B/RTE-BovB	1	0.20
RTE-BovB/BTLTR1E2/RTE-BovB	1	0.20
RTE-BovB/BTLTR1J4/RTE-BovB	1	0.20
RTE-BovB/ERV2-1-LTR_BT/RTE-BovB	1	0.20
RTE-BovB/LTR2_BT/RTE-BovB	1	0.20
L1/BTLTR1J/L1	110	21.51
L1/BTLTR1/L1	3	0.59
L1/BTLTR1F/L1	1	0.20
L2/BTLTR1/L2	1	0.20
L1/BTLTR1/RTE-BovB	4	0.78
RTE-BovB/BTLTR1/L2	2	0.39
RTE-BovB/BTLTR1/L1	2	0.39
RTE-BovB/BTLTR1J/L1	1	0.20
Total	511	

4. Distribution of trinomial domains of mobile genetic elements along 13,436,028 bp fragment of bovine cattle chromosome 1

The number and per cent of domains on 12 equal segments of 1,119,669 bp fragments												
1	2	3	4	5	6	7	8	9	10	11	12	Total
LINE/ERVK/LINE												
33	21	29	40	38	33	41	39	34	56	62	85	511
6.46 %	4.11 %	5.68 %	7.83 %	7.44 %	6.46 %	8.02 %	7.63 %	6.65 %	10.96 %	12.13 %	16.63 %	100 %
В том числе												
RTE-BovB/BTLTR1/RTE-BovB												
26	15	21	29	31	25	28	30	29	40	44	64	382
6.81 %	3.93 %	5.50 %	7.59 %	8.12 %	6.54 %	7.33 %	7.85 %	7.59 %	10.47 %	11.52 %	16.75 %	100 %
L1/BTLTR1J/L1												
7	6	8	9	6	5	11	6	5	16	13	18	110
6.36 %	5.45 %	7.27 %	8.18 %	5.45 %	4.55 %	10.00 %	5.45 %	4.55 %	14.55 %	11.82 %	16.36 %	100 %

5. Positions of trinomial clusters of mobile elements as related to structural genes along 13,436,028 bp fragment of bovine cattle chromosome 1

Structural genes	Cluster number	
	1	2
Bos taurus potassium voltage-gated channel subfamily E regulatory subunit 2 (*kcne2*), mRNA (–)	1	0
Bos taurus phosphoribosylglycinamide formyltransferase, phosphoribosylglycinamide synthetase, phosphoribosylaminoimidazole synthetase (*gart*), mRNA (+)	1	0
Bos taurus transmembrane protein 50B (*tmem50b*), mRNA (+)	1	0
Bos taurus interleukin 10 receptor subunit beta (*il10rb*), mRNA (–)	2	0
Bos taurus interferon alpha and beta receptor subunit 2 (*ifnar2*), mRNA (–)	1	0
Bos taurus URB1 ribosome biogenesis 1 homolog (S. cerevisiae) (*urb1*), mRNA (+)	1	0
Bos taurus glutamate ionotropic receptor kainate type subunit 1 (*grik1*), mRNA (+)	9	1
Bos taurus ubiquitin specific peptidase 16 (*usp16*), mRNA (–)	1	0
Bos taurus listerin E3 ubiquitin protein ligase 1 (*ltn1*), mRNA (+)	1	0
Bos taurus cysteine and tyrosine rich 1 (*cyyr1*), mRNA (+)	1	0
Bos taurus amyloid beta precursor protein (*app*), mRNA (+)	9	3
Bos taurus junctional adhesion molecule 2 (*jam2*), mRNA (–)	2	0

Note. 1 — RTE-BovB/BTLTR1/RTE-BovB, 2 — L1/BTLTR1J/L1; "–" and "+" mean that the gene is located in the reverse or the forward chain.

