AGRICULTURAL BIOLOGY,
2023, V. 58, Iss. 2, pp. 223-248
[SEL'SKOKHOZYAISTVENNAYA BIOLOGIYA]

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

UDC 636:575.174.015.3

doi: 10.15389/agrobiology.2023.2.223eng
doi: 10.15389/agrobiology.2023.2.223rus

DNA MARKERS AND MICROSATELLITE CODE
(review)

V.I. GLAZKO, G.Yu. KOSOVSKY, T.T. GLAZKO✉, L.M. FEDOROVA

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 vigvalery@gmail.com, gkosovsky@mail.ru, tglazko@rambler.ru (✉ corresponding author), felami@mail.ru
ORCID:
Glazko V.I. orcid.org/0000-0003-4242-2239
Kosovsky G.Yu. orcid.org/0000-0003-1889-6063
Glazko T.T. orcid.org/0000-0002-0520-7022
Fedorova L.M. orcid.org/0000-0002-1514-3050
The authors declare no conflict of interests
Final revision received January 10, 2023
Accepted February 6, 2023

Abstract

The search for genetic markers that simplify the selection of animals for crosses, increasing the likelihood of offspring obtaining with the desired manifestation of economically valuable traits is a central problem in modern animal husbandry. Here, we discuss the most successful applications of various types of DNA markers of genomic element polymorphisms for solving specific breeding problems. Microsatellites are used to exclude errors of origin, single nucleotide polymorphisms (SNPs) to create maps of genomic regions in which polymorphism is associated with the variability of phenotypic characteristics (D.J. Rigden, X.M. Fernández, 2023) and to identify the localization of key genes of adaptation to natural selection factors at the natural habitat edges and in areas of animal husbandry risky (E.K. Cheruiyot et al., 2022; L. Buggiotti et al.., 2021, 2022). The loci of increased variability in the copyicity of genome regions (CNV) are used to assess their involvement in responses to natural and artificial selection factors of such polygenic systems as sensory, immune, and transporter (Y. Huang et al., 2021; P. Davoudi et al., 2022). The predominant involvement of regulatory networks including dispersed and tandem repeats, in particular microsatellite repeats, in epigenetic and phenotypic variability is discussed (R.P. Kumar et al., 2010). The structural and functional complexity of microsatellite loci, individual features of variability of specific loci and their participation in evolutionary, recombination, transcription processes are considered. Their involvement in the organization of secondary DNA structures, participation in the formation and variability of the architectonics of the interphase nucleus and regulation of gene expression profiles is noted (R.P. Kumar et al., 2010; X. Tang et al., 2022). The study of regulatory networks is of particular importance, since there is evidence that the size of the genome in animals of different taxa, as well as the distribution and composition of mobile genetic elements (sources of components of regulatory networks) differ significantly, in contrast to the similarity in the number of genes encoding proteins (V.I. Glazko et al., 2022). Accumulating evidence suggests that polylocus genotyping of individual microsatellites and dispersed repeats can contribute to solving practical problems, such as information on the specific features of the population-genetic structure, consolidation and differences between closely related groups of animals.

Keywords: DNA markers, microsatellites, short tandem repeat, STR, single nucleotide polymorphism, SNP, copy number variations, CNV, genome-wide association studies, GWAS

The success of breeding work is determined by the quality of selection and selection of animals for crossing, which directly depends on the ability to predict the manifestation of desirable economically valuable traits in specific environmental conditions. The idea of selection using markers was first put forward by the Soviet geneticist A.S. Serebrovsky who introduced the concept of a signal gene. According to these ideas, the so-called signalia are alternative genes convenient for Mendelistic observations with a more or less known chromosomal localization, which, without directly affecting the transgressive trait being studied and influencing in a fairly specific way, facilitate the genetic analysis of this trait, allowing one

to monitor the inheritance of that trait the region of the chromosome in which these signalia are located [1]. Further, as markers, antibodies to protein antigens of biological objects were considered in animals, then polymorphism of protein products of the same genes, detected by the electrical charge of molecules (biochemical markers) in both plants and animals [2].

The next stage was the development of DNA markers of various types. According to the general definition, DNA markers are markers of polymorphism of genomic regions the variability of which can potentially be associated with its manifestations at the phenotypic level. However, with in-depth study of some DNA markers of various genomic elements, their involvement in complex networks of relationships between them, in ensuring the stability, variability and evolution of biological systems, becomes obvious.

In this review, we presented examples of the effectiveness of different generations of DNA markers as tools in applied research and analyzed data on microsatellite (short tandem repeat, STR) DNA markers. The variety of biological functions of STRs is considered, including, in particular, participation in the formation of secondary DNA structures, the architectonics of the interphase nucleus, elements of gene expression regulatory networks, which, in our opinion, changes the understanding of their biological role and the possibilities of practical application.

Practical application of DNA markers. Types of DNA markers and their significance for genetic and genomic research. Since the 1980s, polymorphism of genomic regions began to be used for genetic marking. In the 1980s, the most common DNA markers were those based on restriction fragment length polymorphism (RFLP); in the 1990s, markers detected using polymerase chain reaction (marker-assisted selection) selection, MAS), after 2000, DNA chips for detecting single nucleotide polymorphisms (SNPs) during whole-genome sequencing became widespread.

Polylocus genotyping using DNA markers (genomic scanning) in agricultural species is widely used i) to determine the parameters of variability within and between breeds; ii) to identify population genetic characteristics in geographically separated groups and/or when mixing animals of different origins in groups; iii) to study evolutionary relationships and search for centers of origin and migration routes; iv) to map the main genes with polymorphism is associated with variability in phenotypic characteristics (including the identification of known alleles for genetically determined diseases and the identification of their carriers, and v) to detect alleles associated with increased resistance to infectious and non-infectious diseases). For multilocus genotyping, microsatellite markers are widely used - tandem repeats, the elementary (repeating) unit of which can be from 2 to 6 bp in length. (simple, or short tandem repeat, STR) [3].

In agreement with the Food and Agricultural Organization (FAO) and the International Society of Animal Genetics (ISAG), genotyping panels for farm animals were initially developed for blood groups, then for biochemical markers, and currently, by microsatellite polymorphism. These panels are specific for each species, include a couple of dozen loci and are designed in such a way that they allow PCR to be carried out simultaneously for several markers and solve a number of current problems (identification of origin, identification of breed population genetic characteristics of animals). For more than two decades, genetic certification of animals, necessary to exclude errors of origin, was carried out using such panels of highly polymorphic markers. The degree of their heterozygosity, despite frequent inbreeding in farm animals, is often 75% or more. Typically, dinucleotide and trinucleotide repeats are used for certification. Along with them, due to the expansion of whole-genome sequencing of representatives of different species, new

generations of DNA markers appear, usually based on SNPs.

To date, more than 30,000 Holstein bulls have been genotyped using BovineSNP50 BeadChip DNA microarrays (Illumina, USA), which allows simultaneous analysis of 54,001 SNPs (approximately one SNP per 50,000 bp). Using such microarrays, maps of SNP distribution across genomes have been constructed for different species [4] and genome-wide association maps of SNP localization sites with variability in phenotypic characteristics (genome-wide association study, GWAS) have been created [5, 6].

DNA microarrays (chips) make it possible to detect not only SNP-linked markers, but also copy number variations (CNVs), including deletions, duplications, translocations and inversions [7]. CNV is attracting increasing attention because it is often associated with variability in phenotypic traits in agricultural animal species [8-13] and with unfavorable phenotypic manifestations in humans [7]. Detailed chromosomal maps of the distribution of CNV markers were created. Specifically, in humans, CNV loci cover 12% of the genome (Database of Genomic Variants, http://projects.tcag.ca/variation/), meaning CNVs involve more nucleotides per genome than SNPs [14]. Spontaneous CNVs are predicted to occur at an average frequency of 10^{-4} bp [15]. The high level of variability gives reason to believe that the use of CNV markers of polymorphism in genomic DNA regions will increase the accuracy of mapping the main genes of such economically important animal characteristics as resistance to biotic and abiotic environmental factors and productive qualities [16-18].

