Journal of Siberian Federal University. Biology 2018 11(1): 60-74
~ ~ ~

УДК 577.21; 577.29

Functional Insights into Genic Neighbourhood Organization of Helitron Transposons in *Bos taurus* Genomes

Anna V. Babii*[a], Svetlana N. Koval'chuk[a], Tat'iana T. Glazko[a,b], Valerii I. Glazko[a,b] and Gleb I. Kosovskii[a]

[a]Federal State Budget Scientific Institution Center of Experimental Embryology and Reproductive Biotechnologies bld. 4, 12 Kostyakova Str., Moscow, 127422, Russia

[b]Russian State Agrarian University – Moscow Timiryazev Agricultural Academy 49 Timiryazevskaya Str., Moscow, 127550, Russia

Received 05.08.2016, received in revised form 08.08.2016, accepted 27.12.2016, published online 24.01.2017

Transposable elements (TEs) represent well-known factors of genomic variability and evolution. TEs are important providers of regulatory elements that are able to significantly influence the architecture and expression of the host genome. Currently, of a special interest are the DNA transposons helitrons. They are supposed to be involved in horizontal transfer of genetic material between distant taxa and to dramatically impact the host genomes via phenomena of exon-shuffling and gene capture. Thereby, and due to their high level of polymorphism and relatively high frequency in the eukaryotic genomes, helitrons can be used as "anchors" for genome scanning of different breeds of farm animals aimed at revealing their "gene pool standards". Currently, there are no comprehensive studies dedicated to helitrons and their interaction and impact on host genomic landscape in cattle (Bos taurus). Earlier we showed the possibility of using the 3'-end consensus sequence of Heligloria helitrons for estimation of consolidation of different cattle breeds via multilocus genotyping. In the present study, in order to investigate the context features of the DNA regions flanked by the inverted repeats of Heligloria helitrons fragments in Bos taurus genomes, we pyrosequenced such fragments (of about 550 bp in length) from three cattle breeds and analyzed the functional implications of the identified genes. Thus, here we provide an insight into the functional organization of the genic neighbourhood of helitron transposons in the genomes of different Bos taurus breeds and an attempt to understand possible consequences of such distribution of helitrons on these genomes.

Keywords: Bos taurus, helitrons, genes, genome, functions, breeds.

* Corresponding author E-mail address: anna.babii@yahoo.com

Citation: Babii A.V., Koval'chuk S.N., Glazko T.T., Glazko V.I., Kosovskii G.I. Functional insights into genic neighbourhood organization of helitron transposons in *Bos taurus* genomes. J. Sib. Fed. Univ. Biol., 2018, 11(1), 60-74. DOI: 10.17516/1997-1389-0003.

Функциональное исследование организации генного окружения транспозонов хелитронов в геномах *Bos taurus*

А.В. Бабий[а], С.Н. Ковальчук[а], Т.Т. Глазко[а,б], В.И. Глазко[а,б], Г.Ю. Косовский[а]

[а]Центр экспериментальной эмбриологии и репродуктивных биотехнологий
Россия, 127422, Москва, ул. Костякова, 12, стр. 4
[б]Российский государственный аграрный университет – МСХА им. К.А. Тимирязева
Россия, 127550, Москва, ул. Тимирязевская, 49

Мобильные элементы (МЭ) являются хорошо известными факторами геномной изменчивости и эволюции. Будучи важными источниками регуляторных элементов, МЭ способны оказывать существенное влияние на архитектуру и экспрессию генома хозяина. В настоящее время особый интерес вызывает исследование ДНК транспозонов хелитронов. Предполагается, что хелитроны вовлечены в горизонтальный перенос генетического материала между удаленными таксонами и способны значительно повлиять на геном хозяина за счет механизмов перетасовки экзонов (exon-shuffling) и «захвата» генов (gene capture). Благодаря этому и в связи с высоким уровнем полиморфизма и относительно высокой плотностью в геномах эукариот, хелитроны могут быть использованы в качестве «якорей» для геномного сканирования различных пород сельскохозяйственных животных с целью выявления их генофондного стандарта. В настоящее время не существует комплексных исследований, посвященных хелитронам и их взаимодействию и влиянию на геномный ландшафт крупного рогатого скота (Bos taurus). Ранее нами была показана возможность использования консенсусной последовательности 3'-конца хелитронов семейства Heligloria для полилокусного генотипирования (геномного сканирования) с целью оценки консолидированности разных пород крупного рогатого скота. В настоящем исследовании для изучения контекстных особенностей участков ДНК, фланкированных инвертированными повторами фрагментов хелитронов Heligloria, были секвенированы подобные геномные участки (около 550 п.н.), принадлежащие трем породам крупного рогатого скота, и проанализированы функциональные особенности выявленных генов. Таким образом, в данной работе представлена функциональная организация генного окружения транспозонов хелитронов в геномах разных пород Bos taurus и обсуждается возможное влияние подобного распределения хелитронов на геномы исследованных групп животных.

Ключевые слова: Bos taurus, хелитроны, гены, геном, функции, породы.

Introduction

Transposable elements (TEs) or transposons represent well-known factors of genomic variability and evolution (Chénais et al., 2012). For a long time TEs had been recognized as major sources of non-coding or "junk" DNA with uncertain functions for the organism's genome. The current data on TEs abundance in the genomes of different taxa of prokaryotes and eukaryotes show a lack of correlation between the level of organism's evolution and the amount of transposons hosted in its genome. Despite being considered as genetic burden for the host genomes, today exhaustive research projects such as ENCODE (Encyclopedia of DNA Elements) are gathering reliable evidence towards the important role of transposons in environmental adaptation of organisms and disease. Due to their structural elements and ability to multiply, transpose and mutate, TEs become important mutators and providers of regulatory elements that are able to significantly influence the architecture and expression of the host genome (Rebollo et al., 2011).