Using Integrated Genome Browser software, we evaluated the positioning of trinomial MGE clusters of the types RTE-BovB/BTLTR1/RTE-BovB and L1/BTLTR1J/L1 as related to structural genes of *Bos taurus*. It was found that these clusters are detected with high frequency within the structural genes which encode glutamate ionotropic receptor kainate type subunit 1 (*grik1*) and amyloid beta precursor protein (*app*). In the *grik1* gene sequences, there were 9 trinomial clusters RTE-BovB/BTLTR1/RTE-BovB out of 30 ones identified in structural genes in the studied chromosome fragment, and one out of 4 L1/BTLTR1J/L1 clusters found. In the *app* gene, 9 clusters RTE-BovB/BTLTR1/RTE-BovB were

found out of 30 clusters RTE-BovB/BTLTR1/RTE-BovB detected, and 3 clusters L1/BTLTR1J/L1 out of 4 clusters L1/BTLTR1J/L1 identified (Table 5). It should be noted that in the sequences of these structural genes, RTE-BovB/BTLTR1/RTE-BovB were encountered with the greatest frequency (see Table 5), despite the increased amount of L1 compared to RTE-BovB in the examined fragment of chromosome 1 (see Table 1). This is consistent with the conclusion that RTE-BovB is an older and more mutated bovine cattle genome element compared to L1 [2]. It is important to emphasize that L1 and RTE-BovB are historically quite distant from each other, although they belong to LINE; nevertheless, in both genes similar trinomial products, the RTE-BovB/BTLTR1/RTE-BovB and L1/BTLTR1J/L1, are present (see Table 5). It can be assumed that such colocalization is associated with the existing in these products recombinations of RT elements with structural and functional similarity, which will be the subject of our further research.

In the literature, there is data on the association of mutations of the subunit 1 of the ionotropic kainate receptor glutamate *grik1* with behavioral pathologies of man, e.g. schizophrenia, epilepsy, depression, bipolar disorder [25-27]. According to available data, the protein of the β-amyloid precursor APP is involved in neuroplasticity processes and is necessary for the survival of nerve cells [28]. The fragment of this protein, the so-called β-amyloid peptide (Aβ), is the main component of senile plaques whish formation is considered to be the main pathomorphological sign of Alzheimer's disease, and the Aβ peptide found in bovine cattle brain shows some similarity with similar human brain peptides in the early stages of aging [29]. High density of localization of trinomial recombination products of species-specific *Bos taurus* retrotransposons BTLINE and BTLTRERV with constant architectonics (i.e. direct repetitions of BTLINE on the flanks of a trinomial structure in one chain and a homology site to BTLTRERV in the center in the alternative chain) found in two genes, closely related to the functions of the central nervous system, let us to assume their certain connection with those signs (reduced aggressiveness toward human) which D.K. Belyaev identified in animals as leading in domestication [30]. Interestingly, the Bov-B insertions, previously identified in the structural gene which is associated with craniofacial peculiarities in bovine cattle, are absent in this gene in humans and mice [31].

In general, the revealed distribution of retrotransposons and their recombination products in the 13,436,028 bp nucleotide sequences of bovine cattle chromosome 1 allows the following conclusions. In the studied of fragment, tRNA-Core-RTE, RTE-BovB, L1 and LTR ERV are often found. Their mutual localization in the genome is complex. The binomial associations, i.e. SINE and LINE, tRNA-Core-RTE and LTR EVR, ERVK/RTE-BovB, ERVK/L1 are the most common. The last two variants serve as the basis for the trinomial clusters RTE-BovB/BTLTR1/RTE-BovB and L1/BTLTR1J/L1. Another RTs of these trinomial clusters are not actually formed. A certain shift was found in the relatively high density of localization of these trinomial clusters to the distal end of the studied fragment of chromosome 1. Localization analysis of the revealed trinomial recombination products between LINE and LTR ERV in relation to structural genes showed that 34 such constructs are detected in 12 structural genes (while the rest are found in intergenic spaces), with 10 and 12 copies in two genes (*grik1* and *app*), closely associated in mammals with central nervous system function. The fact that in each of these two genes there were 9 copies of the trinomial construction RTE-BovB/BTLTR1/RTE-BovB, and the construction L1/BTLTR1J/L1 was found only in one copy in *grik1* and in three copies in *app*, allows us to consider

these genes as ancient targets for insertions and conservation. Note that the construction of L1/BTLTR1J/L1 was found only in these two genes, but not in the remaining 10 in which the RTE-BovB/BTLTR1/RTE-BovB recombination product is present.

So, we obtained data on the existence of regularities in the distribution of retrotransposon fragments and their recombination products in bovine cattle genome. The specific features of the distribution of recombination products between LINE and LTR ERV in the studied fragment of chromosome 1, and their localization in structural genes allow us to assume the possible presence of conserved structural and functional elements that perform a regulatory role. The identification of such elements is the subject of our further research.