Thus, it is worth noting the successful development of methods for sequencing biomolecules, especially using fourth-generation technologies (nanopore-based sequencing) [19], as well as complete reading of genomes in most agricultural animal species, including mammals [20]. However, the directions for practical application of the huge volume of results of these gene and genomic studies are still not sufficiently developed. It is necessary to correlate the data accumulated over decades on the genotypes of representatives of different breeds by microsatellites (STR) and the results of genotyping using SNP panels, which is currently (as previously STR analysis) carried out to control origin, breed identification, and identify relationships between genotypes and variability of phenotypic traits [21-24]. It turned out that, among other complexities, the STR panel recommended by ISAG for genotyping farm animals includes microsatellites that differ significantly from each other in polymorphism and efficiency in differentiating animals within and between breeds [21]. A study of the colocalization of STRs and SNPs in the cattle genome showed that only 57.1% of STRs are in linkage disequilibrium (LD), while the remaining 42.9% are located outside such blocks [25]. In other words, in the cattle genome, a significant number of STRs are not linked to SNPs (probably due to the specific mechanisms of mutations and genomic distribution of STRs, as well as their increased polymorphism). It follows from the fact that each microsatellite has its own characteristics of polymorphism and mutability, and combining different microsatellites into a common panel can lead to erroneous conclusions.

The widespread replacement of whole genome sequencing (WGS) with a relatively limited number of SNPs on DNA chips in order to identify SNPs associated with variability in phenotypic characteristics also often carries sources of significant errors [26]. As in the case of STR, the possibility of using a limited number of SNPs to predict the variability of phenotypic traits will depend on the localization of SNPs in various genomic elements and their structural and functional features [26].

The development of a new direction, the pangenomics (in particular, pangenomics of cattle), due to the accumulation of WGS data for many genomes,

makes it possible to add to the reference genome of cattle presented in GenBank (*Bos taurus*, https://www.ncbi.nlm.nih.gov /data-hub/taxonomy/9913/), appx. 4% of nucleotide sequences [27]. This may influence the shift in the positions of the analyzed SNP allelic variants. Thus, in cattle, in addition to the reference genome, 83,250 SNPs were identified, for which polymorphism is observed both within breeds and between breeds [27].

Search for DNA markers of economically valuable traits. In the last decade, the development of GWAS has allowed advances in understanding the genetic basis of complex traits and diseases in both humans and livestock [28]. An interesting fact is that most SNPs associated with phenotypic variability of properties are localized in non-coding sequences of the genome [28]. In many cases, such sequences are closely related to numerous regulatory elements (RE) influencing gene expression profiles [29]. However, their evolutionary preservation, variability, and involvement in the manifestation of complex polygenic traits still remain insufficiently studied.

STR or SNP polymorphism can be used to map genomic regions of the localization of genomic elements, the variability of which makes a significant contribution to the manifestation of quantitative economically valuable traits in animals of agricultural species, and to detect key genes and/or RE of such traits. As a rule, these genomic elements are detected in animals living in extreme environmental conditions. A striking example of successful searches are studies of the resistance of milk productivity of Australian Holstein cows to high temperatures [30, 31], and of Kholmogory and Yakut cattle to low temperatures [32, 33].

The composition and functional organization of genomic elements involved in the genetically determined control of milk production in cattle may be an example of the complex genetic basis of quantitative economically valuable traits. To date, a chromosomal map of genes involved in the formation of the udder and milk production in cattle (the so-called lactome map) has been created, which includes 197 milk protein genes and more than 6000 genes involved in the development and functioning of the mammary gland [34]. It turned out that these genes are scattered across all 30 chromosomes of cattle. Comparison of the genomes of the platypus, opossum, placental mammals (cattle, dog, human, mouse, and rat) [34] for the genes of milk proteins and mammary gland formation revealed losses and duplications, phylogenetic relationships, conservatism of the sequences of these genes and their evolution. Evidence has been obtained that in cattle, the genes for milk and mammary gland proteins evolve more slowly than in other studied placental species. It was found that in cattle, in comparison with other listed species, the genes of milk proteins that determine its nutritional and immune properties were the most divergent; the genes associated with the processes of milk secretion turned out to be the most conservative.

Analysis of the transcriptome at different stages of lactation showed that 16,892 genes are expressed during the intermediate period of the lactation cycle, 19,094 at the peak of lactation, and 18,070 during the decline of lactation. The expression level of genes encoding caseins, whey proteins and enzymes of the metabolic pathway for lactose synthesis was increased at the beginning of lactation, and most genes of the metabolic pathways of lipid metabolism were increased in the intermediate period and at the peak of lactation [35].

It is obvious that milk production is influenced by the genetic characteristics of the individual, epigenetic processes, nutrition, pathogens, climatic conditions and other external factors. This is especially clearly revealed by the example of differences in selection indices calculated for the same Holstein bulls based on the milk productivity of their daughters born in different ecological and geographical regions, in Luxenburg and Tunisia [36].

Over the past decade, there has been a significant increase in the number of studies assessing the impact of epigenetic variability associated with regulatory networks, which are represented, in particular, by microRNAs (miRNAs, small non-coding RNAs with transcriptional and post-transcriptional effects). The direct involvement of microRNAs in the control of gene expression profiles and in the regulation of the development and functioning of the mammary gland is increasingly being revealed [37, 38]. MicroRNAs have been found to play an important role in many processes associated with breast development and disease, as well as milk secretion. Hundreds of miRNAs have been identified in the mammary gland, but the number of miRNAs whose functions are fully known is very small. The problem is that one miRNA can be involved in the control of hundreds of genes, so functional validation of each target gene for this miRNA is difficult. The situation is further complicated by the fact that the response to the same environmental factors can be provided by different microRNAs, not only in closely related species [39], but also in breeds [40].

The diversity of microRNA spectra is very wide, as is their connection with the regulation of gene expression of different metabolic pathways and their intersection points. Of the 19,994 protein-coding orthologous gene pairs between *Bos taurus* and the extinct species *B. primigenius*, 1,620 genes differ in microRNA binding site polymorphism in the 3′-UTR [41]. These 1,620 genes are primarily involved in the control of pigmentation, reproduction, neural function, general metabolism, immune responses, and variation in animal performance traits, including milk quality and feed efficiency.

Applicability of DNA markers for practical selection. One of the first conclusions that can be drawn based on data from genetic and genomic studies of farm animals is apparently the following. There is no need to look for a universal method to solve all breeding problems. If STR panels have shown success in eliminating origin errors, there is no point in replacing them with more complex and expensive SNP-based test systems. Moreover, a search is already underway for universal STR panels, orthologous in different species, with the desired level of polymorphism [42], which will make it possible to differentiate species and intraspecific variability of biological objects [43].

Obviously, SNP maps are indispensable when searching for genomic regions in which genes for resistance to critical biotic and abiotic external factors are localized. Based on their refinements and expansion of mapping volumes, environmental genomics will be formed as a continuation of environmental genetics - a scientific direction laid down by A.S. Serebrovsky [1].

SNP analysis is important for searching for candidate genes for economically valuable traits when their expression differs significantly in the animals being studied, since in this case it is possible to search for a small number of genes that cause such differences. It is difficult to overestimate the success of using SNPs to reconstruct the genetic dynamics of populations based on the analysis of runs of homozygosity (ROH) [44] or when searching for mutations in genes critical for animal reproduction (loss-of-function or fertility haplotypes) [45].

