AGRICULTURAL BIOLOGY,
2022, V. 57, Iss. 5, pp. 832-851
[SEL'SKOKHOZYAISTVENNAYA BIOLOGIYA]

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

UDC 636.012:575.174.015.3

doi: 10.15389/agrobiology.2022.5.832eng
doi: 10.15389/agrobiology.2022.5.832rus

THE SOURCES OF GENOME VARIABILITY AS DOMESTICATION DRIVERS

(review)

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

[1]*Timiryazev Russian State Agrarian University—Moscow Agrarian Academy*, 49, ul. Timiryazevskaya, Moscow, 127550 Russia, e-mail vigvalery@gmail.com, tglazko@rambler.ru (✉ corresponding author);
[2]*Afanas'ev Research Institute of Fur-Bearing Animal Breeding and Rabbit Breeding*, 6, ul. Trudovaya, pos. Rodniki, Ramenskii Region, Moscow Province, 140143 Russia, e-mail gkosovsky@mail.ru
ORCID:
Glazko V.I. orcid.org/0000-0002-8566-8717
Glazko T.T. orcid.org/0000-0002-3879-6935
Kosovskii G.Yu. orcid.org/0000-0003-3808-3086
The authors declare no conflict of interests
Acknowledgements;
The authors express their sincere gratitude to L.M. Fedorova, Ph.D, for the interest shown in the work, fruitful discussion and useful advice in preparing the article for publication.
Received June 22, 2022

Abstract

Plant and animal domestication is the key event in the history of mankind, its mechanisms have attracted the attention of many researchers, especially in recent decades due to the well-known decline in biodiversity, including in agricultural species. According to the definition proposed by Melinda Zeder (M.A. Zeder, 2015), domestication is the stable mutualistic relationship in a number of generations in which the domesticator influences the reproduction of the domesticates, optimizing their lifestyle for the supply of the needing resource to human, and thanks to which the domesticates gain advantages over other individuals of the species. Such relationships are accompanied by interspecific coevolution, they are present not only in humans and domestic species of plants and animals, but also in representatives of wild species, for example, insects, fungi. As a universal feature of domestic species in comparison with closely related wild ones, a high phenotypic diversity is considered, which was noticed by Charles Darwin (Ch. Darwin, 1951). Pairwise genomic comparisons of such species as domestic dog and wolf, wild and domestic cat, domesticated and wild rabbit reveal a relatively increased density of a number of mobile genetic elements in domesticated animals compared to wild ones. In recent years, mobile genetic elements, or transposons (TEs), have been considered as the main factors of genomic transformations, gene, genomic duplications, genomic and gene reconstructions, as well as horizontal exchanges of genetic information (K.R. Oliver, W.K. Greene, 2009). The number of comparative genomic studies of TEs in domesticated species is small, and the role of such elements in domestication, as a rule, is not discussed. However, it can be expected that universal mechanisms of genome variability underlie all evolutionary events, including in response to the new niche-construction during domestication. The presented review systematizes such mechanisms. TEs providing deep genomic transformations, active and passive forms of their interactions with the host genome are considered (K.R. Oliver et al., 2009). Examples of the emergence of new genes based on TEs, such as the synticin gene, are described (C. Herrera-Úbeda et al., 2021), the synaptic plasticity regulator gene *arc* (Activity Regulated Cytoskeleton Associated Protein) (C. Herrera-Úbeda et al., 2021), the *bex* gene family encoding, in particular, the neuron growth factor receptor (E. Navas-Pérez et al., 2020; R.P. Cabeen et al., 2022). Conflict and cooperative interactions with the host genome during retrotransposon movements and different mechanisms of their effects on gene expression profiles are discussed. The participation of TEs in the formation and variability of networks of genomic regulatory elements, in particular microRNAs, is considered. Examples of the involvement of microRNAs in the control and formation of economically valuable traits in domesticated plants and animals are presented. The accumulated data suggest that the leading source of large phenotypic variability of domesticated species is the relatively high saturation of their genomes with mobile genetic elements and, as a consequence, an increase in the variability of genomic regulatory networks in the formation of a new niche during domestication by humans.

Keywords: domestication, genomics, variability, transposons, regulatory networks, microRNAs

Finding the genetic foundations of domestication as the entry of plants and animals into new habitat and reproduction conditions, purposefully formed by man, has theoretical and practical aspects. Theoretical aspect is related to the fact that only the domestication is a direct experimental model of micro- and macroevolution available for human study [1], as pointed out by Charles Darwin [2]. The practical aspect of the problem is due to the fact that a lack of understanding of the mechanisms of domestication does not allow the development of methods for managing genetic flows in domesticated species, which is becoming increasingly important due to the growth of the Earth population, the reduction of fertile lands and biodiversity [3].

To date, the least controversial definition of domesticators formulated by M.A. Zeder [4] describes it as a stable, multi-generational, mutualistic relationship in which humans (domesticators) ensure, to a significant extent, the reproduction control and care of plants/animals (domesticants) for a more predictable supplying with the resource of plant/animal interest, whereby the plant/animal is able to increase its reproductive success compared to individuals not participating in such a relationship, thereby increasing the fitness of both humans and target domesticated species. The concept of the "domestication syndrome" that crosses taxonomic differences has been formed. For example, in annual plants, this concerns traits associated with seed germination and spreading, e.g., the changes in germination rate, seed size, seed shedding, wall thickness, as well as in the timing and morphology of spreading mechanisms [5, 6]. In animals, these traits include characteristics of the neural crest, which are associated with the behavior of animals (in particular, social activity), adaptive potential, including indicators of fertility, and the variability of those traits that are required by the domesticator and are associated with animal productivity [7].

It should be noted that quite often, both in plants and animals, the characteristics that are classified as domestication syndromes are associated with polymorphism of different genes involved in similar metabolic pathways, which is expected, since most of the signs of domestication syndrome are polygenic [8].

The accumulated data indicate that the only common feature for all domesticated species is a relatively increased phenotypic variability, which serves as a source of response to the factors of natural and artificial selection that appear when representatives of a particular species are involved in the sphere of human interests. In this case, the main issue in studying the mechanisms of domestication becomes the elucidation of the genetic basis for such a large variability [9].

Mobile genetic elements (transposons, transposable elements, TEs). In recent years, mobile genetic elements, or transposons (TEs), have been considered as the main factor in genomic transformations, in gene and genomic duplications, genomic and gene reconstructions, as well as in horizontal exchanges of genetic information [10-12]. In some studies, taking into account the phylogenetic and biological similarity of TEs to viruses, it is proposed to designate the totality of such sequences in genomes as endovir [13]. A certain correlation was found between the density of the genomic distribution, the activity of TEs, and the intensity of speciation (abundance of species) in different taxa [14]. However, the possibility of involvement of TEs in increased phenotypic variability in domesticated species is generally not considered.

Mobile genetic elements are divided into two main classes. Class I consists of retrotransposons whose distribution through the genome involves RNA, while autonomous retrotransposons have a gene encoding reverse transcriptase. Class II includes DNA transposons that do not use RNA as an intermediate for transposition. Each class has autonomous and non-autonomous members. As a result, the following main classification has been adopted for mobile genetic elements

(transposons) (MGEs, TEs) [15]. Class I which includes retrotransposons is represented by autonomous endogenous retroviruses (ERV) with long terminal repeats (LTR), long interspersed nuclear elements (LINE), as well as non-autonomous elements, the short dispersed nuclear elements. elements (short interspersed nuclear elements, SINE) and composite retrotransposon (recombination products between retrotransposons and microsatellites, SVA). Class II represented by DNA transposons includes autonomous elements that carry a DDE-amino acid motif typical for transposase/integrase of most families of autonomous DNA transposons, except for Helitrones (the DNA transposons that replicate via the ring model and use an enzyme with endonuclease and helicase domain), as well as a non-autonomous element, a miniature inverted repeat transposable element (MITE).

Mobile genetic elements are called drivers of evolution, and their significant contribution to evolution, in particular, in primates, is known [16]. TEs have been found in the genomes of various taxonomic groups, from bacteria to mammals. Their ubiquitous presence is due to a pronounced tendency to spread along the genome, as well as to colonize other genomes. Some TEs (e.g., SINE) may arise spontaneously from non-mobile DNA sequences in the genome, while others may be horizontally transferred between species. TEs have an ancient origin dating back to prokaryotes. DNA-TEs (class II) are associated with sequences of bacterial inserts, retro-TEs are associated with introns of the second group [16, 17].