REFERENCS

1. Elsik C.G., Tellam R.L., Worley K.C. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*, 2009, 324(5926): 522-528 (doi: 10.1126/science.1169588).
2. Adelson D.L., Raison J.M., Edgar R.C. Characterization and distribution of retrotransposons and simple sequence repeats in the bovine genome. *PNAS USA*, 2009, 106(31): 12855-12860 (doi: 10.1073/pnas.0901282106).
3. Walsh A.M., Kortschak R.D., Gardner M.G., Bertozzi T., Adelson D.L. Widespread horizontal transfer of retrotransposons. *PNAS USA*, 2013, 110(3): 1012-1016 (doi: 10.1073/pnas.1205856110).
4. Godakova S.A., Sevast'yanova G.A., Semenova S.K. *Molekulyarnaya genetika, mikrobiologiya i virusologiya*, 2016, 34(1): 9-12 (doi: 10.18821/0208-0613-2016-34-1-9-12) (in Russ.).
5. Wang X., Liu X. Close ecological relationship among species facilitated horizontal transfer of retrotransposons. *BMC Evolutionary Biology*, 2016, 1: 201 (doi: 10.1186/s12862-016-0767-0).
6. Chen L., Chamberlain A.J., Reich C.M., Daetwyler H.D., Hayes B.J. Detection and validation of structural variations in bovine whole-genome sequence data. *Genet. Sel. Evol.*, 2017, 49: 13 (doi: 10.1186/s12711-017-0286-5).
7. Krull M., Petrusma M., Makalowski W., Brosius J., Schmitz J. Functional persistence of exonized mammalian-wide interspersed repeat elements (MIRs). *Genome Res.*, 2007, 17(8): 1139-1145 (doi: 10.1101/gr.6320607).
8. Schütz E., Wehrhahn C., Wanjek M., Bortfeld R., Wemheuer W.E., Beck J., Brenig B. The Holstein Friesian Lethal Haplotype 5 (HH5) results from a complete deletion of TBF1M and cholesterol deficiency (CDH) from an ERV-(LTR) insertion into the coding region of APOB. *PLoS ONE*, 2016, 11(4): e0154602 (doi: 10.1371/journal.pone.0154602).
9. Kordis D., Gubensek F. Horizontal transfer of non-LTR retrotransposons in vertebrates. *Genetica*, 1999, 107(1-3): 121-128.
10. Kordis D., Gubensek F. Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes. *PNAS USA*, 1998, 95(18): 10704-10709.
11. Gentles A.J., Wakefield M.J., Kohany O., Gu W., Batzer M.A., Pollock D.D., Jurka J. Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. *Genome Res.*, 2007, 17(7): 992-1004 (doi: 10.1101/gr.6070707).
12. Carelli F.N., Hayakawa T., Go Y., Imai H., Warnefors M., Kaessmann H. The life history of retrocopies illuminates the evolution of new mammalian genes. *Genome Res.*, 2016, 26(3): 301-314 (doi: 10.1101/gr.198473.115).
13. Tan S., Cardoso-Moreira M., Shi W., Zhang D., Huang J., Mao Y., Jia H., Zhang Y., Chen C., Shao Y., Leng L., Liu Z., Huang X., Long M., Zhang Y.E.. LTR-mediated retroposition as a mechanism of RNA-based duplication in metazoans. *Genome Res.*, 2016, 26(12): 1663-1675 (doi: 10.1101/gr.204925.116).
14. Garcia-Etxebarria K., Sistiaga-Poveda M., Jugo B.M. Endogenous retroviruses in domestic animals. *Curr. Genomics*, 2014, 15(4): 256-265 (doi: 10.2174/1389202915666140520003503).
15. Mei L., Ding X., Tsang S.-Y., Pun F.W., Ng S.-K., Yang J., Zhao C., Li D., Wan W., Yu C.-H., Tan T.-C., Poon W.-S., Leung G.K.-K., Ng H.-K., Zhang L., Xue H. AluScan: a method for genome-wide scanning of sequence and structure variations in the human genome. *BMC Genomics*, 2011, 12: 564 (doi: 10.1186/1471-2164-12-564).
16. Ng S.-K., Hu T., Long X., Chan C.-H., Tsang S.-Y., Xue H. Feature co-localization landscape of the human genome. *Sci. Rep.*, 2016, 6: 20650 (doi: 10.1038/srep20650).
17. Liu M., Eiden M.V. Role of human endogenous retroviral long terminal repeats (LTRs) in maintaining the integrity of the human germ line. *Viruses*, 2011, 3(6): 901-905 (doi: 10.3390/v3060901).