A significant amount of data on the distribution of SNPs has allowed genome-wide association analyzes (GWAS) of SNPs with various traits in plant and animal species to be performed [46]. The results are presented on the resources of the National Genomics Data Center, China National Center for Bioinformation [47] and in the database collection of the journal Nucleic Acids Research [47]. For such a complex trait as milk production, it is difficult to expect obvious and reliable success [49]. L. Flori et al. [49) reported associations between SNP haplotypes (in linkage disequilibrium) with variability in milk production parameters in the three main dairy breeds of France — Holsteins, Normans, and Montbeliards. In areas of

increased density of such haplotypes in three breeds, a total of 40 genes were identified, mainly differing in the studied breeds. Perhaps the observed contradictions are due to epistatic interactions between gene ensembles involved in the manifestation of such economically valuable traits as the amount of total milk yield, the percentage and amount of milk fat and protein. In a GWAS analysis using 76,109 SNPs in 294,079 Holstein cows of the first lactation, the effect of pairwise epistasis on indicators of milk productivity and reproduction (total milk yield, yield of milk fat, protein, percentage of milk fat and protein, pregnancy rate of daughters) was assessed [50]. Of the top 50,000 identified pairwise epistasis effects for each trait, five involve large chromosomal regions with intrachromosomal epistasis [50]. In fact, this can explain the well-known undesirable correlations between total milk yield and milk protein content, milk fat content and reproduction in cows. A clear demonstration of the difficulty of identifying gene elements whose polymorphism is associated with variability in an economically valuable trait is the quantitative trait locus on chromosome 18 (BTA18), associated with ease of calving and stillbirth in Holstein-Friesian cattle and its crosses (51). This fact has been known for more than 20 years, but its genetic basis has not yet been identified. To identify it, based on genotyping of 2697 Holstein Friesians, a detailed analysis of the corresponding BTA18 region was performed and an assessment of linkage disequilibrium in this region was performed. As a result, the connection of the polymorphism with the described pathology was confirmed, 4 SNPs with almost perfect linkage disequilibrium were identified, but not a single candidate gene associated with the specified pathology was identified. An abundance of segmental duplications was found within and around the region [51].

The method of genotyping using CNV markers appears to be very effective in identifying physiological systems involved in the direction of selection, primarily at the interspecific level. Indirect support for this assumption is the fact that in farm animals, CNVs are most often detected in the case of genes involved in various functions of the immune system [52-55].

MicroRNAs as elements of regulatory networks have attracted attention due to their involvement in the control and modulation of the functional activity of genes [37-40]. However, it must be taken into account that the degree of variability in the expression level of different genes is not the same. Our analysis of gene expression profiles in the liver of pigs [56] revealed two groups of genes, with and without individual differences between the animals studied. In pigs, using the KEGG Metabolic pathway database (https://www.genome.jp/kegg/path-way.html), we assessed the involvement in different metabolic pathways for 17 genes with similar expression values in individuals (these genes conditionally designated as a group with constitutive expression) and 18 genes that make up the variable part of gene expression profiles with pronounced individual differences between animals. In our experiment, the products of 17 constitutively expressed genes were involved in 25 metabolic pathways, or an average of 1.5 pathways per gene, and each of the 18 genes with variable expression levels accounted for 3 metabolic pathways [56]. That is, the greater the number of metabolic pathways in which the gene product is involved, the more complex the potential regulatory elements that unite and control them, the higher the individual variability of gene expression. Thus, although the role of microRNAs in regulatory networks is indisputable (as is the role of the regulatory networks themselves in the formation of traits and changes in their manifestations under the influence of influencing factors), the promise of microRNAs as molecular markers is ambiguous. As noted above, one microRNA can affect the transcripts of several dozen genes, and different microRNAs can affect the activity of one gene [37-40]. Thousands of transcription regulatory factors also change it by interacting with non-coding nucleotide sequences of the

gene [40, 41]. In addition, the interaction of gene products of the same metabolic pathway and/or different metabolic pathways can also activate processes leading to changes in gene activity profiles [40, 41].

Microsatellites, structural and functional diversity. *Structural and functional features of STR markers.* Microsatellites belong to the first generation of DNA markers, which have been widely used in the genetics and genomics of humans, farm animals and plants for more than 30 years [42]. The experimental data accumulated over this period made it possible to obtain a fairly complete pattern of the complex structural and functional organization of DNA markers of this type. To date, species-specific features of the number and genomic distribution of microsatellites have already been studied. Maps of the distribution of microsatellites in the genomes of different species have been created [57].

Using the example of the distribution of STRs in the genome of the domestic rabbit (*Oryctolagus cuniculus*), one can note a pronounced difference in the frequency of occurrence of STRs due to the length of the elementary repeat: 579097 mono-, 927755 di-, 122482 tri-, 767458 tetra-, 614173 penta- and 1739144 hexanucleotide repeats [57]. The large number of hexanucleotide repeats is explained by the fact that in mammals the conserved telomeric repeat TTAGGG is duplicated several thousand times [58]. It is interesting that in total in the genome of both the rabbit and a number of other mammalian species, including humans, there are significantly fewer trinucleotide repeats than di-, tetra- and pentanucleotide repeats, that is, this difference is not related to the length of the elementary repeat unit [57]. Probably because the structure of trinucleotide repeats corresponds to the triplet principle of the genetic code, they are under selection pressure.

There are pronounced differences between the frequency of occurrence of microsatellites with the same length of the elementary unit, but with different core motifs, for example AG and AC. In the rabbit, the AG motif occurs approximately 5 times more often than the AC motif and more often than all other microsatellites, while in humans, on the contrary, there are more microsatellites with the AC core motif than with the AG motif [57, 59].

STRs also occur in prokaryotes, but at low frequency [56). Trinucleotide STRs are common in nematodes and insects, and dinucleotide STRs are common in fish, a relative deficiency of which is observed in birds [60]. Closely related species can differ significantly in the number of microsatellite loci [60]. We have already noted that, in general, in many mammalian species, di- and tetranucleotide STR motifs are more frequent than trinucleotide ones, but the frequency of the latter varies significantly depending on the microsatellite core motifs, and these differences may be species-specific [61].

Different STRs are generally found most frequently in intergenic spaces, followed by introns, promoter regions, and least frequently in exons [56]. The microsatellite database of different species [56] shows that trinucleotide STRs are found more often in exons than di- and tetranucleotide STRs. Moreover, the frequency of occurrence of trinucleotide STRs in exons can be quite high.

Despite the relatively reduced frequency of occurrence of STRs in exons, their contribution to the polymorphism of encoded proteins can be significant. Thus, recently, by comparing the results of whole-genome sequencing in laboratory mice (71 lines), STR alleles present in the coding regions of 562 genes were identified and evidence was provided that these alleles can change the folding of the encoded protein and thus have a significant impact on its function [62].

Expansion of microsatellites during pathologies and adaptations. It is known that many human pathologies (most often neurocognitive and neurodegenerative) are associated with polymorphism in the copy number (length) of trinucleotide microsatellites [63-66]. In many studies of the genetic basis of diseases in humans,

amplification of triplets $(CCG)_n$, $(CGG)_n$, $(GCC)_n$, $(GCG)_n$ and $(CAG)_n$ was found in the coding regions of the genome, which determine the synthesis of polyproline, polyarginine, polyalanine and polyglutamine [67]. The development of a number of diseases is also associated with changes in the lengths of dinucleotide microsatellites with the GA core motif at their specific genomic localization [67].

Variability in the lengths of trinucleotide STRs is also found in some examples of adaptations. Thus, a connection between an increase in the copy number of microsatellites and an increase in adaptive potential was found in the giant panda during adaptation to food rich in carbohydrates, and in polar and brown bears - to the temperature of the habitat [68]. Variation in STR associated with these adaptations has mainly been identified in regulatory genes (e.g., transcription regulatory factor genes, insulin-like growth factor receptor signaling pathway). These genes are mainly involved in two metabolic pathways involved in key physiological processes (cardiovascular function and regulation of energy metabolism).

The presence of STRs in coding sequences can affect protein folding and change its flexibility, which allows it to bind to various substrates, be it nucleotides, lipids or proteins. Such proteins, containing amino acid repeats encoded by STRs, are expected to participate in the regulation of gene expression, are often multifunctional and have pleiotropic effects, increasing the resistance of cells and multicellular organisms to variability in environmental factors, which seems to justify the complexity associated with the potential high mutability of microsatellites [69, 70].

Non-coding STRs are also capable of significantly influencing phenotypic variation. In humans, it has been reported [71] that 10-15% of heritable variation in gene expression is associated with the presence of STRs. STRs were identified alleles of which differed significantly between ethnic groups. Fifteen STRs were found in which the repeat length correlates with the level of gene expression, two of these genes (*Glutathione Peroxidase 7* and *Glutathione S-Transferase Mu 3*) are involved in glutathione metabolism [72].