Currently, of a special interest are the DNA transposons helitrons. Discovered in 2001 by computational analysis of whole genome sequences in plant genomes (Kapitonov, Jurka, 2001), the *Helitron* superfamily represents class II TEs that are hypothesized to replicate via a "rolling circle" mechanism (Kapitonov, Jurka, 2001; Kapitonov, Jurka, 2007). This mechanism involves a helitron-encoded Rep/Helicase, which is predicted to act both as HUH endonuclease and 5' to 3' helicase, and generation of a single-stranded DNA intermediate without any subsequent creation of target site duplications (Kapitonov, Jurka, 2001; Kapitonov, Jurka, 2007; Thomas, Pritham, 2015). The "rolling circle" mechanism of replication, typical for single-stranded DNA viruses or bacterial plasmids, has not been revealed in other known DNA transposons that migrate through genomes via "cut-and-paste" transposition (Kapitonov, Jurka, 2001; Kapitonov, Jurka, 2007; Thomas, Pritham, 2015). Also, all known helitrons share in common several structural features: the 5' TC and 3' CTRR termini and a palindrome sequence (16 to 20 bp) upstream the 3'-end (Feschotte, Wessler, 2001; Kapitonov, Jurka, 2001).

First discovered in plants, helitrons have soon been shown to exist in many other eukaryotic genomes, including mammals (Thomas et al., 2010; Thomas et al., 2014). These DNA transposons are supposed to be involved in horizontal transfer of genetic material between distant taxa (Thomas et al., 2010; Guo et al., 2014; Thomas et al., 2014; Coates, 2015) and to impact the genome architecture of their host as "DNA shuttles" (Feschotte, Wessler, 2001) via their implication in processes of exon-shuffling and gene capture (Gupta et al., 2005).

Due to their relatively high frequency in the eukaryotic genomes that ranges from 1.6% in *Arabidopsis thaliana* (Xiong et al., 2014) to up to 5.8% in little brown bat *(Myotis lucifugus)* (Thomas et al., 2014) and even 6.6% in maize *(Zea mays)* (Xiong et al., 2014), their high level of polymorphism and their considerable implication in genomic rearrangements (Kapitonov, Jurka, 2001; Kapitonov, Jurka, 2007; Thomas, Pritham, 2015), helitrons become promising candidate markers of highly polymorphic genomic regions. One possible application of helitron fragments is their use as "anchors" for multilocus genotyping (genome scanning) aimed at revealing the "gene pool standards" of different breeds of farm animals. Currently, there are no comprehensive studies dedicated to helitrons and their interaction and impact on host genomic landscape in cattle (*Bos taurus*). Earlier we showed the possibility of using the 3'-end consensus sequence of *Heligloria* helitrons for estimation of consolidation of different

cattle breeds via multilocus genotyping (Babii et al., 2015). In the present study, in order to investigate the context features of the DNA regions flanked by the inverted repeats of *Heligloria* helitrons fragments in the *Bos taurus* genomes, we pyrosequenced such fragments (of about 550 bp in length) from three cattle breeds and analyzed the identified genes, based on the existing data on their functions. Thus, here we provide an insight into the organization of the genic neighbourhood of helitron transposons in the genomes of different *Bos taurus* breeds and an attempt to understand possible consequences of such distribution of helitrons on these genomes.

Materials and Methods

Sample preparation and DNA pyrosequencing

The genomic regions of about 550 bp, flanked by the inverted repeats of 3'-end consensus sequence of *Heligoria* family of helitrons were previously obtained via multilocus genotyping by ISSR-PCR-like (inter-simple sequence repeat polymerase chain reaction) amplification (Babii et al., 2015) of 3 cattle breeds: the local beef Kalmyk breed (30 animals) and 2 factory-grown dairy breeds – the Ayrshire breed (15 animals) and the holsteinized Black-and-White cattle (15 animals). Briefly, the PCR samples were prepared using *qPCRmix-HS (5x)* mix (Evrogen, Russia) and the 3'-end of the *Heligloria* helitrons as a primer (5'-GCAACGCGTGGCCGG-3'). The ISSR-PCR-like amplification was run under optimized conditions (95 °C for 2 min; 94 °C for 15 sec, 56 °C for 15 sec, 72 °C for 2 min, 40 cycles; 72 °C for 2 min). Subsequently, the obtained amplicons were separated in ethidium bromide stained agarose gel (1.2%) for 120 min, and their length was determined using the DNA marker (*O'Gene Ruler DNA Ladder mix 100-10 000 bp*, Thermo Scientific, USA).

The amplicons of about 550 bp, obtained after the multilocus genotyping of the above-mentioned animals, were further purified from the agarose gel and prepared for DNA sequencing. For that purpose we used the 454 pyrosequencing technology (GS Junior from Roche, Switzerland) and the manufacturer's protocols for sample preparation and sequencing.

Sequence analysis

Only reliable sequences, in terms of their coverage (at least 10x), length (≥ 400 bp) and primer sequence at flanks (at least 10 primer nucleotides), were considered for further research. The selected DNA sequences were analyzed using the NCBI BLASTn tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (NCBI Genomes (chromosome) database, Bos_taurus_UMD_3.1.1 reference genome, default parameters; cutoff – 90% identity, E-value 0.0). We used NCBI's Gene database (for *Bos taurus*) to map the genomic regions – like untranslated regions (UTRs), exons or introns – flanked by the inverted sequences of 3'-end consensus of *Heligloria* helitrons, on the corresponding position within the genes. The identified genes were then functionally annotated, based on available information on their ontologies – in PANTHER Gene List Analysis tool (http://pantherdb.org/) and known functional implications – in UniProt (http://www.uniprot.org/) and GeneCards (http://www.genecards.org/) – for *Homo sapiens* as search species, as the corresponding data for the studied species *Bos taurus* are insufficient or missing.