18. Glazko V.I., Feofilov A.V., Bardukov N.V., Glazko T.T. *Izvestiya TSKhA*, 2012, 1: 118-125 (in Russ.).
19. Smit A.F.A., Hubley R., Green P. *RepeatMasker*. Available http://repeatmasker.org. No date.
20. *bosTau7.fa.out.gz, Oct 2011. RepeatMasker open-4.0.5, Repeat Library 20140131*. Available http://www.repeatmasker.org/species/bosTau.html. No date.
21. Nikol J.W., Helt G.A., Blanchard S.G. Jr., Raja A., Loraine A.E. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics*, 2009, 25(20): 2730-2731 (doi: 10.1093/bioinformatics/btp472).
22. *GenBank*. Available https://www.ncbi.nlm.nih.gov/genbank/. No date.
23. Ivancevic A.M., Kortschak R.D., Bertozzi T., Adelson D.L. LINEs between species: evolutionary dynamics of LINE-1 retrotransposons across the eukaryotic tree of life. *Genome Biol. Evol.*, 2016, 8(11): 3301-3322 (doi: 10.1093/gbe/evw243).
24. Saylor B., Elliott T.A., Linquist S., Kremer S.C., Gregory T.R., Cottenie K. A novel application of ecological analyses to assess transposable element distributions in the genome of the domestic cow, *Bos taurus*. *Genome*, 2013, 56(9): 521-533 (doi: 10.1139/gen-2012-0162).
25. Hirata Y., Zai C.C., Souza R.P., Lieberman J.A., Meltzer H.Y., Kennedy J.L. Association study of *GRIK1* gene polymorphisms in schizophrenia: case-control and family-based studies. *Hum. Psychopharmacol.*, 2012, 27(4): 345-351 (doi: 10.1002/hup.2233).
26. Le-Niculescu H., Patel S.D., Bhat M., Kuczenski R., Faraone S.V., Tsuang M.T., McMahon F.J., Schork N.J., Nurnberger J.I. Jr., Niculescu III A.B. Convergent functional genomics of genome-wide association data for bipolar disorder: comprehensive identification of candidate genes, pathways and mechanisms. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, 2009, 150B(2): 155-181 (doi: 10.1002/ajmg.b.30887).
27. Bashkatov S.A., Nurgalieva A.Kh., Enikeeva R.F., Kazantseva A.V., Khusnutdinova E.K. *Vestnik YuUrGU. Seriya Psikhologiya*, 2016, 9(4): 25-39 (doi: 10.14529/psy160403) (in Russ.).
28. Tatarnikova O.G., Orlov M.A., Bobkova N.V. *Uspekhi biologicheskoi khimii*, 2015, 55: 351-390 (in Russ.).
29. Costassa E.V., Fiorini M., Zanusso G., Peletto S., Acutis P., Baioni E., Maurella C., Tagliavini F., Catania M., Gallo M., Lo Faro M., Chieppa M.N., Meloni D., D'Angelo A., Paciello O., Ghidoni R., Tonoli E., Casalone C., Corona C. Characterization of amyloid-β deposits in bovine brains. *Journal of Alzheimer's Disease*, 2016, 51: 875-887 (doi: 10.3233/JAD-151007).
30. Trut L.N. *Vestnik VOGiS*, 2007, 11(2): 273-289 (in Russ.).
31. Iwashita S., Itoh T., Takeda H., Sugimoto Y., Takahashi I., Nobukuni T., Sezaki M., Masui T., Hashimoto K. Gene organization of bovine BCNT that contains a portion corresponding to an endonuclease domain derived from an RTE-1 (Bov-B LINE), non-LTR retrotransposable element: duplication of an intramolecular repeat unit downstream of the truncated RTE-1. *Gene*, 2001, 268(1-2): 59-66 (doi: 10.1016/S0378-1119(01)00422-X).