The complexity of the mechanisms of influence on transcription and translation with increasing STR lengths is discussed. An increase in the copy number of trinucleotide STRs in coding sequences can lead to pathologies due to the appearance of polyglutamine or polyalanine regions incorrectly localized in proteins, which leads, in particular, to disruption of protein-protein interactions, many of which are involved in the regulation of transcription, DNA repair and/or interfere formation of molecular condensates [73]. The formation of a condensation of a certain transcript with transcription regulatory factors affects the expression of a number of genes, and the accumulation of amino acid repeats can lead to disruption of such condensation and changes in transcription regulation. For example, the polyglutamine protein TBP (TATA-box Binding Protein) binds to the TATA box of gene promoters to initiate transcription, but when the length of the polyglutamine repeats are enlarged, its ability to condense with transcriptional activators is altered, leading to the transcriptional dysregulation observed in many polyglutamine diseases [73]. In other cases, such as with Ataxin-2, the RNA Binding Protein (RBP), involved in the assembly of condensates and stress granules and in RNA processing, carries a polyglutamine sequence. An increase in its copy number leads to neurodegenerative diseases (Spinocerebellar ataxia type 2, SCA2) [74, 75]. Let us recall that RNA-binding proteins also play a significant role in nuclear-cytoplasmic transport, the disruption of which also leads to neurodegenerative diseases [76].

For the studied cases, as noted above [57], expansion of trinucleotide repeats is typical for exons, while for introns, promoters, 3'- and 5'-UTRs, variability in the lengths of not only trinucleotides, but also tetra-, penta-, hexa- and decanucleotide

repeats occurs [77]. When an STR is amplified, the methylation of the corresponding DNA region may change; the STR elongation itself can change the distance between regulatory motifs in promoters, which will significantly affect expression [78]. Transcription of long STRs leads to the formation of RNA aggregates that serve as a trap for proteins and to multimolecular interactions with other transcripts, which, in turn, can affect splicing, gene expression profiles, and RNA interference [79-81].

STRs can be transcribed in the sense and antisense directions. This leads to the appearance of toxic peptides due to translation from non-ATG initiated triplets (repeat-associated non-ATG translation, RAN) transcribed from the STR [82].

Some studies note that many pathologies associated with STR amplification are detected in genes encoding transcription regulatory factors, which is accompanied by impaired formation of multimolecular condensates and interferes with their interaction with RNA polymerase II [77].

Features of the occurrence of mutations in STR. It has been found [83] that the localization of STRs and recombination hotspots in meiosis, which are usually located in gene promoters, often coincide. STRs can influence recombination processes at such points. This has been shown for STRs with core motifs GA, CA, GT, CT due to their high affinity for recombination enzymes (84). STR mutation rates vary widely, from 10^{-2} to 10^{-8} per locus per generation, but vary widely from locus to locus [85-87]. The dependence of the frequency of STR mutations on the action of environmental factors was described in model objects (*Caenorhabditis elegans*) when comparing mutability in laboratory and natural conditions [88].

In some cases, mechanisms have been discovered which compensate the adverse effects of STR mutations. Thus, there is a close relationship between the mutability of STRs and the polymorphism of the chromatin proteins that package them, which mitigate the adverse effects of changes in the length and organization of STRs on the processes involved in the suppression of transposon expression, accurate transmission of chromosomes, and ensuring their integrity [89].

Of particular structural and functional significance, including for mutability, is the ability of STR to form secondary DNA structures. The formation of G4 quadruplexes in tracks enriched with G/C, triplexes in purine-pyrimidine tracks, R-loops (DNA-RNA duplexes with displacement of the second DNA strand, which is not complementary to RNA), and other loop and hairpin structures affect gene expression profiles and enzyme function repairs, DNA polymerase function, STR instability [90-92].

STRs can result in non-canonical DNA structures that differ from the classical dextrorotatory B-helix, as determined by the primary nucleotide sequence. For example, levorotatory Z-DNA contains alternating GC-rich purine-pyrimidine sequences, and supercoiling is critical for the formation and stabilization of the Z-form of DNA. Z-DNA is thought to regulate the level of supercoiling and thus plays important roles in transcription, gene expression, recombination, translocation, and deletion [93]. Thus, Z-DNA formation induces instability in the region of trinucleotide repeats (CAG, CGG, and GAC), which are associated with various neurodegenerative diseases [93].

The emergence of non-canonical multistranded structures in the regions of purine-pyrimidine tracks consisting of microsatellites, for example AG/TC, GAG/CTC, includes DNA-DNA interactions with the release of one DNA strand, DNA-RNA interactions with the same effect, and interactions of double-stranded RNA with RNA [94-96]. The interaction of the third strand with duplex DNA or RNA in a double-stranded sequence-specific manner, leading to the formation of an intermolecular triplex, has a significant impact on transcription, post-transcriptional modifications, and mutagenesis [95].

It should be emphasized that all secondary structures are dynamic in nature, appear under certain conditions, and disappear when they change [93].

Special attention is attracted by STR sequences predisposed to the formation of cruciform structures in the regions of localization of inverted repeats, since many proteins involved in the control of cell division, for example, topoisomerases, p53, Rif1 (Replication Timing Regulatory Factor 1), can induce the formation of such structures [97]. Cross-shaped DNA forms play an important role in the regulation of replication and gene expression, are involved in the formation of nucleosome structure and recombination [98], and serve as targets for many architectural and regulatory proteins, e.g., histones H1 and H5, topoisomerase IIβ, proteins HMG, HU, p53, proto-oncogenic protein DEK [97, 98]. A number of DNA-binding proteins (eg, members of the HMGB-box family, Rad54, the BRCA1 protein, and the polymerase PARP-1) preferentially bind to cruciform structures [97]. It is assumed that, according to their function, proteins that interact with cruciform structures are mainly divided into four main groups: topoisomerases; DNA repair proteins and transcription regulatory factors; proteins involved in replication; chromatin-associated proteins [98]. The prevalence of cruciform structures formed by inverted STR repeats and their role in epigenetic regulation and maintenance of cellular homeostasis allow us to consider inverted repeats as essential components of regulatory systems [98].

In mammals, sex differences in the frequency of STR mutations have been described. Since oocytes, unlike spermatozoa, in particular in mammals, are in a state of rest for a long time, mutations that arise in STR during homologous recombinations, unequal crossing over, and double-strand breaks can accumulate and have more pronounced manifestations [99]. In general, STR mutations inherited through the maternal germline have a slightly higher frequency than those inherited from the father [100]. However, with age, the number of STR mutations in oocytes remains virtually unchanged, while in sperm it increases 2-fold (studies were conducted in a group of men from 20 to 58 years old) [101].

Complete sequencing of the genomes of 544 people from 29 families in three generations (database of the Center for the Study of Human Polymorphism (Center d'Etude du Polymorphisme Humain, CEPH, https://uofuhealth.utah.edu/center-genomic-medicine/research/ceph-resources) showed a high diversity of new types of STR mutations occurring in different families and at different STR loci [102]. A relationship was found between repeat length and mutation frequency (with the exception of mononucleotide STRs). The lowest frequency of mutations was detected in exons (only two in trinucleotide STRs), the largest part of them (53.38%) occurs in intergenic regions, slightly less than half (44.87%) are located in introns, apprx. 1.6% in 5′ - and 3′-UTR. The average calculated mutation rate (5.24×10^{-5}) is consistent with that for other types of mutations, including averages for SNPs, but in this family-based analysis using data from CEPH there was no correlation between the occurrence of new alleles at STRs and SNPs. It turned out that approximately 30% of STR mutations occur in Alu elements (short interspersed element, SINE), comprising 11% of the genome, while LINE-1 (long interspersed element-1, LINE-1, or L1) insertions, covering 17% of the human genome, only 10% of STR mutations are found [102]. That is, a family analysis of three generations revealed that a fifth of all new STR mutations occur in retrotransposon sequences.