Some TEs appear to have been present in eukaryotes since the very beginning of their existence, perhaps even more than a billion years ago. TEs often make up the largest, if not the largest, portion of the eukaryotic genome. For example, sequenced mammalian genomes are composed of at least a third of TEs in non-primates and about half in primates. TEs appear to be an important determinant of genome size, with organisms with very large genomes (e.g. plants) often having a much higher abundance of TEs (> 60%) compared to species with relatively small genomes (yeasts, nematodes, insects and birds), in which the proportion of TEs is significantly lower.

Thanks to whole genome sequencing powered by software for the analysis of nucleotide sequences, data on the presence of TEs in different taxa have been accumulated (Fig. 1) [18]. These results indicate a wide distribution of TEs and that their number varies significantly between species, making an obvious contribution to the differences in the size of their haploid genome. Significant differences can be found even between closely related organisms. For example, in the genus *Entamoeba*, the *E. histolytica* genome is ~ 20% TEs, while the *E. dispar* genome is ~ 10%. Similarly, in the genus *Oryza*, the genome of the wild rice species *O. australiensis* more than doubled in size for 3 million years due to amplification of TEs, resulting in approx. 3 times the genomes of some of its closest relatives. On the contrary, a number of taxonomic groups show only minor differences in the content of TEs, which may reflect certain restrictions on the size of the genome or, conversely, a relatively high tolerance for genome expansion. For example, birds have a relatively conserved genome size (possibly due to the metabolic costs associated with active flight), despite the fact that TEs are active in most studied species of this taxonomic group. In salamanders, on the contrary, extreme and independent amplification of TEs occurred, which led to the formation of giant genomes (see Fig. 1).

It is interesting to note that the largest genome size and, accordingly, the contribution of TEs to it was found in lungfish and amphibian, that is, species in which aromorphosis (survival in two environments) is realized (see Fig. 1). Also noteworthy are the data on the unequal representation of different TEs variants in plants and animals: in plants, endogenous retroviruses are more common than in animals, and in the latter, the frequency of occurrence of SINEs is higher [18]

(see Fig. 1).

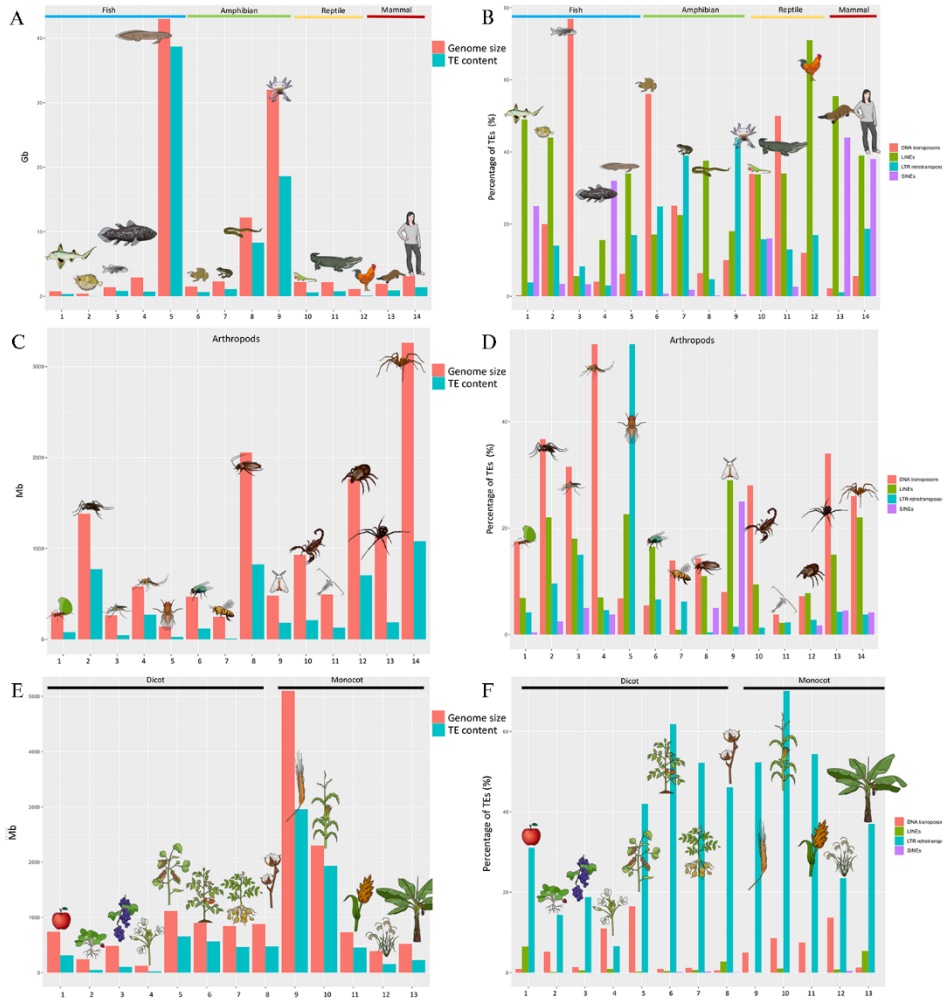

Fig. 1. The prevalence of TEs transposons in the genomes of various species [18].

A. Size of the genome (red) and the TEs it contains (blue) in vertebrates from left to right: 1 — *Callorhinchus milii*, 2 — *Tetraodon nigroviridis,* 3 — *Danio rerio*, 4 — *Latimeria chalumnae*, 5 — *Neoceratodus forsteri*, 6 — *Xenopus tropicalis*, 7 — *Nanorana parkeri*, 8 — *Chthyophis bannanicus*, 9 — *Ambystoma mexicanum*, 10 — *Anolis carolinensis*, 11 — *Alligator mississipiensis*, 12 — *Gallus gallus*, 13 — *Ornithorhynchus anatinus*, 14 — *Homo sapiens.*

B. Proportion (%) of different types of TEs: red — DNA transposons, green — LINE, lilac — SINE, blue — endogenous retroviruses (ERV) of vertebrates (see species A).

C. Genome size (red) and TEs it contains (blue) in arthropods:1 — *Acromyrmex echinatior*, 2 — *Aedes aegypti*, 3 — *Anopheles gambiae*, 4 — *Culex quinquefasciatus* (москиты), 5 — *Drosophila melanogaster* (плодовая мушка), 6 — *Lucilia cuprina*, 7 — *Apis mellifera*, 8 — *Blattella germanica*, 9 — *Bombyx mori*, 10 — *Centruroides exilicauda*, 11— *Eurytemora affinis*, 12 — *Ixodes scapularis*, 13 — *Latrodectus hesperus*, 14 — *Loxosceles reclusa*.

D. Proportion (%) of different types of TEs in arthropods: red — DNA transposons, green — LINE, lilac — SINE, blue — endogenous retroviruses (ERV) of vertebrates (species see C).

E. Genome size (red) and TEs it contains (blue) in plants: 1 — *Malus domestica*, 2 — *Fragaria vesca*, 3 — *Vitus vinifera*, 4 — *Arabidopsis thaliana*, 5 — *Glycine max*, 6 — *Solanum lycopersicum*, 7 — *Solanum tuberosum*, 8 — *Gossypium raimondii*, 9 — *Hordeum vulgare*, 10 — *Zea mays*, 11 — *Sorghum bicolor*, 12 — *Oryza sativa*, 13 — *Musa acuminate.*

F. Proportion (%) of different types of TEs in plants: red - DNA transposons, green - LINE, lilac - SINE, blue - endogenous retroviruses (ERV) of vertebrates (see species E).

Active and passive mechanisms of interaction of transposons with the host genome, the emergence of new genes. TEs,

as already noted, are powerful factors in the evolution of the genome (and hence phenotypic diversity) because they are able to induce genetic changes on a large scale. TEs can change the way genomes work both actively and passively. Species with active TEs or abundant uniform inactive TEs that can passively influence genome function by inducing ectopic recombination are potentially fertile and adaptable. Conversely, taxa deficient in TEs or having heterogeneous populations of inactive TEs are often well adapted in their niche but tend to be stagnant for a long time and may be at risk of extinction due to a lack of adaptability to changing conditions or diversification [17].

The mechanisms of influence of TEs on the genome functioning include a number of genomic effects of TEs, divided into active and passive influences [17].