Results and Discussion

Sequencing results – genes identified in the studied Bos taurus breeds

In the sequenced genomic regions, flanked by the inverted sequences of 3'-end of *Heligloria* helitrons from 3 studied cattle breeds, we identified sequence homology to 14 cattle genes.

According to our results, such fragments are differently distributed across the analyzed groups of cattle. For instance, DOCK5 and TMEM41A genes have been shown to be flanked by 3'-end helitron fragments in all 3 breeds (Table 1). In contrast, such fragments of the following genes were found in the sequences belonging to only one breed: SLC16A6 and PHC2 – in the representatives of Kalmyk breed; UBASH3A, NRG3, ARL4C, LOC784305 and BRD9 – in cows of Ayrshire breed; MYADML2, LOC61431 and LOC107131227 – in the analyzed population of holsteinized Black-and-White cattle (Table 1). Also, the helitrons flank the LOC100848239 gene in Kalmyk and holsteinized cattle and PTPRN2 gene in the representatives of the studied dairy breeds (Table 1).

Functional implications of the identified genes

In order to investigate the possible functional implications of the 14 genes identified above, we first looked for the available gene ontology data. The results of PANTHER search showed that, generally, the identified genes represent proteins with nucleic acid binding, enzyme modulator and transporter activities, involved in diverse metabolic and cellular processes (Fig. 1). The details on ontologies of the identified genes are represented in Table 2.

Further, we looked for the information on these 14 genes in other databases, like GeneCards and UniProt (see Materials and Methods section) with regard to their known functions (in *Homo sapiens* as search species, due to the lack of data for the studied species *Bos taurus*). The findings allowed us to group the all 14 genes under study in 4 functional groups. Thereby, 4 genes (MYADML2, DOCK5, TMEM41A, UBASH3A) are involved in the immune system functioning, 5 genes (DLGAP2, PLCXD1, NRG3, PTPRN2, ARL4C) are components of the membrane signaling system, 2 genes (SLC16A6, ABCC4) allow the membrane transport, and 3 genes

Table 1. Genes identified in the sequenced genomic regions, flanked by the inverted sequences of 3'-end of *Heligloria* helitrons from two commercial dairy breeds of Ayrshire and holsteinized cattle and one local beef breed of Kalmyk cattle

Gene symbol	Breed	Breed "status"
SLC16A6 PHC2	Kalmyk	Aborigine breed
MYADML2 LOC61431 LOC107131227	Holsteinized Black-and-White cattle	Commercial dairy breed
UBASH3A NRG3 ARL4C LOC784305 BRD9	Ayrshire	Commercial dairy breed
PTPRN2	Ayrshire Holsteinized Black-and-White cattle	Commercial dairy breeds
LOC100848239	Kalmyk	Aborigine breed
	Holsteinized Black-and-White cattle	Commercial dairy breed
DOCK5 TMEM41A	Kalmyk	Aborigine breed
	Ayrshire Holsteinized Black-and-White cattle	Commercial dairy breeds