Involvement of microsatellites in genomic variability and organization of the interphase nucleus. Relationship between microsatellites and mobile genetic elements. Close relationships between STRs and mobile genetic elements—transposons (TEs) have been identified quite a long time ago [103]. Many microsatellites arose from genomic TE insertions. It is assumed that

this may be the result of a number of events: tandem insertions of TEs into certain regions of the host genome (duplicated target sites, target site duplication, TSD), the presence of direct and inverted repeats in TEs and interactions between them, captures of host genome sequences on the flanks of TEs , recombinations between different TEs. The extensive interactions of TEs with each other lead to the conclusion that they can be reconstructed into repeated noncoding or coding sequences. This suggests an evolutionary relationship between such DNA sequences and that the evolution of genomes involved frequent shuffling of repetitive sequences, a process referred to as DNA remodeling [103-105]. Multiple TE insertions are caused by the presence of preferred sites for such integration in target genomes, resulting in the appearance of new TE recombination products formed at a high rate during periods of active transposition. In other words, the TE transpositions itself regularly generates sequences from which new microsatellites can arise [58, 105].

Thus, non-autonomous short dispersed Alu retrotransposons (SINEs), containing a poly(A) tail and a central linker region rich in adenines, are widespread in the human genome. Significant connections are shown between the 3' ends of Alu sequences not only with mononucleotide repeats $(A)_n$, but also with $(AAC)_n$, $(AAT)_n$ and tetra- and hexanucleotide repeats enriched in A-nucleotides, while for dinucleotide repeats $(AT)_n$ such a connection is significantly weaker [106]. The localization of dinucleotide repeats $(AC)_n$ is also preferentially associated with Alu elements, with 75% of them identified at the 3′-end of the element, while the rest are in the central region. Interestingly, $(GAA)_n$, a trinucleotide repeat whose amplification is observed in Friedreich's ataxia, may have arisen with the participation of the Alu element. Of the 788 loci in the human genome containing $(GAA)_n$ repeats, 63% (501 loci) have homology with Alu of at least 25 bp. Among them, 94% are associated with the poly(A) tail, and the rest are associated with either the 5′-end of the element or the central region [106]. In Carnivore species studied, several hundred tRNALys-derived SINEs have been identified that contain microsatellite repeats predominantly enriched in AG and A [107]. In a number of species, including fur seal (*Phoca vitulina concolour*), cattle (Bos taurus), CA/GT microsatellite sequences of varying lengths are found flanking the most common autonomous TE, the L1 [108].

It is known that in genomes TEs form areas of preferential localization, so-called nests [109, 110], and both TEs themselves and their integration sites contain microsatellite sequences, due to which recombination occurs here and new TE variants arise.

It is important to highlight that the ability for horizontal transfer between different taxa has been described for some TEs (e.g., L1 and BovB). They are especially common in mammals, and their representation in the genome can vary significantly even among groups of animals that are relatively close in origin [111].

Some STRs show a direct relationship between their sequence and TE. For example, microsatellite $(AGC)_n$ is more common in cattle and sheep than in other mammalian species. In particular, in cattle, the representation of microsatellite loci with the AGC core is 90 and 142 times higher, respectively, than in humans and dogs [112]. Moreover, in the cattle genome, 39% of such microsatellite loci are directly associated with the Bov-A2 retrotransposon (part of BovB) [112, 113], the evolutionarily young and species-specific for cattle. Interestingly, in cattle, Bov-A2 functions as an enhancer of type II interferon gene expression [113]. The close association of $(AGC)_n$ with Bov-A2 turned out to be specific for this STR and TE; in approximately 60% of other STRs, no genetic linkage with TE was detected [112].

Thus, both the TEs themselves and their flanks contain STR sequences,

and in some cases they are closely related to each other. The variability of STR and TE is determined by various events and mechanisms (recombination processes, the prevalence of the nesting principle of TE localization in each other), but each time the STR and TE loci have a pronounced individuality in the speed and features of evolution, which is apparently due to structural and functional characteristics and the action of selection factors.

Microsatellites, sites of increased chromosome fragility — karyotype evolution. In humans, approximately 230 sites of potential increased fragility containing STR have been described [114]. Ten fragile regions of human chromosomes have been identified, for which expansions of gene-specific tandem repeats with core motifs CGG and CCG are shown. The authors of this study suggest that increasing STR lengths may lead to the emergence of new sites of increased fragility [114].

The tuco-tuco genus (*Ctenomys*) of subterranean rodents (*Rodentia*: *Ctenomyidae*) contains about 65 species, which exhibit the most significant chromosomal variations among mammals (from $2n = 10$ to $2n = 70$). Moreover, karyotypic variability is possible even within a species, for example, $2n$ in *C. minutus* can vary from 42 to 50, in *C. talarum* from 44 to 48, and in *C. lami* from 54 to 58 [115]. Among them, *C. minutus* stands out, with 45 different cytotypes already identified, of which seven are believed to be the original ones (they are common in the coastal plains of Southern Brazil). In tuco-tuco, repeating DNA regions, including microsatellites and LINE-1, were mapped, and a direct connection was revealed between the localization of STRs with different core motifs and LINE-1, on the one hand, and karyotypic variability (formation of cytotypes) on the other in different populations within species [115]. It is important to emphasize that the described cytotypes included not only Robertsonian translocations, chromosome fusions and splits, but also tandem repeats, paracentric and pericentric inversions. The involvement of centromeric and telomeric repeats in intraspecific variability and the formation of intraspecific chromosomal races is well known in a number of shrew species [116]; three variants of intraspecific chromosomal races ("Robertsonian fans") were described by N.N. Vorontsov [117] in mice of the genus *Leggada*, house mice *Mus domesticus* of the superspecies *Mus musculus* and mole voles of the group *Ellobius tancrei*, belonging to the superspecies *Ellobius talpinus*. The involvement of STR and TE in karyotypic variation in fish has been described [118]. Thus, a comparative analysis of the sequenced genomes of three fish species showed that the commercial species *Solea senegalensis* has undergone extensive chromosomal evolution associated with the localization of STR and TE in areas with an increased density of chromosomal rearrangements.

Bursts of rapid karyotype evolution, often referred to as karyotypic megaevolution or chromosomal tachythely, have been found across taxa [119]. Apparently, the most obvious example is provided by two species of deer, the karyotypes of which differ sharply, these are the Indian muntjac *Muntiacus muntjak* ($2n = 6$) and the Chinese muntjac *M. reevesi* ($2n = 46$). Comparative analysis of the sequenced genomes of muntjac, red deer (*Cervus elaphus)* and cattle (*Bos taurus*) confirms the evolutionary sequence of chromosome divisions and fusions described cytogenetically. It was found that since the divergence of deer and cattle species (apprx. 20 million years ago), the rapid evolution of the Indian muntjac karyotype has not been accompanied by major inversions or other internal rearrangements (except for discrete events of splitting and fusion of chromosomes) [119].

Chromosome-level genome comparisons made for *Hydropotes inermi*s (water deer, $2n = 70$), *Muntiacus reevesi* ($2n = 46$), female and male *M. crinifrons* (black mujac, $2n = 8$ or 9) and *M. gongshanensis* (Gongshan Mountains deer, $2n = 8$ or 9) [120], led the authors to conclusion that unique centromeric satellite

repeats, including STRs, telomeric STRs, and palindromic repeats, could be responsible for repeated chromosome fusions in deer of these species [120].

In order to reconstruct the karyotypes of 16 phylogenetic nodes of mammals, including the karyotype of their common ancestor, large-scale studies of genomic sequences were carried out in 32 species belonging to eutherians (19 orders), marsupials and monotremes (3 orders each) as representatives of three superorders of mammals, the *Euarchonta*, *Laurasiatheria* and *Xenarthra*, respectively [121]. Modern species in which the number of chromosomal rearrangements have been estimated relative to the putative common ancestor of mammals are humans, sloths (*Choloepus didactylus*), and cattle. The findings suggest that the common ancestor of mammals probably had 19 pairs of autosomes. However, nine of the smallest chromosomes were shared by the common ancestor of mammals and the common ancestor of all amniotes (of which three chromosomes are still conserved in extant mammals) [121]. The number and types of chromosomal rearrangements for transitions between the karyotype of mammalian ancestors, descendant ancestors and existing species were determined. Common regions of increased rates of evolutionary transformations of chromosomes (evolutionary breakpoint regions, EBRs) and evolutionarily conserved blocks (homologous synteny blocks, HSBs) have been identified [121]. It turned out that EBR regions differ from HSB regions in the increased density of actively transcribed genes and repeating elements. There is a non-random distribution of EBRs in genomes and their association with fragile sites during tumorigenesis [121]. The high density of expressed genes detected in the EBR, as the authors suggest, may explain the increased tendency for DNA double strand breaks (DNA in open transcriptionally active regions of chromatin is more sensitive to damage). Analysis showed that the EBR regions are enriched in genes whose products are involved in sensory perception and transcriptional regulation, whereas the HSB blocks have an increased density of genes involved in the formation of anatomical characteristics and the development of the central nervous system [121]. EBR has a significantly higher density of repeats of all types, segmental duplications, SINE (SINE; all SINE and Alu), LINE (LINE; L1), and long terminal repeats (LTR; all LTR and endogenous retrovirus 1, ERV1) than HSB [121].