Approx. 50 cases of neogenes have been described the nucleotide sequnsces of which are largely the TEs derivatives [17]. These genes include *TERT*, *CENPB*, *RAG1/2*. These neogenes made possible some extraordinarily complex evolutionary events that otherwise might not have occurred, such as the formation of recombination signaling sequences involved in rearrangements of the V(D)J Ag receptor. These sequences, like the RAG1/2 recombinase genes themselves, seem to be descended from the ancient DNA-TE.

*Exons (partial exons), protein-coding region*s. TEs often form independent exons within genes [17].

Extragenic sequences. TEs contribute to the formation of various extragenic sequences, such as centromeres, telomeres in Drosophila and some protozoa, sites of DNA replication initiation in yeast, regions associated with the framework of the interphase nucleus, and chromosome matrix in humans [17].

*Direct contribution to gene regulatio*n. *Full and partial promoters, enhancers, silencers.* Many TEs control gene expression, often tissue-specific [17]. In addition to influencing individual genes, TEs, apparently, turned out to be mobile carriers of ready-made promoters (enhancers) for the wide distribution of discrete regulatory elements throughout the genome. This provides a regulatory network by which an entire set of genes can be co-regulated to create new pathways for cellular development or improve existing ones.

Regulatory (micro)RNAs. Many exonized TEs encode miRNAs. Fifty-five human miRNA genes have been identified that derived from TEs and are capable of regulating thousands of genes [17].

*Indirect regulation: retrotransposition/transduction of gene sequences, gene duplication, exon shuffling, distribution of regulatory elem*ents. Some families of retro-TEs (e.g., LINE and LTR elements) tend to transduce host DNA due to their weak transcriptional termination sites. Gene duplication can also occur through the assignment of the TEs (*pol* reverse transcriptase gene) retrotransposition apparatus by host mRNA transcripts. In human, there are more than 1000 retrogens that have arisen in this way, some of which have developed very useful functions, for example, the *GLUD2* (*Glutamate Dehydrogenase 2*) gene is important for the utilization of the main excitatory neurotransmitter glutamate during neurotransmission [17].

Passive mechanisms. Stimulation of duplications (or losses) of DNA sections, gene duplication, exon duplication, segmental duplication. The mere presence in the genome of a large number of inactive TEs-like elements creates many highly homologous sites, which typically causes ectopic (non-allelic) DNA recombination due to such homology. This probably explains most of the ongoing effects of TEs in organisms with low TEs transcription activity, high TEs abundance and low diversity [17].

DNA duplication events are especially important in evolution because they

create functional redundancy and the potential for enhancing gene function and/or expression.

Stimulation of karyotypic changes by ectopic recombination, intra- and interchromosomal DNA rearrangements. TEs can passively support major chromosomal rearrangements by creating highly homologous regions scattered throughout the genome that are prone to ectopic recombination. For example, Alu-mediated translocation t(11;22)(q23;q11) is the most common constitutional translocation in humans [17].

TEs and new host genes. Active influences of TEs on evolutionary events include, in particular, examples of the incorporation of proviral genes in the host genome. For example, approximately 8% of the human genome consists of endogenous retroviral sequences. In the course of evolution, most of the genes from these sequences lost their function, but some of them were captured and "domesticated" via so-called exaptation. Among the domesticated viral genes, there is a group of syncytin genes that most clearly influenced the evolution of mammals [19]. Syncytins are captured viral proteins, products of the envelope gene of hereditarily endogenized retroviruses. The envelope glycoprotein (*Env* gene) is crucial in the process of virus entry into the cell and induces the fusion of the virion envelope with the plasma membrane of the cell. Several *Env* genes are found in the human genome, but only two of them, which induce the formation of syncytium, have placental-specific expression. Since it is the presence of the placenta that underlies the allocation of placental mammals to a separate taxonomic group, one would expect that the syncytins responsible for the development of this unique organ, which is formed only during pregnancy, would be orthologues in different species, but this is not the case. Primate and mouse syncytins are not orthologues, and there is evidence pointing to independent uptake events of the respective provius genes in the ancestors of each clade, as well as in *Scincidae* of the genus *Mabuya*. In fact, in mammals, gene capture events of various retroviruses can be associated with four main types of placental structures. It can be seen that the differences between the lizard placenta of the genus *Mabuya* and the mammalian placenta are due to different *Env* genes uptake events [19]. It is important to emphasize that the formation of syncytium with the participation of syncytins can occur due to different syncytins, which provide intercellular fusions of the envelope proteins of many retroviruses (including HIV, bovine leukemia virus), which is necessary for the reproduction of the virus [19].

Another example is the *Arc* gene encoding the Activity Regulated Cytoskeleton Associated Protein which is of particular interest because it seems to be the main regulator of synaptic plasticity [19]. The Activity Regulated Cytoskeleton Associated Protein is released from neurons in extracellular vesicles that mediate the transfer of *Arc* mRNA to new target cells where *Arc* mRNA can be translated. It was found that this protein is necessary for forms of long-term memory dependent on its synthesis and is involved in the development of depression in humans. The protein accumulates in weak synapses (probably to prevent their undesirable strengthening), participates in postsynaptic transport and processing of beta-amyloid A4 (APP). In addition to its role in synapses, it is involved in the regulation of the immune system: it is specifically expressed in migrating dendritic cells, thereby participating in the activation of T cells [19].

The authors of these studies [19] note that *arc* gene products mediate intercellular communication and synaptic plasticity through extracellular vesicles and are largely homologous to group-specific retroviral antigen (Gag) polyproteins. In retroviruses, capsids are essential for cellular infection and their assembly is mainly mediated by Gag. The similarity between Arc and Gag is not limited to amino acid sequence, as Arc is able to spontaneously assemble into a capsid-like

structure. In fact, Arc not only forms these capsid-like structures, but also encapsulates any mRNA present. Such processes ensure the movement of RNA molecules between the cells of the nervous system. Regarding the uptake and evolution of this viral protein, phylogenetic analyzes have identified at least two independent uptake events that occurred in tetrapod ancestors and in schizophorans. In both cases, arc co-optation resulted in similar functions of the RNA transporter protein in the nervous system. In both lines, the Ty3/gypsy retrotransposon was the closest to the putative initial ancestral variant, but according to the homology of the Ty3/gypsy sequences, tetrapods clustered with fish, while flies clustered with insects, which, apparently, indicates the independent origin of arc from retrotransposons. Ty3/gypsy in each line, despite the significant homology of its product with the Gag protein [19].

The results of transposon domestication also include the appearwnce of a new *Bex/Tceal* cluster (consisting of 14 genes, located in the X chromosome of the placental ancestor) after the divergence of the marsupial and placental clades. The *bex* gene family, encoding, in particular, the neuronal growth factor receptor, emerged through the introduction of placental retrotransposons HAL1b, L1ME-like, and the Hnmph1 endogenous retrovirus in placental progenitors [20].

Analysis of the functional activity of the *Bex/Tceal* genes was performed in vitro and on mouse lines mutant for the *bex3* gene. The lines homozygous for the mutation of this gene showed deviations in the morphology of the skull, the size of the cerebellum and brain decreased, which may be associated with the behavioral defects observed in mutant mice. They showed impairments in social interactions, nest building, working memory, and object recognition memory. This mutant phenotype may mean that the *bex3* gene subproduct interferes with the interaction of the TSC1/2 complex (TSC Complex Subunit 1, a tumor-suppressing gene encoding the growth-inhibiting protein hamartin) with the target of rapamycin in small-feeding mTORC1/mTORC2, inhibiting this pathway. mTORC2 belongs to the phosphatidylinositol kinase family. They mediate cellular responses to stress (in particular, in response to DNA damage). This kinase is a component of two different complexes, the mTORC1 which controls overall protein synthesis, cell growth and proliferation, and mTORC2 which serves as a regulator of the actin cytoskeleton and promotes cell survival [21].

mTOR inhibitors are used in organ transplantation as immunosuppressants and are being evaluated for therapeutic potential in SARS-CoV-2 infections [20]. Mutations in the *mTOR* protein gene are associated with Smith-Kingsmore syndrome (characterized by macrocephaly, mental retardation, seizures) and somatic focal cortical dysplasia type II [21]. The *ANGPTL7* gene (Angiopoietin Like 7, involved in the negative regulation of carpillar network development) is located in the *mTOR* intron. It has been suggested that *ANGPTL7* dysregulation under the influence of *Bex3* is associated with autism-type disorder in humans.