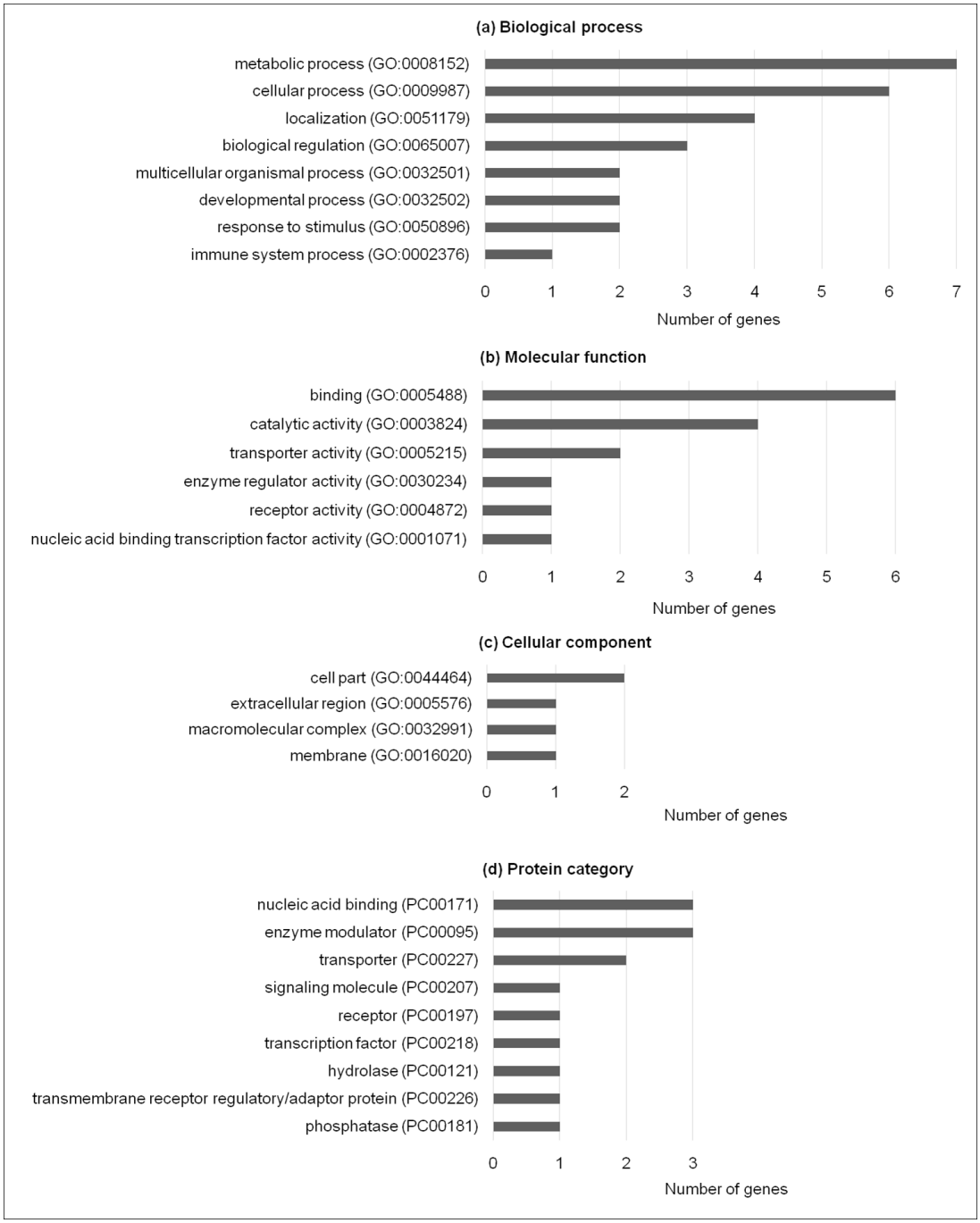

Fig 1. Functional investigation of genes, identified in the sequenced genomic regions flanked by the inverted sequences of 3'-end of *Heligloria* helitrons, and their number, according to their ontological categories and protein class from PANTHER database: (a) Biological process; (b) Molecular function; (c) Cellular component; (d) Protein category

Table 2. Ontologies and protein class data extracted from PANTHER for the 14 genes, identified in the sequenced genomic regions flanked by the inverted repeats of 3'-end consensus sequence of *Heligloria* helitrons in the genomes of 3 studied cattle breeds. Data presented for *Homo sapiens* as search species

Gene	PANTHER protein class	PANTHER GO-Slim Biological process	PANTHER GO-Slim Molecular function	PANTHER GO-Slim Cellular component
1	2	3	4	5
UBASH3A	None	metabolic process (GO:0008152)	None	None
ZRSR2	ribonucleoprotein (PC00171)	mRNA splicing, via spliceosome (GO:0008152)	RNA binding (GO:0005488)	ribonucleoprotein complex (GO:0032991)
PHC2	transcription factor (PC00218); chromatin/ chromatin-binding protein (PC00171)	transcription from RNA polymerase II promoter (GO:0008152); cell cycle (GO:0044238); ectoderm development (GO:0006139); mesoderm development (GO:0016070); regulation of transcription from RNA polymerase II promoter (GO:0006351)	sequence-specific DNA binding transcription factor activity (GO:0001071); sequence-specific DNA binding transcription factor activity (GO:0003700); chromatin binding (GO:0005488)	None
ABCC4	ATP-binding cassette (ABC) transporter (PC00227)	immune system process (GO:0002376); metabolic process (GO:0008152); response to toxic substance (GO:0050896); extracellular transport (GO:0009636)	ATPase activity, coupled to transmembrane movement of substances (GO:0003824); transmembrane transporter activity (GO:0016787)	None
DOCK5	guanyl-nucleotide exchange factor (PC00095)	metabolic process (GO:0008152); cellular component movement (GO:0009987); cell communication (GO:0006928); intracellular protein transport (GO:0007154); phagocytosis (GO:0051179); regulation of catalytic activity (GO:0006810)	catalytic activity (GO:0003824); protein binding (GO:0005488); small GTPase regulator activity (GO:0005515); guanyl-nucleotide exchange factor activity (GO:0030234)	None

Continuation of Table 2

1	2	3	4	5
NRG3	growth factor (PC00207); membrane-bound signaling molecule (PC00112); kinase activator (PC00152)	cell communication (GO:0009987); single-multicellular organism process (GO:0007154); system development (GO:0032501); response to stimulus (GO:0044707); regulation of biological process (GO:0032502)	protein binding (GO:0005488)	extracellular region (GO:0005576)
ARL4C	small GTPase (PC00095)	metabolic process (GO:0008152); cell communication (GO:0009987); intracellular protein transport (GO:0007154); vesicle-mediated transport (GO:0051179)	GTPase activity (GO:0003824); protein binding (GO:0016787)	None
SLC16A6	Transporter (PC00227)	cellular process (GO:0009987); anion transport (GO:0051179)	transmembrane transporter activity (GO:0005215)	plasma membrane (GO:0016020); integral to membrane (GO:0005886); cell part (GO:0016021)
TMEM41A	None	None	None	None
PLCXD1	None	None	None	None
MYADML2	None	None	None	None
DLGAP2	transmembrane receptor regulatory/ adaptor protein (PC00226)	neurological system process (GO:0032501)	None	None
BRD9	chromatin/ chromatin-binding protein (PC00171)	None	nucleic acid binding (GO:0005488); chromatin binding (GO:0003676)	None
PTPRN2	receptor (PC00197); protein phosphatase (PC00181); protein phosphatase (PC00195)	cellular protein modification process (GO:0008152); cell communication (GO:0044238)	phosphoprotein phosphatase activity (GO:0003824); phosphoprotein phosphatase activity (GO:0016787); receptor activity (GO:0016788)	cytoplasm (GO:0044464)

(BRD9, PHC2, ZRSR2) regulate transcription (Table 3). Here we have determined differences in localization of helitron 3'-end identity sequence in the 3 studied cattle breeds. Given this fact, we can expect that identification of such differences in different breeds and inbred groups may help control the dynamics of genetic structures, as well as reveal the "gene pool standards" of breeds and inbred groups and predict the reproduction of animals with desired phenotypes.