According to the same authors [121], the evolutionary history of chromosomes of ancestral mammals (mammalian ancestor chromosomes, MAM) varied significantly depending on the size of the chromosomes. Large long MAMs were more often involved in chromosomal rearrangements than short ones, and some extant species (e.g., *Mus musculus*, *Equus caballus*, *Canis familiaris*, *Bos taurus*, and *Capra hircus*) maintained gene synteny at the chromosome level. Nine of the 14 small MAMs in the listed mammalian species turned out to be orthologous in gene synteny with the chromosomes of *Gallus gallus* and the reconstructed chromosomes of the ancestors of birds and amniotes. Some MAMs were conserved as individual chromosomes or as closed units (i.e., entire chromosomes fused to one or more chromosomes without breaking synteny) in mammalian genomes. For example, MAM7 was conserved as an entire chromosome in *Oryctolagus cuniculus*, *Rhinolophus luctus*, and *Procavia capensis*, which represent three orders of mammals [122]. MAM13 and MAM14 are present as distinct chromosomal units in more than 15 extant mammalian species [121].

Taken together, these results demonstrate striking conservation of synteny over the approximately 320 million years of vertebrate evolution since the common ancestor of all amniotes. The reconstructed genome of the ancestors of mammals showed that the existing genomes of mammals are a mosaic obtained as a result of the evolutionary shuffling of 2557 syntenic segments, which is from 69 to 94% of the genome size in the analyzed species, and EBR sequences enriched in TE

act as links between such segments and STR [121]. Analyzing the data obtained, the authors suggested [121] that evolutionarily conserved syntenic segments, the HSBs serve as the main building blocks (genomic elements similar to elements of the periodic system of chemical elements) for the genomes of all mammals, preserving, along with synteny, biological functions [121]. A similar assumption was formulated in studies of the characteristics of the organization and distribution of microchromosomes in reptiles, birds and mammals [122]. It has been shown that the genomes of birds and reptiles, but not mammals, consist of several large macrochromosomes from 3 to 6 microns in length and many tiny microchromosomes, less than 0.5 μm. Microchromosomes have centromeric and telomeric regions, carry a large number of genes, are enriched in GC nucleotides, are highly conserved among birds and reptiles, and have homology with one or more tiny chromosomes of invertebrates that diverged from vertebrates more than 680 million years ago. Microchromosomes associate with each other, are TE-poor and cluster together in the center of interphase nuclei, which, according to the authors, indicates functional coherence. In turtles, snakes and lizards, many microchromosomes disappeared due to fusion into macrochromosomes; in most mammals, microchromosomes disappeared completely, but some platypus chromosomes coincide with several previously described microchromosomes of reptiles and birds. This suggests that such chromosomes represent the building blocks of mammalian chromosomes, the connection between which is formed by the participation of TE and STR [121, 122].

STR and TE are directly involved in the architecture of the interphase nucleus. In the interphase nucleus, chromatin is organized in the form of a hierarchy from nucleosomes to chromatin domains (CD), then to the formation of topologically associated domains (TADs) and to higher-level compartments; The top of the hierarchy is the so-called chromosomal territories (CT) [123]. According to modern concepts, chromatin organization is a critical factor regulating gene expression [123-125]. Enhancers interact with target genes almost exclusively within the TAD, the distally located co-expressed genes are recruited into common protein clusters upon activation, and compact domains exhibit movement and configurational changes in vivo [124, 125].

The non-random radial positioning of CTs in the nucleus indicates the possibility of preferential patterns of interaction between chromosomal territories. Their ability to form specific interchromosomal networks has been discovered, which change during the cell cycle, during cell differentiation, and during neoplastic transformation [125]. It is assumed that the dynamics of these networks correlate with the global control of structural changes and regulation of the functional activity of the genome. The tendency to various translocations in pathologies can be explained by the close and normally demonstrating specificity of the location of certain chromosomal regions during the co-expression of genes localized in them. It is possible that genomic regions from the EBR are characterized by predominantly open euchromatin, which promotes epigenetic modifications due to the availability of DNA for regulation and active gene function and/or their interactions [124].

The participation of STR and TE in the regulation of changes in gene expression programs through dynamic changes in the architecture of the interphase nucleus has been reported in many studies [123-125]. In our opinion, the most striking example is a study performed on the interphase nuclei of columnar photoreceptor cells in nocturnal mammals [125]. The authors revealed an inversion in the localization of heterochromatin and euchromatin compared to the nuclei of ganglion cells. The heterochromatin was located inside while the euchromatin on the periphery of the nucleus under the lamina [125]. Typically, SINEs are

associated with actively transcribed regions, LINEs with heterochromatized ones, localized at the periphery of the interphase nucleus under the nuclear envelope, and only in the case of cylindrical photoreceptor cells of nocturnal mammals is LINE, together with heterochomatin, concentrated in the center of the interphase nucleus, SINEs at the periphery. According to the authors of the study [125], this arrangement of hetero- and euchromatin is largely due to contacts between homologous dispersed repeats, which are localized in different regions of chromatin.

It has been suggested that STRs can be considered as components of the so-called "genome packaging code" which determined the features of its condensation depending on the cell type [126]. It has been noted that in some organisms STRs are grouped in certain regions of chromosomes, for example, in pericentromeric or subtelomeric regions; in complex genomes, STRs are distributed throughout its entire length in a non-random manner, located predominantly in intergenic spaces [57, 126]. Several proteins are known that specifically bind to STR [125, 126]. Therefore, STRs can be "anchors" for the involvement of groups of loci with which they are linked in the processes of intra- and interchromosomal interactions, the formation of TAD and CD. The involvement of STR in intercellular interactions in complex organs through its putative influence on the architecture of chromatin packaging and, as a consequence, on gene expression programs, has been discussed in a number of works, in particular in the case of microsatellites with the GATA core motif in animals and plants [126, 128]. It has been shown that GAGA repeats, which are abundantly present in the eukaryotic genome, are recognized by the GAGA-associated factor GAF, which influences chromatin packaging (GAGA pioneer factor, GAF) in Drosophila, BBR proteins (barley B recombinant protein) in barley, GBP (GAGA- binding protein) in soybean and rice [126, 128].

The hypothesis of the existence of such a "microsatellite code" of gene expression programs [126] is also important for understanding the molecular genetic mechanisms of the formation of convergent characters in evolutionarily distant taxa. There have been numerous attempts to find common structural genes for such traits [129]. In particular, a comparison was made of transcriptomes in eight vertebrate species (lizards, mammals, sharks) that carry embryos in the uterus [129]. It turned out that in all viviparous groups the basic set of physiological functions of the uterus does not differ, but in the same set of genes, none is expressed specifically for all viviparous lineages or even in all lineages of viviparous amniotes that form the embryonic membranes [129]. Thus, the morphological and physiological traits necessary for successful pregnancy in distantly related vertebrates turn out to be controlled by different genes. Apparently, evolutionary changes in viviparity as a mode of procreation occurred multiple times, involving different genes due to cooperation and collaterality of metabolic pathways, but the set of such genes was still initially limited by their composition in the ancestors of each lineage [129].