The following scheme was proposed to describe the stages of gene cooptation [19]. At the first stage, the proto-BGW motif (Bex/GASP/Wex element common to the genes *bex*, *gasp*, and *wex*) [21] was present in the X chromosome of the placental and marsupial ancestor in a position upstream of the alpha-galactosidase gla gene promoter (pα). At the second stage, in the placental lineage, the retrotranscribed endogenous retrovirus Hnrnph1 was inserted next to the *bgw* motif in positions above *gla*, which led to the appearance of the retrogen *hnrnph2*. At the third stage, the region containing the co-opted motifs *bgw* and *hnrnph2* underwent duplication, and retrotransposons similar to HAL1b and L1ME were inserted nearby. At the fourth stage, *bgw* and the open reading frame (ORF) appeared upon the introduction of retrotransposons formed the nucleotide sequences that corresponded to proto-*Bex/Tceal* with the preserved YY1 binding site from

HAL1b. At the next stage, the BGW motif and the YY1 binding site of the *Bex/Tceal* gene were duplicated in a position upstream of the retrocopy of the *armc10* gene. This led to the appearance of the inherited *armcX* gene belonging to the family of signaling genes located on the X chromosome, which encode the Armadillo repeat in proteins and possess tumor- overwhelming activity. At the final sixth stage (even before the diversification of the placental lineage of mammals), the *bex/tceal* and armcx gene families expanded, forming the BGW cluster.

The X chromosome contains a disproportionate number of genes associated with mental functions, as evidenced by the prevalence of mental retardation in men. However, all nine X-linked genes that, when mutated, lead to mental impairment have orthologues in fish or even earlier eukaryotes. It also turned out that all bex, wex, and gasp genes are expressed in the brain [21]. Thus, these placental species-specific genes can be considered as possible candidates for the adaptive evolution of the neocortex, a region of the forebrain that is unique in mammals [22].

The presence of the conserved BGW element within the 5'UTR of the bex, wex, and gasp genes suggests its involvement in the regulation of translation [21]. The rationale is that, as previously shown, the rate of translation is influenced by regulatory regions, including the consensus sequence, secondary structures preceding the AUG (site of translation initiation), internal ribosome entry sites, and the site of recognition of sequences specific for regulatory factors, such as protein or RNA. The presence of translational control for the bex, wex, and gasp genes may indicate that the proteins encoded by them are used under certain physiological conditions, at certain developmental stages, or in subcellular compartments. Another possible role for the BGW element could be the regulation of alternative splicing [21].

The same authors identified genes with a coding sequence overlapping more than 50% with annotated TEs and present in more than one species (28 genes in humans and 9 genes in mice) [21].

In plants, various genes have been described whose mutations are involved in artificial selection and are caused by the incorporation of TEs either into coding sequences or into regulatory motifs [23-25]. The same events are found in the genes of different animal species, in particular in chickens [26, 27].

Retrotransposons LINE and SINE in genomic changes. The scale of involvement of retrotransposons in genomic rearrangements was recently revealed by comparing the results of whole genome sequencing of the human genome [28]. This paper considers the transduction of human genomic elements associated with such retrotransposons as L1 and SVA. During their transcription, the signal of its termination at the 3'-end can be ignored by RNA polymerase. As a result, transcription is completed at the host genomic element, creating a chimeric transcript. The work analyzed 3202 sequenced genomes from 26 population groups from different countries and identified 7103 polymorphic L1 and 3040 polymorphic SVA. As a result, 268 and 162 transduction variants of 3'-regions from 7 to 997 nucleotides in length were found involving sequences homologous to L1 and SVA, respectively. In the chromosomes X, 6 and 7, specific loci with the most widely represented L1 and SVA were identified, which, among other things, determined the largest number of transductions.

Of particular importance are the processes associated with TEs which lead to a change in the copy number of genes (loss or increase in their number) [29].

Interactions between TEs and the host genome are complex and include a large number of different mechanisms in each specific case [30]. Multiple variants of TEs insertion into the host genome and interactions with its various genomic elements can lead to different consequences, including both conflicting and

favorable mutually beneficial relationships [31]. Examples of conflicting relationships include insertion of TEs into exons, which leads to frameshift mutations and disturbances in protein structure and function. TEs are also able to increase genomic instability by forming structures that create regions of homology along chromosomes, which leads to chromosomal rearrangements, i.e., duplications, deletions, inversions and translocations. When TEs are inserted into regulatory regions, such as 5'-, 3'-regions or gene introns, epigenetic modifications occur, which cause inappropriate activation or suppression of gene expression [31].

An example of cooperative relationships [31] is the use of TEs by the host genome to generate new regulatory signals or coding sequences. This process is called molecular domestication [31]. TEs are able to introduce new enhancer sequences for transcription factors that change the spatiotemporal regulation of gene expression. In Drosophila, after the loss of telomerase, autonomous (LTR-deprived) retrotransposons HeT-A, TAHRE, and TART are actively involved in telomere maintenance. TEs can help maintain the architecture of the genome by providing binding sites for the CTCF protein, which is responsible for the creation of topologically associated domains (TADs). Topologically associated domains are directly involved in the assembly of gene expression programs [32)], which significantly depends on the distribution and movement of various TEs throughout the genome.

It should be noted that LINEs are widely represented in almost all eukaryotes [33]. LINE1 long dispersed nuclear elements are the most successful family of TEs in terrestrial mammals. The length of LINE1 varies within 6 kb. LINE1s carry the genes for two main proteins, ORF1 and ORF2 (similar to the *gag* and *pol* genes of exogenous retroviruses), which are involved in transposition mechanisms. The product of ORF1 is an RNA-binding protein, ORF2 has endonuclease (EN) and reverse transcriptase (RT) enzymatic activities. Despite their abundance, most LINE1 elements are not transposed due to the accumulation of mutations. Of the approximately 8,868,000 and 599,000 LINE1 elements in the human and mouse genomes, respectively, only 80-100 and approximately 2300 LINE1s are transpositionally competent [34].

Insertion of LINE1 occurs mainly through endonuclease-dependent reverse transcription. LINE1 RNA forms a complex in the cytoplasm with several ORF1p homotrimers and at least one ORF2p dimer to form a ribonucleoprotein (LINE1 RNP). The LINE1 RNP enters the cell nucleus, where the ORF2p endonuclease releases the 3'-hydroxyl group. The free 3'-hydroxyl group is then used by the ORF2p reverse transcriptase as a primer for the synthesis of the LINE1 cDNA, starting from the LINE1 polyA-mRNA tail. The non-autonomous short dispersed genomic element (SINE) polyA-mRNA tail can compete with LINE1 polyA-mRNA for the LINE1 ORF2p reverse transcriptase, exploiting the LINE1 mechanism for transposition. In addition, LINE1 ORF2p can also retrotranspose unique protein-coding mRNAs and small nuclear RNAs.

LINE1 retrotranspositions depend on other cellular proteins. A number of positive transposition regulators have been identified, including, in particular, nucleolin and heterogeneous nuclear ribonucleoproteins (hnRNPs), mitogen-activated protein kinases, and cyclin-dependent kinases. To complete the LINE1 insertion, a DNA repair mechanism is required. There are several mechanisms for limiting LINE1 transposition in mammalian genomes at the transcriptional, post-transcriptional, and posttranslational levels. Transcription of LINE1 is downregulated by methylation of CpG DNA regions and repressive histone modifications at the LINE1 promoter. Various KRAB-ZFPs (zinc finger DNA interaction proteins) selectively recognize ERV and LINE1 and recruit KAP1 (heterochromatin protein). Post-transcriptional suppression of LINE1 mRNA is mediated through

RNA interference by small RNAs. LINE1 retrotransposition may also be post-translationally limited by a number of interferon-stimulated genes [34].

The estimated frequency of new LINE1 inserts is approximately 1 per 100 births in humans and 8 in mice. The possibilities of LINE1 influence on the work of the genome are numerous and include, in particular, the induction of genomic insertions, the control of retrotransposition of SINE/Al elements (which themselves act as the main modulators of genomic variability). The endonuclease activity of LINE1 ORF2 is a potential inducer of mutagenic effects, regardless of retrotransposition. The LINE1 and SINE elements can contribute to changes in the number of repeats in microsatellite loci, especially those rich in AT, which are abundant in genomes.