Helitron insertions and their impact on gene expression

It is a well-known fact that insertions of TEs are able to greatly alter the gene structure and expression (Chénais et al., 2012). Such changes are available in the case of helitron transposons, as well, as there have been described many examples of how these DNA transposons can impact the gene integrity and function (Thomas, Pritham, 2015). It has been shown that helitron insertions can take place at different sites with corresponding consequences. Insertion in the promoter can disrupt the gene transcription (Gupta et al., 2005) or even provide a *de novo* promoter (Miller et al., 1995), while integration of helitrons upstream can enhance transcription (Inagaki et al., 2009). Integration of helitrons in the 5' UTR determines the diversification of the transcripts (Han et al., 2013), while their positioning in the 3' UTR can cause loss of function, due to disruption of polyadenylation (Tsukamoto et al., 2010). Helitrons are also able to jump into introns, thus disrupting or altering splicing (Barbaglia et al., 2012), or "capture" exons and other structural elements from different genes, creating chimeric transcripts (Lai et al., 2005; Barbaglia et al., 2012; Grabudzija et al., 2016). Besides, one possible consequence of gene capture by helitrons is the implication in host gene regulation by generation of small RNAs, which target both the helitron and the parental gene (Li et al., 2013). Moreover, helitrons are able to promote the generation of protogene families through gene capture at RNA level (Thomas et al., 2014). In our study we were able to locate the DNA fragments, flanked by the inverted sequences of 3'-end of *Heligloria* helitrons in the genomes of the studied cattle breeds, on *Bos taurus* reference genome (Table 4). According to our results, in most of the cases the helitron-captured gene fragments are located in intronic regions. There are also gene regions, flanked by 3'-end of helitrons, that match the positions like 5' UTRs neighbouring exons or parts of exons and introns altogether (Table 4).

The dramatic changes triggered by transpositional activity of helitrons (Thomas et al., 2014), the link between the activity of transposable elements and occurrence of copy number variations (CNVs) and the fact that CNVs overlap with cattle genes involved in immune response and milk production (da Silva et al., 2016; Xu et al., 2016), particularly, are consistent with our findings on proximity of localization of the inverted repeats of helitron identity sequences (consensus sequence of 3'-end) and the corresponding functional groups of genes (Table 3). Findings are of a particular interest, given the fact that the cattle genome differs considerably from other mammalian genomes, including human, by the increased number of innate immunity gene families (Elsik et al., 2009). It is worth noting that preferable helitron localization in introns (Table 4) may indicate at the phenomenon of purifying selection, which prevents keeping of these transposons in gene coding regions. Given the fact that we have identified interbreed differences in helitron localization even in closely related breeds, like Ayrshire and holsteinized cattle, indicates at intense involvement of helitrons in complicate processes of genomic variability.

Table 3. Genes identified in the sequenced genomic regions, flanked by the inverted sequences of 3'-end of *Heligloria* helitrons from 3 studied cattle breeds, and grouped according to their known functional implications

Gene symbol	Gene name	Protein accession	Function reference (in *H. sapiens*)	Breed
Immune system functioning				
MYADML2	Myeloid-Associated Differentiation Marker-Like 2	NP_001070416	(König et al., 2008; Reversade et al., 2009; Aranda et al., 2011)	Holsteinized Black-and-White cattle
DOCK5	Dedicator Of Cytokinesis 5	NP_001309739	(Sanders et al., 2009; El-Sayed Moustafa et al., 2012)	Kalmyk Ayrshire Holsteinized Black-and-White cattle
TMEM41A	Transmembrane Protein 41A	NP_001068668	(Korfali et al., 2010)	Kalmyk Ayrshire Holsteinized Black-and-White cattle
UBASH3A	Ubiquitin Associated And SH3 Domain Containing A	NP_001015599	(Diaz-Gallo et al., 2013; Cai et al., 2014; Liu et al., 2015)	Ayrshire
Membrane signaling system				
LOC100848239 (DLGAP2*)	Disks large-associated protein 2-like	XP_015324048	(Wu et al., 2013; Li et al., 2014)	Kalmyk Holsteinized Black-and-White cattle
LOC61431 (PLCXD1*)	PI-PLC X domain-containing protein 1-like	NP_001096774	(Tarpey et al., 2009)	Holsteinized Black-and-White cattle
NRG3	Neuregulin 3	NP_001192407	(Yang et al., 2013; Tost et al., 2014; Wang et al., 2014)	Ayrshire
PTPRN2	Protein Tyrosine Phosphatase, Receptor Type, N	XP_010803008	(Wu et al., 2007; Ross, 2014; Sorokin et al., 2015)	Ayrshire Holsteinized Black-and-White cattle
ARL4C	ADP-Ribosylation Factor-Like 4C	NP_001095818	(Jacobs et al., 1999; Fujii et al., 2015)	Ayrshire
Membrane transport				
SLC16A6	Solute carrier family 16, member 6	NP_001179601	(Murakami et al., 2005; Halestrap, 2013)	Kalmyk
LOC784305 (ABCC4*)	Multidrug resistance-associated protein 4-like	XP_005213957	(Borst et al., 2000)	Ayrshire
Regulation of transcription				
BRD9	Bromodomain Containing 9	NP_001180021	(Filippakopoulos et al., 2012; Flynn et al., 2015; Huang et al., 2015)	Ayrshire
PHC2	Polyhomeotic Homolog 2	NP_001179837	(Wei et al., 2006)	Kalmyk
LOC107131227 (ZRSR2*)	U2 small nuclear ribonucleoprotein auxiliary factor 35 kDa subunit-related protein 2-like	XP_015317295	(Shen et al., 2010; Hong et al., 2015)	Holsteinized Black-and-White cattle