The example of viviparity can clearly explain the relatively low efficiency of marker-assisted selection (MAS), where the selection and selection of animals for crosses is based on genotypes for a small number of protein-coding genes [130]. That is why the search for DNA markers of regulatory networks, the variability of which underlies the organization of gene expression profiles, is of particular importance. Accumulated data indicate that, despite chromosomal rearrangements and structural transformations due to the interaction between macro- and microchromosomes, fairly high evolutionary conservation in the gene composition of chromosomal regions in TAD A (actively transcribed domains of interphase chromatin) and TAD B (heterochromatized domains) remains [121, 123-125]. This suggests the presence of a spectrum of regulatory elements involved in such a

division. One of these elements is STR involved in the structural interactions of macro- and microchromosomes and in the formation of the architecture of the interphase nucleus, which, in turn, is closely related to the modulation of gene expression profiles.

Summarizing the discussion of the use of DNA markers, it is important to note that many issues that are significant for breeding are already being resolved using modern molecular genetic methods. The key stages of the breeding process are the selection and selection of animals for crossing and the assessment of the breeding value of the parents based on the characteristics of the offspring. DNA markers make it possible to exclude errors of origin and facilitate the identification of mutations associated with phenotypic and reproductive defects, resistance to biotic factors and environmental stress. If the desired development of economically valuable traits is controlled by a small number of key genes, DNA marking is applicable to search for the corresponding allelic variants, usually associated with variability in the quality of the final product. Attempts to detect associations between sets of SNP genotypes and variability of phenotypic traits are not always successful due to the complex design of gene networks, competitive relationships between molecular genetic structures that serve as targets for natural and artificial selection factors, and variability in the contribution of elements of regulatory networks to interactions between metabolic pathways depending on the genotypic environment and the influence of external factors. "Elusive master genes" of quantitative traits and differences in the genetic control of similar phenotypic characteristics are also examples of the fact that DNA marking of genes and loci associated with a desired trait is not always sufficient for successful selection.

Note that among animal species, the number of protein-coding genes varies relatively little, while variations in genome size are significant and are due to differences in the prevalence of dispersed and tandem repeats in them [110, 126]. Tandem repeats (particularly microsatellites) account for more nucleotides in the mammalian genome than protein-coding genes [126]. The division of the genomes of various animal taxa into evolutionarily conserved (HSB) and evolutionarily unstable (EBR) gene blocks, and the enrichment of the latter with dispersed repeats [121], suggests the direct involvement of dispersed repeats and their derivatives, such as STRs, in the regulation of gene networks. The wide representation of microsatellite repeats in genomes and the diversity of their biological effects distinguish these DNA markers as elements of regulatory networks and, possibly, independent targets of variability and selection. Polylocus genotyping of these particular genomic elements will make it possible to analyze the population genetic structure of animal groups with high resolution, assess the degree of their consolidation and identify differences from closely related groups.

So, one of the central problems in modern animal husbandry remains the search for genetic markers that would simplify the selection and selection of animals for crossing and increase the likelihood of obtaining offspring with desirable economically valuable traits. Developed DNA markers of different types and generations are successfully used to solve a number of issues that are significant for breeding. Microsatellites (STR) are used to exclude errors of origin, mononucleotide polymorphisms (SNP) for mapping genomic regions associated with phenotypic characteristics and adaptation to the pressure of natural selection on edges of habitats and in areas of risky livestock farming. Areas with increased copy number variation (CNV) are involved in analysis of polygenic system responses to the factors of natural and artificial selection. With a small number of main genes that determine the manifestation of a trait, DNA marking is applicable for searching for allelic variants of structural genes and analyzing the variability of elements of regulatory networks and the relationships between metabolic pathways. Long-term

studies have revealed the multiple involvement of STR in basic processes (replication, repair, transcription, translation, adaptation and morphogenesis, epigenetic effects) that determine the stability, variability and evolution of biological systems. In this regard, STRs can be considered as elements of regulatory networks, being the main targets of natural and artificial selection. Polylocus genotyping based on microsatellite and dispersed repeats seems promising for analyzing the population genetic structure and consolidation of animal groups and their differences from closely related groups.

REFERENCES

1. Serebrovskiy A.S. *Geneticheskiy analiz* [Genetic analysis]. Moscow, 1970 (in Russ.).
2. **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).
3. Thakur J., Packiaraj J., Henikoff S. Sequence, chromatin and evolution of satellite DNA. *Int. J. Mol. Sci.,* 2021, 22: 4309 (doi: 10.3390/ijms22094309).
4. Li C., Tian D., Tang B., Liu X., Teng X., Zhao W., Zhang Z., Song S. Genome variation map: a worldwide collection of genome variations across multiple species. *Nucleic Acids Res.*, 2021, 49(D1): D1186-D1191 (doi: 10.1093/nar/gkaa1005).
5. Wang Z.H., Zhu Q.H., Li X., Zhu J.W., Tian D.M., Zhang S.S., Kang H.L., Li C.P., Dong L.L., Zhao W.M., Li M.H. iSheep: an integrated resource for sheep genome, variant and phenotype. *Front. Genet.*, 2021, 12: 714852 (doi: 10.3389/fgene.2021.714852).
6. Kirichenko A.V., Zlobin A.S., Shashkova T.I., Volkova N.A., Iolchiev B.S., Bagirov V.A., Borodin P.M., Karssen L.S., Tsepilov Y.A., Aulchenko Y.S. The GWAS-MAP|ovis platform for aggregation and analysis of genome-wide association study results in sheep. *Vavilovskii Zhurnal Genetiki i Selektsii*, 2022, 26(4): 378-384 (doi: 10.18699/VJGB-22-46).
7. Pös O., Radvanszky J., Buglyy G., Pös Z., Rusnakova D., Nagy B., Szemes T. DNA copy number variation: main characteristics, evolutionary significance, and pathological aspects. *Biomed. J.*, 2021, 44(5): 548-559 (doi: 10.1016/j.bj.2021.02.003).
8. Strillacci M.G., Gorla E., Cozzi M.C., Vevey M., Genova F., Scienski K., Longeri M., Bagnato A. A copy number variant scan in the autochthonous Valdostana Red Pied cattle breed and comparison with specialized dairy populations. *PLoS ONE*, 2018, 3(9): e0204669 (doi: 10.1371/journal.pone.0204669).
9. Lee Y.L., Bosse M., Mullaart E., Groenen M.A.M., Veerkamp R.F., Bouwman A.C. Functional and population genetic features of copy number variations in two dairy cattle populations. *BMC Genomics*, 2020, 21(1): 89 (doi: 10.1186/s12864-020-6496-1).
10. Qiu Y., Ding R., Zhuang Z., Wu J., Yang M., Zhou S., Ye Y., Geng Q., Xu Z., Huang S., Cai G., Wu Z., Yang J. Genome-wide detection of CNV regions and their potential association with growth and fatness traits in Duroc pigs. *BMC Genomics*, 2021, 22(1): 332 (doi: 10.1186/s12864-021-07654-7).
11. Davoudi P., Do D.N., Rathgeber B., Colombo S.M., Sargolzaei M., Plastow G., Wang Z., Karimi K., Hu G., Valipour S., Miar Y. Genome-wide detection of copy number variation in American mink using whole-genome sequencing. *BMC Genomics*, 2022, 23(1): 649 (doi: 10.1186/s12864-022-08874-1)
12. Yuan C., Lu Z., Guo T., Yue Y., Wang X., Wang T., Zhang Y., Hou F., Niu C., Sun X., Zhao H., Zhu S., Liu J., Yang B. A global analysis of CNVs in Chinese indigenous fine-wool sheep populations using whole-genome resequencing. *BMC Genomics*, 2021, 22(1): 78 (doi: 10.1186/s12864-021-07387-7).
13. Laseca N., Molina A., Valera M., Antonini A., Demyda-Peyrás S. Copy number variation (CNV): a new genomic insight in horses. *Animals (Basel)*, 2022, 12(11): 1435 (doi: 10.3390/ani12111435).
14. Bhanuprakash V., Chhotaray S., Pruthviraj D.R., Rawat C., Karthikeyan A., Panigrahi M. Copy number variation in livestock: a mini review. *Vet. World*, 2018, 11(4): 535-541 (doi: 10.14202/vetworld.2018.535-541).
15. Seroussi E., Glick G., Shirak A., Yakobson E., Weller J.I., Ezra E., Zeron Y. Analysis of copy loss and gain variations in Holstein cattle autosomes using BeadChip SNPs. *BMC Genomics*, 2010, 11: 673 (doi: 10.1186/1471-2164-11-673).
16. Wang Z., Guo J., Guo Y., Yang Y., Teng T., Yu Q., Wang T., Zhou M., Zhu Q., Wang W., Zhang Q., Yang H. Genome-wide detection of CNVs and association with body weight in sheep based on 600K SNP Arrays. *Front. Genet.*, 2020, 11: 558 (doi: 10.3389/fgene.2020.00558).
17. Taghizadeh S., Gholizadeh M., Rahimi-Mianji G., Moradi M.H., Costilla R., Moore S., Di Gerlando R. Genome-wide identification of copy number variation and association with fat deposition in thin and fat-tailed sheep breeds. *Sci. Rep.*, 2022, 12(1): 8834 (doi: 10.1038/s41598-022-12778-1).