Thus, due to the abundance in genomes, LINE1 and SINE elements can induce large-scale genomic changes, such as duplications and inversions. It is generally accepted that, on average, any two human haploid genomes differ by approx. 1000 TEs, mainly from the LINE1 or Alu families [34].

Epigenetic effects of LINE and SINE transpositions. The inherent self-assembly property of L1 (LINE) and B1/Alu (SINE) repeats provides multiple trigger points for the formation of nuclear subregions in the interphase nucleus. Repetitive DNA sequences also serve as anchor sites necessary for the functioning of transcription mechanisms, for the binding of regulatory proteins and RNA. TEs can influence gene expression profiles due to the fact that their DNA or RNA transcripts can interact with DNA and/or RNA binding proteins. The accumulation of molecules can lead to aggregation of repeats containing the same type, forming separate compartments in the nucleus and, thus, changing the genome packaging. Nuclear segregation of compartments rich in L1 or B1/Alu can be further enhanced by binding their DNA sequences to subnuclear structures (nuclear speckles or nucleolus, respectively) that serve as scaffolds to stabilize the nuclear architecture [32].

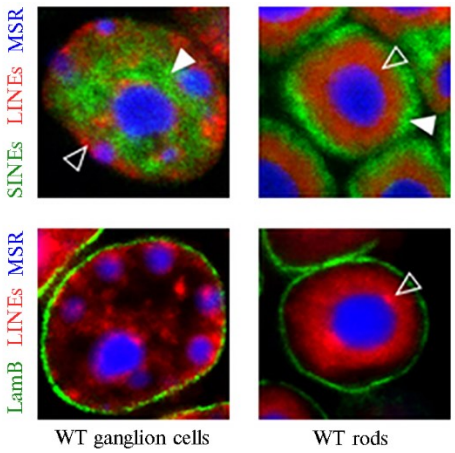

Fig. 2. Distribution of heterochromatin markers LINE (red)**, euchromatin SINE** (green)**, microsatellite sequences MSR** (blue)**, LamB** (lamina protein - a protein network underlying the nuclear envelope, green) **in the nuclei of nerve ganglion cells** (on the left, WT ganglion cells) **and in cylindrical cells of retinal photoreceptors** (on the right, WT rods) (35).

Apparently, the most illustrative example of the involvement of tandem and dispersed repeats in the regulation of changes in gene expression programs through dynamic changes in the architectonics of the interphase nucleus was described in a study performed on the interphase nuclei of cylindrical photoreceptor cells in the retina of nocturnal mammals.

They had an inverted pattern of localization of heterochromatin and euchromatin compared to other nuclei: heterochromatin was located inside, euchromatin was located on the periphery of the nucleus under the lamina (Fig. 2) [35]. The authors of this study came to the conclusion that a significant contribution to such differences can be made, in particular, by contacts between homologous dispersed repeats localized in different regions of chromatin.

In a group of closely related species of the bristled rat, pronounced interspecies differences were found in the chromosomal positioning of the LINE and SINE retrotansposons, which made it possible to suggest their involvement in

speciation [36].

It is interesting to note that the involvement of the mutual positioning of chromosomes and their regions (domains) in interphase nuclei in interspecific differentiation was discussed quite a long time ago by V.N. Stegniy [37] on the example of polytene chromosomes of the malarial mosquito. He also revealed the participation of mobile genetic elements, in particular LINE, in the contacts of chromosomes with the lamina of the interphase nucleus [38].

The wide distribution of LINE1 elements in mammalian genomes makes them good candidates for centers of local chromatin organization and higher-order chromatin architecture. Interestingly, the LINE1 and SINE elements, whose number in the genome depends on the LINE1 ORF2 activity, have opposite genomic distributions. LINE1s are more common in silent, almost gene-free, heterochromatic, AT-rich regions, while SINE elements predominate in gene-rich, expressed, euchromatic, GC-rich regions. A similar general subdivision of genomic regions is observed between compartments B and A (non-expressed and expressed regions of the genome, respectively) in the interphase nucleus. It is assumed that the presence of LINE1 and SINE elements correlates with the presence of compartments B and A, respectively, and LINE1 and SINE can directly participate in the formation of chromosomal regions of these two compartments. The high association of RNA repeat sequences, including LINE1 RNA, with chromatin, and evidence that chromatin-associated RNA promotes global chromatin organization, support this hypothesis [35].

Iinterestinly, it is shown that the regulatory effects of LINE insertions into the host genome may differ depending on the function of the genes [39]. A comparative analysis of the localization of LINE1 and LINE2 in regulatory sequences (promoters, enhancers) in genes whose expression is tissue-specific and common genes (housekeeping genes) revealed a relatively increased rate of evolution of tissue-specific genes, in which the inserted LINE1 retrotransposon is actively involved. The more ancient retrotransposon LINE2 is more often present in the regulatory sequences of housekeeping genes [39].

TEs, in particular LINE1 and various ERVs, are globally demethylated and expressed during major waves of epigenetic erasure that occur during preimplantation and germline development [34, 39]. Similarly, weakening of TEs repression in somatic cells upon demethylation of H3K9me2 occurs in artificially induced pluripotent stem cells [39]. Many families of TEs act as important cis-elements of gene regulation [39]. However, it becomes clear that TEs gene products also play a significant role in the early development of an organism. RNA derived from the transcription of elements of the endogenous retrovirus HERVH is important for maintaining the undifferentiated state of human embryonic stem cells (perhaps by acting as a long non-coding RNA, the lncRNAs involved in regulatory networks of gene expression). Rec protein encoded by the endogenous retrovirus HERVK is highly expressed in human embryos where it increases antiviral resistance by upregulating the IFITM1 (interferon-induced transmembrane protein 1) proteins [39]. Rec also forms complexes with a subset of endogenous RNAs, thus regulating their binding to ribosomes and, as a result, protein expression levels, which may be important for early embryonic development.

It is assumed that the mechanisms of limiting the mutagenic activity of TEs can be considered from the point of view of the balance between the beneficial and negative functions of TEs, which ensures the reproducible development of the organism up to reproductive age, while maintaining the ability to quickly generate genetic novelty in a changing environment. However, such metastability can in some cases lead to disease as a result of adverse retrotransposition events. In other words, the potential reactivation of TEs in adult somatic cells and their association

with diseases such as cancer may be the price to pay for the importance of TEs for the development and evolution of organisms [40].

Endogenous retroviruses. Endogenous retroviruses exhibit many functions that affect the normal biology of host cells [41]. Some of these functions are directly related to interactions with exogenous retroviruses. These may include receptor interference, immune self-tolerance, recombination, and the simultaneous action of restricting and stimulating exogenous retroviral infection using various mechanisms.

Endogenous retroviruses (ERVs) are divided into different classes and are direct descendants of exogenous retroviruses. Retroviruses (exogenous and endogenous) are divided into three classes (I, II and III) each of which corresponds to the expanded groups ERV1, ERV2 and ERV3. Class I retroviruses are endogenous (ERV1) and exogenous gamma and epsilon viruses, class II includes ERV2 and exogenous alpha, beta, delta retroviruses and lentiviruses, class III are spumaviruses and ERV3 (ERV3/ERVL mammals). The ERV4 group now includes the previously identified and described CrocERV in crocodilians, as this ERV differs from any other known groups of retroviruses and, at the same time, shares homologous regions with all [42].

Invertebrate eukaryotic genomes also contain retrovirus-like LTR-containing elements, the so-called LTR retrotransposons. They are divided into three groups, the *Pseudoviridae* (Copia/Ty1) group, found in plants and fungi, the *Metaviridae* (Gypsy/Ty3) group, which is also found in plants and fungi, and the *Semotivirus* (Bel/Pao) group, found in metazoans. The most diverse group is the *Metaviridae*, which includes about 10 subgroups. One of them (chromoviruses) has a wider range of hosts (plants, fungi and vertebrates). Chromoviruses got their name because their pol gene encodes an integrase with a chromodomain (chromatin organization modifier domain), the nucleosome-binding portion of which can mediate a provirus insertion specific to the host genome sequence [43].

There are regiond of homology between various taxonomically distant retroviruses, including pathogenic ones, in particular with human immunodeficiency viruses, bovine leukemia viruses, chromoviruses (in centromeres more often in plants), polychaete retroviruses, which indicates a wide scale of recombinations between these viruses [43].