* – Gene symbols of the corresponding orthologs in *H. sapiens* genome

Table 4. Genomic location of the DNA fragments, flanked by the inverted sequences of 3'-end of *Heligloria* helitrons in the genomes of the studied cattle breeds, and the corresponding "captured" gene regions (according to *Bos taurus* reference genome assembly Bos_taurus_UMD_3.1.1)

Gene	Chromosome	Accession	*Helitron*-flanked sequence coordinates on chromosome		"Captured" gene regions
			From	To	
MYADML2	19	AC_000176	51562400	51562868	5' UTR-exon
DOCK5	8	AC_000165	73478005	73478485	Intron
TMEM41A	1	AC_000158	82285812	82286326	5' UTR-exon
UBASH3A	1	AC_000158	144296391	144296855	Intron
LOC100848239	21	AC_000178	59998461	59998944	Intron
LOC61431	X	AC_000187	143661532	143662030	Exon-intron
NRG3	28	AC_000185	37251244	37251734	Intron
PTPRN2	4	AC_000161	119785732	119786262	Intron
ARL4C	3	AC_000160	114716061	114716524	Exon
SLC16A6	19	AC_000176	62492807	62493336	Intron
LOC784305	12	AC_000169	70524501	70524981	Intron
BRD9	20	AC_000177	71517715	71518191	Intron-exon-intron
PHC2	2	AC_000159	121218094	121218535	Exon-intron
LOC107131227	X	AC_000187	143060449	143060948	Intron

Conclusion

In the current work we explored the localization of DNA tranposons helitrons in the genomes of 3 cattle breeds (one aborigine breed of Kalmyks and 2 commercial dairy breeds of Ayrshire and holsteinized Black-and-White cattle), provided an analysis of the preferential sites of insertion of helitrons within cattle genes and determined the functional implications of such genes. We determined that the greatest number of genes, identified in the sequenced genomic regions flanked by the inverted repeats of *Heligloria* helitrons in the genomes of the studied breeds, were involved in cell signaling systems and immune system functioning. Moreover, the results revealed that helitrons tend to flank different gene regions in different breeds, according to their origin (local mixed or factory-grown dairy cattle). It is worth noting that the cattle genome differs from other mammalian genomes, including human, by over-representation of genes, whose products are engaged in the immune system functioning (Elsik et al., 2009). According to this fact and considering our findings, seems that the intense artificial selection of *Bos taurus* as a species targets different signaling and immune system genes. The evolution of these genes under helitron and other TEs mobilization, along with factors of natural and/or artificial selection, and subsequent gene expression may promote the emergence of cattle breeds of different trends of productivity (beef, dairy or mixed).

References

Aranda J.F., Reglero-Real N., Kremer L., Marcos-Ramiro B., Ruiz-Sáenz A., Calvo M., Enrich C., Correas I., Millán J., Alonso M.A. (2011) MYADM regulates Rac1 targeting to ordered

membranes required for cell spreading and migration. *Molecular Biology of the Cell*, 22(8): 1252–1262

Babii A.V., Kovalchuk S.N., Glazko T.T., Kosovsky G.Y., Glazko V.I. (2015) DNA transposon helitron in cattle genome. *Veterinary and Feeding* [Veterinariia i Kromlenie], 3: 30–33 (in Russian)

Barbaglia A.M., Klusman K.M., Higgins J., Shaw J.R., Hannah L.C., Lal S.K. (2012) Gene capture by Helitron transposons reshuffles the transcriptome of maize. *Genetics*, 190(3): 965–975

Borst P., Evers R., Kool M., Wijnholds J. (2000) A family of drug transporters: the multidrug resistance-associated proteins. *Journal of the National Cancer Institute*, 92(16): 1295–1302

Cai T., Wang X., Muhali F.-S., Song R., Shi X., Jiang W., Xiao L., Li D., Zhang J. (2014) Lack of association between polymorphisms in the UBASH3A gene and autoimmune thyroid disease: a case control study. *Arquivos Brasileiros de Endocrinologia E Metabologia*, 58(6): 640–645

Chénais B., Caruso A., Hiard S., Casse N. (2012) The impact of transposable elements on eukaryotic genomes: From genome size increase to genetic adaptation to stressful environments. *Gene*, 509(1): 7-15. doi: 10.1016/j.gene.2012.07.042

Coates B.S. (2015) Horizontal transfer of a non-autonomous Helitron among insect and viral genomes. *BMC Genomics*, 16(1): 137. doi: 10.1186/s12864-015-1318-6

da Silva J.M., Giachetto P.F., da Silva L.O., Cintra L.C., Paiva S.R., Yamagishi M.E., Caetano A.R. (2016) Genome-wide copy number variation (CNV) detection in Nelore cattle reveals highly frequent variants in genome regions harboring QTLs affecting production traits. *BMC Genomics*, 17(1): 454. doi: 10.1186/s12864-016-2752-9

Diaz-Gallo L.M., Sánchez E., Ortego-Centeno N., Sabio J.M., García-Hernández F.J., de Ramón E., González-Gay M.A., Witte T., Anders H.J., González-Escribano M.F., Martin J. (2013) Evidence of new risk genetic factor to systemic lupus erythematosus: the UBASH3A gene. *PLoS ONE*, 8(4): e60646. doi: 10.1371/journal.pone.0060646