A unique example of the integration of an exogenous retrovirus into the koala genome, which occurred under the supervision of researchers, has been described [44]. It is clearly shown how endogenous retroviruses arise as a result of infection of cells that form gametes. During retroviral infection, the RNA of the viral genome undergoes reverse transcription into proviral DNA, which is subsequently integrated into the host genome. Sometimes integration can also occur in primary embryonic cells, which can lead to the formation of an embryo and then progeny with an integrated provirus in all cells. Over many generations of the host, proviral DNA undergoes significant mutational changes (single nucleotide polymorphisms, insertions and deletions), which usually result in an inability to produce an infectious virus [45].

It should be noted that TEs are widely involved in epigenomic variability, including changes in methylation pattern, histone modification, miRNA formation, and transgenerational inheritance [46, 47]. It can be expected that the construction of new niches in which humans and domesticated species participate contributes to the activation of TEs and the formation of new regulatory networks based on them [48-50].

Non-coding RNA (ncRNA), miRNA. *Major sources of non-coding RNA.* In addition to the functions of messenger, ribosomal, and transfer RNAs, many other RNAs (non-coding RNAs, ncRNAs, ncRNAs) play a regulatory role

in eukaryotes [51]. They act as regulators of the functional activity of nucleic acids, recognizing their specific target sequences by homology, and are involved in the regulation of growth, development, and stress reactions in animals and plants. Regulatory ncRNAs ranging from short to long (lncRNAs) control a wide range of biological processes. Depending on the mode of biogenesis and function, ncRNAs have evolved into various forms, including microRNAs (microRNAs), small interfering RNAs (siRNAs), microRNA variants (isomiRNAs), long non-coding RNAs (lncRNAs), circular RNAs (circRNAs), and derivatives of non-coding RNAs [51]. One of the elements of regulatory networks is microRNA, which has a hairpin structure and is a derivative of TE and other genomic elements [51].

In eukaryotic cells, gene expression is regulated at several levels. At the post-transcriptional level, regulation is modulated by various trans-acting factors that bind to specific sequences in mRNA. This affects various processes such as the rate of degradation and the efficiency of mRNA translation, splicing, and localization [52]. MicroRNAs in combination with the Argonaute enzyme form an RNA-induced silencing complex (miRISC), which uses a complementary nucleotide sequence to suppress the target transcript. RNA-binding proteins (RBPs) promote post-transcriptional control by affecting mRNA stability and translation when bound to cis elements in the mRNA transcript. RBPs influence gene expression through miRISC or its interaction with the target mRNA [52].

LINE1 and endogenous retroviruses (ERVs) induce the formation of dsRNAs formed from convergent transcripts or hairpin structures as a result of end-to-end transcription of head-to-head/tail-to-tail elements. Alu elements (SINEs) are much shorter and form hairpin structures as well as open dsRNA hybrids [53].

MicroRNAs have become the object of in-depth research in recent years. A large amount of data has been accumulated on the involvement of various microRNAs in the regulation of developmental stages and responses to stress in plants [51], as well as in the control of feed payment in some main species of farm animals [54, 55].

Epigenetic mechanisms such as DNA methylation, histone modification, and expression of non-coding RNAs appear to be particularly important in the response of multicellular organisms to environmental stressors [54]. In addition, abiotic stress, such as heat shock, can induce suppression of TEs, causing destruction of RISC by the inducible Hsp70 chaperone, which directs the RISC complex to lysosomes.

In connection with the involvement of miRNA molecules in epigenetic variability, studies of their genomic distribution in a number of domesticated mammalian species (cow, dog, horse, pig, and rabbit) were performed. Data have been obtained on the predominant localization of miRNAs in introns and intergenic spaces [55]. It has been found that in the process of evolution, new miRNAs appear and the existing ones are lost [55].

The results of a comparative analysis of the so-called young and old orthogroups (common in origin) of miRNAs in different tissues in the studied species indicated that the expression of young groups has more pronounced tissue-specific features compared to old groups [55]. Approximately 20% of the new orthogroups are localized to the brain, and their target targets appear to be enriched in genomic elements to ensure neuronal activity and processes of their differentiation.

Changes in the microRNA regulatory network were also found during domestication. Thus, a comparison of the extinct predecessors of cattle *Bos primigenius* with representatives of the modern species *B. taurus* showed that noncoding miRNAs became key regulators of the spatiotemporal expression of target genes that control the growth and development of mammals [56]. During the domestication process, the selection of mutational changes in microRNAs and/or microRNA

binding sites could provide a mechanism for generating some of the traits that distinguish domesticated cattle from their wild predecessors. An open reading frame DNA sequence analysis was performed for 19,994 pairs of orthologous protein-coding genes from existing *Bos taurus* genomes and one extinct *B. primigenius* genome. Polymorphisms of miRNA-binding sites in the 3'-UTR were identified in 1620 of these orthologous genes. The identified 1620 genes with miRNA binding sites that differ between *B. taurus* and wild progenitors are candidate genes associated with domestication. These 1620 candidate genes have been found to be involved in the control of pigmentation, fertility, neurobiological processes, metabolism, immunity, and animal performance characteristics (including milk quality and feed conversion efficiency) [56]. These results suggest that the directed selection of miRNA regulatory variants was important in the domestication and subsequent artificial selection that gave rise to modern European cattle [56].

It has been shown that domesticated species differ from closely related wild ones in terms of an increased frequency of deletions [57], i.e. mutations, the nature of which is also closely related to retrotransposons [58].

There are also direct results of experimental studies indicating significant differences in the distribution of TEs in species such as the domestic dog (*Canis lupus familiaris*), gray wolf (*Canis lupus*) and red wolf (*Cuon alpinus*) which are associated with domestication processes [59]. Differences between the genomes of these species were revealed. Thus, TEs in the dog account for 41.75% of the nucleotide sequences in the genome which is higher than that of the gray wolf (39.26%) and red wolf (38.51%). The most divergent components of TEs in these genomes are long dispersed nuclear elements LINE1 (L1) and microsatellites, which distinguish the dog from the gray wolf by 86.1%, from the red wolf by 83.2%.

A comparison of the distribution of transposable genetic elements between the domestic rabbit (*Oryctolagus cuniculus domesticus*) and the pika (*Ochotona princeps*) showed a markedly higher frequency of occurrence in the domestic species compared to the wild one [60]. Similar differences were found when comparing a domestic cat (*Felis silvestris catus*) and a closely related wild species (Felis silvestris silvestris) [61].

Therefore, key questions of domestication are where does the large range of phenotypic variability come from and why some species are amenable to domestication unless others. In our opinion, the answer may be that during formation of a new multicomponent niche human (domesticator)—plants/animals (domesticants), an increased density of mobile genetic elements occurres in the genome of the domesticant due to a reduced resistance of genomes to transposon integration. This assumption is supported by the involvement of transposons in epigenetic variability [31, 51] and in the organization of the interphase nucleus architectonics [35], by the high rate of TEs evolution [34], the frequency of recombinations between TEs [43, 58], and the differences between domestic and wild species in terms of prevalence of some transposons in genomes [56, 59-61]. In our previous studies, we attracted attention to the fact that the main unresolved issue in the search for leading genomic features in domesticated species when compared to closely related wild species is the source of variability that distinguishes these closely related forms [62, 63]. The accumulated data on the leading role of transposons in evolutionary transformations, in the formation and modulation of regulatory networks that control gene expression profiles, suggest that mobile genetic elements are also essential for domestication processes.

Thus, in our opinion, the high ability of mobile genetic elements (TEs) to generate variability, their relatively increased frequency of occurrence in the genomes of domesticated species compared to closely related wild ones, and the

involvement of TEs in the formation of regulatory networks of gene expression profiles suggests that the formation of new habitat conditions and human-controlled reproduction leads to increased contacts of domesticated forms with a wide range of exogenous viruses new to them. The events ultimately cause an increased variability (including regulatory networks) which contributes to phenotypic diversity and the effectiveness of artificial selection. This or a close mechanism, apparently, should be involved in all evolutionary processes associated with the emergence of new forms.

REFERENCES

1. Andersson L., Purugganan M. Molecular genetic variation of animals and plants under domestication. *Proc. Natl. Acad. Sci. USA*, 2022, 119(30): e2122150119 (doi: 10.1073/pnas.2122150119).
2. Darvin Ch. *Izmeneniya domashnikh zhivotnykh i kul'turnykh rasteniy. Sochineniya. Tom 4* [Changes in domestic animals and cultivated plants. Works. V. 4]. Moscow-Leningrad, 1951 (in Russ.).
3. Diamond J. Evolution, consequences and future of plant and animal domestication. *Nature*, 2002, 418(6898): 700-707 (doi: 10.1038/nature01019).
4. Zeder M.A. Core questions in domestication research. *Proc. Natl. Acad. Sci. USA*, 2015, 112(11): 3191-3198 (doi: 10.1073/pnas.1501711112).
5. Denham T., Barton H., Castillo C., Crowther A., Dotte-Sarout E., Florin S.A., Pritchard J., Barron A., Zhang Y., Fuller D.Q. The domestication syndrome in vegetatively propagated field crops. *Ann. Bot.*, 2020, 125(4): 581-597 (doi: 10.1093/aob/mcz212).
6. Burban E., Tenaillon M.I., Le Rouzic A. Gene network simulations provide testable predictions for the molecular domestication syndrome. *Genetics*, 2022, 220(2): iyab214 (doi: 10.1093/genetics/iyab214).
7. Rubio A.O., Summers K. Neural crest cell genes and the domestication syndrome: a comparative analysis of selection. *PLoS ONE*, 2022, 17(2): e0263830 (doi: 10.1371/journal.pone.0263830).
8. Glazko V.I., Kosovsky G.Yu., Glazko T.T. Human and domesticated species (review). *Biogeosystem Technique*, 2021, 8(1): 34-45 (doi: 10.13187/bgt.2021.1.34).
9. Glazko V., Zybailov B., Glazko T. Asking the right question about the genetic basis of domestication: what is the source of genetic diversity of domesticated species? *Adv. Genet. Eng.*, 2015, 4: 2 (doi: 10.4172/2169-0111.1000125).
10. Shapiro J.A. Nothing in evolution makes sense except in the light of genomics: read-write genome evolution as an active biological process. *Biology (Basel)*, 2016, 5(2): 27 (doi: 10.3390/biology5020027).
11. Shapiro J.A. Living organisms author their read-write genomes in evolution. *Biology (Basel)*, 2017, 6(4): 42 (doi: 10.3390/biology6040042).
12. Shapiro J.A. What we have learned about evolutionary genome change in the past 7 decades. *Biosystems*, 2022, 215-216: 104669 (doi: 10.1016/j.biosystems.2022.104669).
13. Trono D. Transposable elements, polydactyl proteins, and the genesis of human-specific transcription networks. *Cold Spring Harb. Symp. Quant. Biol.*, 2015, 80: 281-288 (doi: 10.1101/sqb.2015.80.027573).
14. Ricci M., Peona V., Guichard E., Taccioli C., Boattini A. Transposable elements activity is positively related to rate of speciation in mammals. *J. Mol. Evol.*, 2018, 86(5): 303-310 (doi: 10.1007/s00239-018-9847-7).
15. Rebollo R., Horard B., Hubert B., Vieira C. Jumping genes and epigenetics: towards new species. *Gene*, 2010, 454(1-2): 1-7 (doi: 10.1016/j.gene.2010.01.003).
16. Oliver K.R., Greene W.K. Mobile DNA and the TE-Thrust hypothesis: supporting evidence from the primates. *Mobile DNA*, 2011, 2: 8 (doi: 10.1186/1759-8753-2-8).
17. Oliver K.R., Greene W.K. Transposable elements: powerful facilitators of evolution. *BioEssays*, 2009, 31(7): 703-714 (doi: 10.1002/bies.200800219).
18. Almojil D., Bourgeois Y., Falis M., Hariyani I., Wilcox J., Boissinot S. The structural, functional and evolutionary impact of transposable elements in eukaryotes. *Genes (Basel)*, 2021, 12(6): 918 (doi: 10.3390/genes12060918).
19. Herrera-Úbeda C., Garcia-Fernandez J. New genes born-in or invading vertebrate genomes. *Front. Cell Dev. Biol.*, 2021, 9: 713918 (doi: 10.3389/fcell.2021.713918).
20. Navas-Pérez E., Vicente-García C., Mirra S., Burguera D., Fernandez-Castillo N., Ferrán J.L., López-Mayorga M., Alaiz-Noya M., Suárez-Pereira I., Antón-Galindo E., Ulloa F., Herrera-Úbeda C., Cuscy P., Falcón-Moya R., Rodríguez-Moreno A., D'Aniello S., Cormand B., Marfany G., Soriano E., Carriyn Á.M., Carvajal J.J., Garcia-Fernandez J. Characterization of an eutherian gene cluster generated after transposon domestication identifies *Bex3* as relevant for advanced neurological functions. *Genome Biol.*, 2020, 21(1): 267 (doi: 10.1186/s13059-020-02172-3).

21. Winter E.E., Ponting C.P. Mammalian *BEX*, *WEX* and *GASP* genes: coding and non-coding chimaerism sustained by gene conversion events. *BMC Evol. Biol.*, 2005, 5: 54 (doi: 10.1186/1471-2148-5-54).
22. Cabeen R.P., Toga A.W., Allman J.M. Mapping frontoinsular cortex from diffusion microstructure. *Cerebral Cortex*, 2022: bhac237 (doi: 10.1093/cercor/bhac237).
23. Bhattacharyya M.K., Smith A.M., Ellis T.H., Hedley C., Martin C. The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell*, 1990, 60(1): 115-122 (doi: 10.1016/0092-8674(90)90721-p).
24. Hazzouri K.M., Flowers J.M., Visser H.J., Khierallah H.S.M., Rosas U., Pham G.M., Meyer R.S., Johansen C.K., Fresquez Z.A., Masmoudi K., Haider N., El Kadri N., Idaghdour Y., Malek J.A., Thirkhill D., Markhand G.S., Krueger R.R., Zaid A., Purugganan M.D. Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. *Nat. Commun.*, 2015, 6: 8824 (doi: 10.1038/ncomms9824).
25. Feschotte C., Jiang N., Wessler S.R. Plant transposable elements: where genetics meets genomics. *Nat. Rev. Genet.*, 2002, 3(5): 329-341 (doi: 10.1038/nrg793).
26. Matsumine H., Herbst M.A., Ou S.H., Wilson J.D., McPhaul M.J. Aromatase mRNA in the extragonadal tissues of chickens with the henny-feathering trait is derived from a distinctive promoter structure that contains a segment of a retroviral long terminal repeat. Functional organization of the Sebright, Leghorn, and Campine aromatase genes. *J. Biol. Chem.*, 1991, 266(30): 19900-19907.
27. Li J., Davis B.W., Jern P., Dorshorst B.J., Siegel P.B., Andersson L. Characterization of the endogenous retrovirus insertion in *CYP19A1* associated with henny feathering in chicken. *Mobile DNA*, 2019, 10: 38 (doi: 10.1186/s13100-019-0181-4).
28. Halabian R., Makałowski W. A map of 3′ DNA transduction variants mediated by non-LTR retroelements on 3202 human genomes. *Biology*, 2022, 11: 1032 (doi: 10.3390/biology11071032).
29. Crow M., Suresh H., Lee J., Gillis J. Coexpression reveals conserved gene programs that covary with cell type across kingdoms. *Nucleic Acids Research*, 2022, 50(8): 4302-4314 (doi: 10.1093/nar/gkac276).
30. Nicolau M., Picault N., Moissiard G. The evolutionary volte-face of transposable elements: from harmful jumping genes to major drivers of genetic innovation. *Cells*, 2021, 10(11): 2952 (doi: 10.3390/cells10112952).
31. Colonna Romano N., Fanti L. Transposable elements: major players in shaping genomic and evolutionary patterns. *Cells*, 2022, 11(6): 1048 (doi: 10.3390/cells11061048).
32. Lu J.Y., Chang L., Li T., Wang T., Yin Y., Zhan G., Han X., Zhang K., Tao Y., Percharde M., Wang L., Peng Q., Yan P., Zhang H., Bi X., Shao W., Hong Y., Wu Z., Ma R., Wang P., Li W., Zhang J., Chang Z., Hou Y., Zhu B., Ramalho-Santos M., Li P., Xie W., Na J., Sun Y., Shen X. Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome. *Cell Research*, 2021, 31(6): 613-630 (doi: 10.1038/s41422-020-00466-6).
33. 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).
34. Percharde M., Sultana T., Ramalho-Santos M. What doesn't kill you makes you stronger: transposons as dual players in chromatin regulation and genomic variation. *BioEssays*, 2020, 42(4): e1900232 (doi: 10.1002/bies.201900232).
35. Falk M., Feodorova Y., Naumova N., Imakaev M., Lajoie B.R., Leonhardt H. Joffe B., Dekker J., Fudenberg G., Solovei I., Mirny L.A. Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature*, 2019, 570(7761): 395-399 (doi: 10.1038/s41586-019-1275-3).
36. Soares S.C., Eler E.S., E Silva C.E.F., da Silva M.N.F., Araújo N.P., Svartman M., Feldberg E. LINE-1 and SINE-B1 mapping and genome diversification in *Proechimys* species (*Rodentia: Echimyidae*). *Life Science Alliance*, 2022, 5(6): e202101104 (doi: 10.26508/lsa.202101104).
37. Stegniy V.N. *Populyatsionnaya genetika i evolyutsiya malyariynykh komarov* [Population genetics and evolution of malarial mosquitoes]. Tomsk, 1991 (in Russ.).
38. Artemov G.N., Stegniy V.N. *Vestnik Tomskogo gosudarstvennogo universiteta. Biologiya*, 2010, 2(10): 123-131 (in Russ.).
39. Roller M., Stamper E., Villar D., Izuogu O., Martin F., Redmond A.M., Ramachanderan R., Harewood L., Odom D.T., Flicek P. LINE retrotransposons characterize mammalian tissue-specific and evolutionarily dynamic regulatory regions. *Genome Biology*, 2021, 22(1): 62 (doi: 10.1186/s13059-021-02260-y).
40. Percharde M., Sultana T., Ramalho-Santos M. What doesn't kill you makes you stronger: transposons as dual players in chromatin regulation and genomic variation. *BioEssays*, 2020, 42(4): e1900232 (doi: 10.1002/bies.201900232).
41. Chiu E.S., VandeWoude S. Endogenous retroviruses drive resistance and promotion of exogenous retroviral homologs. *Annu. Rev. Anim. Biosci.*, 2021, 9: 225-248 (doi: 10.1146/annurev-animal-050620-101416).