El-Sayed Moustafa J.S., Eleftherohorinou H., De Smith A.J. et al. (2012) Novel association approach for variable number tandem repeats (VNTRs) identifies DOCK5 as a susceptibility gene for severe obesity. *Human Molecular Genetics*, 21(16): 3727–3738

Elsik C.G., Tellam R.L., Worley K.C. (2009) The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*, 324 (5926): 522–528

Feschotte C., Wessler S.R. (2001) Treasures in the attic: rolling circle transposons discovered in eukaryotic genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 98(16): 8923–8924

Filippakopoulos P., Picaud S., Mangos M., Keates T., Lambert J.P., Barsyte-Lovejoy D., Felletar I., Volkmer R., Müller S., Pawson T., Gingras A.C., Arrowsmith C.H., Knapp S. (2012) Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell*, 149(1): 214–231

Flynn M., Huang O., Poy F., Oppikofer M., Bellon S.F., Tang Y., Cochran A.G. (2015) A subset of human bromodomains recognizes butyryllysine and crotonyllysine histone peptide modifications. *Structure*, 23(10): 1801–1814

Fujii S., Matsumoto S., Nojima S., Morii E., Kikuchi A. (2015) Arl4c expression in colorectal and lung cancers promotes tumorigenesis and may represent a novel therapeutic target. *Oncogene*, 34(37): 4834–4844

Grabudzija I., Messing S., Thomas J., Cosby R., Bilic I., Miskey C., Gogol-Döring A., Kapitonov V., Diem T., Dalda A., Jurka J., Pritham E.J., Dyda F., Izsvák Z., Ivics Z. (2016) A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. *Nature Communications*, 7: 10716. doi: 10.1038/ncomms10716

Guo X., Gao J., Li F., Wang J. (2014) Evidence of horizontal transfer of non-autonomous Lep1 helitrons facilitated by host-parasite interactions. *Scientific Reports*, 4: 5119. doi: 10.1038/srep05119

Gupta S., Gallavotti A., Stryker G.A., Schmidt R.J., Lal S.K. (2005) A novel class of Helitron-related transposable elements in maize contain portions of multiple pseudogenes. *Plant Molecular Biology*, 57(1): 115–127

Halestrap A.P. (2013) The SLC16 gene family – structure, role and regulation in health and disease. *Molecular Aspects of Medicine*, 34*:* 337–349

Han M.J., Shen Y.H., Xu M.S., Liang H.Y., Zhang H.-H., Zhang Z. (2013) Identification and evolution of the silkworm helitrons and their contribution to transcripts. *DNA Research*, 20(5): 471–484

Hong J., Seo J., Kim S., Jung H., Park S., Kim K., Jung C.W., Kim J.S., Park J.S., Kim H.J., Jang J.H. (2015) Mutations in the spliceosomal machinery genes SRSF2, U2AF1, and ZRSR2 and response to decitabine in myelodysplastic syndrome. *Anticancer Research*, 35(5): 3081–3089

Huang H.T., Chen S.M., Pan L.B., Yao J., Ma H.T. (2015) Loss of function of SWI/SNF chromatin remodeling genes leads to genome instability of human lung cancer. *Oncology Reports*, 33(1): 283–291

Inagaki S., Nakamura K., Morikami A. (2009) A link among DNA replication, recombination, and gene expression revealed by genetic and genomic analysis of TEBICHI gene of *Arabidopsis thaliana*. *PLoS Genetics*, 5(8): e1000613. doi: 10.1371/journal.pgen.1000613

Jacobs S., Schilf C., Fliegert F., Koling S., Weber Y., Schürmann A., Joost H.G. (1999) ADP-ribosylation factor (ARF)-like 4, 6, and 7 represent a subgroup of the ARF family characterization by rapid nucleotide exchange and a nuclear localization signal. *FEBS Letters*, 456(3): 384–388

Kapitonov V.V., Jurka J. (2001) Rolling-circle transposons in eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, 98(15): 8714–8719

Kapitonov V.V., Jurka J. (2007) Helitrons on a roll: eukaryotic rolling-circle transposons. *Trends in Genetics*, 23(10): 521-529

König R., Zhou Y., Elleder D., Diamond T.L., Bonamy G.M., Irelan J.T., Chiang C.Y., Tu B.P., De Jesus P.D., Lilley C.E., Seidel S., Opaluch A.M., Caldwell J.S., Weitzman M.D., Kuhen K.L., Bandyopadhyay S., Ideker T., Orth A.P., Miraglia L.J., Bushman F.D., Young J.A., Chanda S.K. (2008) Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell*, 135(1): 49–60

Korfali N., Wilkie G.S., Swanson S.K., Srsen V., Batrakou D.G., Fairley E.A., Malik P., Zuleger N., Goncharevich A., de Las Heras J., Kelly D.A., Kerr A.R., Florens L., Schirmer E.C. (2010) The leukocyte nuclear envelope proteome varies with cell activation and contains novel transmembrane proteins that affect genome architecture. *Molecular & Cellular Proteomics*, 9(12): 2571–2585

Lai J., Li Y., Messing J., Dooner H.K. (2005) Gene movement by Helitron transposons contributes to the haplotype variability of maize. *Proceedings of the National Academy of Sciences of the United States of America*, 102(25): 9068–9073

Li J.M., Lu C.L., Cheng M.C., Luu S.U., Hsu S.H., Hu T.M., Tsai H.Y., Chen C.H. (2014) Role of the DLGAP2 gene encoding the SAP90/PSD-95-associated protein 2 in schizophrenia. *PLoS ONE*, 9(1): e85373. doi: 10.1371/journal.pone.0085373