42. Chong A.Y., Kojima K.K., Jurka J., Ray D.A., Smit A.F., Isberg S.R., Gongora J. Evolution and gene capture in ancient endogenous retroviruses — insights from the crocodilian genomes. *Retrovirology*, 2014, 11: 71 (doi: 10.1186/s12977-014-0071-2).
43. Benachenhou F., Sperber G.O., Bongcam-Rudloff E., Andersson G., Boeke J.D., Blomberg J. Conserved structure and inferred evolutionary history of long terminal repeats (LTRs). *Mobile DNA*, 2013, 4(1): 5 (doi: 10.1186/1759-8753-4-5).
44. Chiu E.S., Gagne R.B. A natural laboratory to elucidate the evolution of endogenous-exogenous retroviral interactions. *Mol. Ecol.*, 2021, 30(11): 2473-2476 (doi: 10.1111/mec.15850).
45. Chiu E.S., VandeWoude S. Endogenous retroviruses drive resistance and promotion of exogenous retroviral homologs. *Annu. Rev. Anim. Biosci.*, 2021, 9: 225-248 (doi: 10.1146/annurev-animal-050620-101416).
46. Lanciano S., Mirouze M. Transposable elements: all mobile, all different, some stress responsive, some adaptive? *Opinion in Genetics and Development*, 2018, 49: 106-114 (doi: 10.1016/j.gde.2018.04.002).
47. Hosaka A., Kakutani T. Transposable elements, genome evolution and transgenerational epigenetic variation. *Curr. Opin. Genet. Dev.*, 2018, 49: 43-48 (doi: 10.1016/j.gde.2018.02.012).
48. Cho J. Transposon-derived non-coding RNAs and their function in plants. *Front. Plant Sci.*, 2018, 9: 600 (doi: 10.3389/fpls.2018.00600).;
49. Shapiro J.A. Living organisms author their read-write genomes in evolution. Biology, 2017, 6(4): 42 (doi: 10.3390/biology6040042).
50. Hosaka A., Kakutani T. Transposable elements, genome evolution and transgenerational epigenetic variation. *Curr. Opin. Genet. Dev.*, 2018, 49: 43-48 (doi: 10.1016/j.gde.2018.02.012).
51. Bhogireddy S., Mangrauthia S.K., Kumar R., Pandey A.K., Singh S., Jain A., Budak H., Varshney R.K., Kudapa H. Regulatory non-coding RNAs: a new frontier in regulation of plant biology. *Funct. Integr. Genomics*, 2021, 21(3-4): 313-330 (doi: 10.1007/s10142-021-00787-8).
52. Jungers C.F., Djuranovic S. Modulation of miRISC-mediated gene silencing in eukaryotes. *Front. Mol. Biosci.*, 2022, 9: 832916 (doi: 10.3389/fmolb.2022.832916).
53. Sadeq S., Al-Hashimi S., Cusack C.M., Werner A. Endogenous double-stranded RNA. *Noncoding RNA*, 2021, 7(1): 15 (doi: 10.3390/ncrna7010015).
54. Hu G., Do D.N., Davoudi P., Miar Y. Emerging roles of non-coding RNAs in the feed efficiency of livestock species. *Genes (Basel)*, 2022, 13(2): 297 (doi: 10.3390/genes13020297).
55. Penso-Dolfin L., Moxon S., Haerty W., Di Palma F. The evolutionary dynamics of microRNAs in domestic mammals. Sci. Rep., 2018, 8(1): 17050 (doi: 10.1038/s41598-018-34243-8).
56. Braud M., Magee D.A., Park S.D., Sonstegard T.S., Waters S.M., MacHugh D.E., Spillane C. Genome-wide microRNA binding site variation between extinct wild aurochs and modern cattle identifies candidate microRNA-regulated domestication genes. *Front. Genet.*, 2017, 8: 3 (doi: 10.3389/fgene.2017.00003).
57. Makino T., Rubin C.J., Carneiro M., Axelsson E., Andersson L., Webster M.T. Elevated proportions of deleterious genetic variation in domestic animals and plants. *Genome Biol. Evol.*, 2018, 10(1): 276-290 (doi: 10.1093/gbe/evy004).
58. Startek M., Szafranski P., Gambin T., Campbell I.M., Hixson P., Shaw C.A., Stankiewicz P., Gambin A.. Genome-wide analyses of LINE-LINE-mediated nonallelic homologous recombination. *Nucleic Acids Research*, 2015, 43(4): 2188-2198 (doi: 10.1093/nar/gku1394).
59. Wang G.D., Shao X.J., Bai B., Wang J., Wang X., Cao X., Liu Y.H., Wang X., Yin T.T., Zhang S.J., Lu Y., Wang Z., Wang L., Zhao W., Zhang B., Ruan J., Zhang Y.P. Structural variation during dog domestication: insights from grey wolf and dhole genomes. *National Science Review*, 2019, 6(1): 110-122 (doi: 10.1093/nsr/nwy076).
60. Yang N., Zhao B., Chen Y., D'Alessandro E., Chen C., Ji T., Wu X., Song C. Distinct retrotransposon evolution profile in the genome of rabbit (*Oryctolagus cuniculus*). *Genome Biol. Evol.*, 2021, 13(8): evab168 (doi: 10.1093/gbe/evab168).
61. Ngo M.H., Arnal M., Sumi R., Kawasaki J., Miyake A., Grant C.K., Otoi T., Fernández de Luco D., Nishigaki K. Tracking the fate of endogenous retrovirus segregation in wild and domestic cats. J. Virol., 2019, 93(24): e01324-19 (doi: 10.1128/JVI.01324-19).
62. Glazko V.I. Gene and genomic levels of domestication signature (review). *Sel'skokhozyaistvennaya biologiya* [*Agricultural Biology*], 2018, 53(4): 659-672 (doi: 10.15389/agrobiology.2018.4.659eng).
63. Glazko V., Zybailov B., Glazko T. Asking the right question about the genetic basis of domestication: what is the source of genetic diversity of domesticated species? *Adv. Genet. Eng.*, 2015, 4: 2 (doi: 10.4172/2169-0111.1000125).