Li L., Petsch K., Shimizu R., Liu S., Xu W.W., Ying K., Yu J., Scanlon M.J., Schnable P.S., Timmermans M.C., Springer N.M., Muehlbauer G.J. (2013) Mendelian and non-Mendelian regulation of gene expression in maize. *PLoS Genetics*, 9(1): e1003202. doi: 10.1371/journal.pgen.1003202

Liu J., Liu J., Ni J., Leng R.X., Pan H.F., Ye D.Q. (2015) Association of UBASH3A gene polymorphisms and systemic lupus erythematosus in a Chinese population. *Gene*, 565(1): 116–121

Miller W.J., Paricio N., Hagemann S., Martinez-Sebastián M.J., Pinsker W., de Frutos R. (1995) Structure and expression of clustered P element homologues in *Drosophila subobscura* and *Drosophila guanche*. *Gene*, 156(2): 167–174

Murakami Y., Kohyama N., Kobayashi Y., Ohbayashi M., Ohtani H., Sawada Y., Yamamoto T. (2005) Functional characterization of human monocarboxylate transporter 6 (SLC16A5). *Drug Metabolism and Disposition*, 33(12): 1845–1851

Rebollo R., Romanish M.T., Mager D.L. (2011) Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annual Review of Genetics*, 46(1): 21-42

Reversade B., Escande-Beillard N., Dimopoulou A. et al. (2009) Mutations in PYCR1 cause cutis laxa with progeroid features. *Nature Genetics*, 41(9): 1016–1021

Ross K.A. (2014) Coherent somatic mutation in autoimmune disease. *PLoS ONE*, 9(7): e101093. doi: 10.1371/journal.pone.0101093

Sanders M.A., Ampasala D., Basson M.D. (2009) DOCK5 and DOCK1 regulate Caco-2 intestinal epithelial cell spreading and migration on collagen IV. *Journal of Biological Chemistry*, 284(1): 27–35

Shen H., Zheng X., Luecke S., Green M.R. (2010) The U2AF35-related protein Urp contacts the 3' splice site to promote U12-type intron splicing and the second step of U2-type intron splicing. *Genes and Development*, 24(21): 2389–2394

Sorokin A.V., Nair B.C., Wei Y., Aziz K.E., Evdokimova V., Hung M.C., Chen J. (2015) Aberrant expression of proPTPRN2 in cancer cells confers resistance to apoptosis. *Cancer Research*, 75(9): 1846–1858

Tarpey P.S., Smith R., Pleasance E. et al. (2009) A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nature Genetics*, 41(5): 535–543

Thomas J., Phillips C.D., Baker R.J., Pritham E.J. (2014) Rolling-circle transposons catalyze genomic innovation in a mammalian lineage. *Genome Biology and Evolution*, 6(10): 2595–2610

Thomas J., Pritham E.J. (2015) Helitrons, the eukaryotic rolling-circle transposable elements. *Microbiology Spectrum*, 3(4): 1–32

Thomas J., Schaack S., Pritham E.J. (2010) Pervasive horizontal transfer of rolling-circle transposons among animals. *Genome Biology and Evolution*, 2(1): 656–664

Tost H., Callicott J.H., Rasetti R., Vakkalanka R., Mattay V.S., Weinberger D.R., Law A.J. (2014) Effects of neuregulin 3 genotype on human prefrontal cortex physiology. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 34(3): 1051–1056

Tsukamoto T., Hauck N.R., Tao R., Jiang N., Iezzoni A.F. (2010) Molecular and genetic analyses of four nonfunctional S haplotype variants derived from a common ancestral S haplotype identified in sour cherry (*Prunus cerasus* L.). *Genetics*, 184(2): 411–427

Wang K.S., Xu N., Wang L., Aragon L., Ciubuc R., Arana T.B., Mao C., Petty L., Briones D., Su B.B., Luo X., Camarillo C., Escamilla M.A., Xu C. (2014) NRG3 gene is associated with the risk and age at onset of Alzheimer disease. *Journal of Neural Transmission*, 121(2): 183–192

Wei J., Zhai L., Xu J., Wang H. (2006) Role of Bmi1 in H2A ubiquitylation and Hox gene silencing. *Journal of Biological Chemistry*, 281(32): 22537–22544

Wu C., Ma M.H., Brown K.R., Geisler M., Li L., Tzeng E., Jia C.Y., Jurisica I., Li S.S. (2007) Systematic identification of SH3 domain-mediated human protein-protein interactions by peptide array target screening. *Proteomics*, 7(11): 1775–1785

Wu K., Hanna G.L., Easter P., Kennedy J.L., Rosenberg D.R., Arnold P.D. (2013) Glutamate system genes and brain volume alterations in pediatric obsessive-compulsive disorder: a preliminary study. *Psychiatry Research*, 211: 214–220

Xiong W., He L., Lai J., Dooner H.K., Du C. (2014) HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proceedings of the National Academy of Sciences*, 111(28): 10263–10268

Xu L., Hou Y., Bickhart D.M., Zhou Y., Hay el H.A., Song J., Sonstegard T.S., Van Tassell C.P., Liu G.E. (2016) Population-genetic properties of differentiated copy number variations in cattle. *Scientific Reports*, 6: 23161. doi: 10.1038/srep23161

Yang J., Duan S., Zhong R., Yin J., Pu J., Ke J., Lu X., Zou L., Zhang H., Zhu Z., Wang D., Xiao H., Guo A., Xia J., Miao X., Tang S., Wang G. (2013) Exome sequencing identified NRG3 as a novel susceptible gene of Hirschsprung's disease in a Chinese population. *Molecular Neurobiology*, 47(3): 957–966