18. Kalds P., Zhou S., Gao Y., Cai B., Huang S., Chen Y., Wang X. Genetics of the phenotypic evolution in sheep: a molecular look at diversity-driving genes. *Genet. Sel. Evol.*, 2022, 54(1): 61 (doi: 10.1186/s12711-022-00753-3).
19. Feng Y., Zhang Y., Ying C., Wang D., Du C. Nanopore-based fourth-generation DNA sequencing technology. *Genom. Proteom. Bioinform.*, 2015, 13: 4-16 (doi: 10.1016/j.gpb.2015.01.009).
20. Suminda G.G.D., Ghosh M., Son Y.O. The innovative informatics approaches of high-throughput technologies in livestock: spearheading the sustainability and resiliency of agrigenomics research. *Life (Basel),* 2022, 12(11):1893 (doi: 10.3390/life12111893).
21. Nolte W., Alkhoder H., Wobbe M., Stock K.F., Kalm E., Vosgerau S., Krattenmacher N., Thaller G., Tetens J., Kühn C. Replacement of microsatellite markers by imputed medium-density SNP arrays for parentage control in German warmblood horses. *J. Appl .Genet.,* 2022, 63(4): 783-792 (doi: 10.1007/s13353-022-00725-9).
22. Marina H., Suarez-Vega A., Pelayo R., Gutiérrez-Gil B., Reverter A., Esteban-Blanco C., Arranz J.J. Accuracy of imputation of microsatellite markers from a 50K SNP chip in Spanish Assaf sheep. *Animals (Basel)*, 2021, 11(1): 86 (doi: 10.3390/ani11010086).
23. McClure M.C., Sonstegard T.S., Wiggans G.R., Van Eenennaam A.L., Weber K.L., Penedo C.T., Berry D.P., Flynn J., Garcia J.F., Carmo A.S., Regitano L.C., Albuquerque M., Silva M.V., Machado M.A., Coffey M., Moore K., Boscher M.Y., Genestout L., Mazza R., Taylor J.F., Schnabel R.D., Simpson B., Marques E., McEwan J.C., Cromie A., Coutinho L.L., Kuehn L.A., Keele J.W., Piper E.K., Cook J., Williams R., Bovine HapMap Consortium, Van Tassell C.P. Imputation of microsatellite alleles from dense SNP genotypes for parentage verification across multiple *Bos taurus* and *Bos indicus* breeds. *Front Genet.*, 2013, 4: 176 (doi: 10.3389/fgene.2013.00176).
24. McClure M., Sonstegard T., Wiggans G., Van Tassell C. Imputation of microsatellite alleles from dense SNP genotypes for parental verification. *Front. Genet.,* 2012, 3: 140 (doi: 10.3389/fgene.2012.00140).
25. Xu L., Haasl R.J., Sun J., Zhou Y., Bickhart D.M., Li J., Song J., Sonstegard T.S., Van Tassell C.P., Lewin H.A., Liu G.E. Systematic profiling of short tandem repeats in the cattle genome. *Genome Biol. Evol.*, 2017, 9(1): 20-31 (doi: 10.1093/gbe/evw256).
26. Jiang Y., Song H., Gao H., Zhang Q., Ding X. Exploring the optimal strategy of imputation from SNP array to whole-genome sequencing data in farm animals. *Front. Genet.*, 2022, 13: 963654 (doi: 10.3389/fgene.2022.963654).
27. Crysnanto D., Leonard A.S., Fang Z.H., Pausch H. Novel functional sequences uncovered through a bovine multiassembly graph. *Proc. Natl. Acad. Sci. USA*, 2021, 118(20): e2101056118 (doi: 10.1073/pnas.2101056118).
28. Chen S., Liu S., Shi S., Jiang Y., Cao M., Tang Y., Li W., Liu J., Fang L., Yu Y., Zhang S. Comparative epigenomics reveals the impact of ruminant-specific regulatory elements on complex traits. *BMC Biol.*, 2022, 20(1): 273 (doi: 10.1186/s12915-022-01459-0).
29. Li J., Xiang Y., Zhang L., Qi X., Zheng Z., Zhou P., Tang Z., Jin Y., Zhao Q., Fu Y., Zhao Y., Li X., Fu L., Zhao S. Enhancer-promoter interaction maps provide insights into skeletal muscle-related traits in pig genome. *BMC Biol.*, 2022, 20(1): 136 (doi: 10.1186/s12915-022-01322-2).
30. Hayes B.J., Bowman P.J., Chamberlain A.J., Savin K., van Tassell C.P., Sonstegard T.S., Goddard M.E. A validated genome wide association study to breed cattle adapted to an environment altered by climate change. *PLoS ONE,* 2009, 4(8): e6676 (doi: 10.1371/journal.pone.0006676).
31. Cheruiyot E.K., Haile-Mariam M., Cocks B.G., Pryce J.E. Improving genomic selection for heat tolerance in dairy cattle: current opportunities and future directions. *Front. Genet.*, 2022, 13: 894067 (doi: 10.3389/fgene.2022.894067).
32. Buggiotti L., Yurchenko A.A., Yudin N.S., Vander Jagt C.J., Vorobieva N.V., Kusliy M.A., Vasiliev S.K., Rodionov A.N., Boronetskaya O.I., Zinovieva N.A., Graphodatsky A.S., Daetwyler H.D., Larkin D.M. Demographic history, adaptation, and NRAP convergent evolution at amino acid residue 100 in the world northernmost cattle from Siberia. *Mol. Biol. Evol.*, 2021, 38(8): 3093-3110 (doi: 10.1093/molbev/msab078).
33. Buggiotti L., Yudin N.S., Larkin D.M. Copy number variants in two northernmost cattle breeds are related to their adaptive phenotypes. *Genes*, 2022, 13: 1595 (doi: 10.3390/ genes13091595).
34. Lemay D.G., Lynn D.J., Martin W.F., Neville M.C., Casey T.M., Rincon G., Kriventseva E.V., Barris W.C., Hinrichs A.S., Molenaar A.J., Pollard K.S., Maqbool N.J., Singh K., Murney R., Zdobnov E.M., Tellam R.L., Medrano J.F., German J.B., Rijnkels M. The bovine lactation genome: insights into the evolution of mammalian milk. *Genome Biol.*, 2009, 10(4): R43 (doi: 10.1186/gb-2009-10-4-r43).
35. Wickramasinghe S., Rincon G., Islas-Trejo A., Medrano J.F. Transcriptional profiling of bovine milk using RNA sequencing. *BMC Genomics*, 2012, 13: 45 (doi: 10.1186/1471-2164-13-45).
36. Hammami H., Rekik B., Bastin C., Soyeurt H., Bormann J., Stoll J, Gengler N. Environmental sensitivity for milk yield in Luxembourg and Tunisian Holsteins by herd management level. *J. Dairy Sci.*, 2009, 92(9): 4604-4612 (doi: 10.3168/jds.2008-1513).
37. Cai W., Li C., Li J., Song J., Zhang S. Integrated small RNA sequencing, transcriptome and GWAS data reveal microRNA regulation in response to milk protein traits in Chinese Holstein cattle. *Front. Genet.,* 2021, 12: 726706 (doi: 10.3389/fgene.2021.726706).

38. Dysin A.P., Barkova O.Y., Pozovnikova M.V. The role of microRNAs in the mammary gland development, health, and function of cattle, goats, and sheep. *Noncoding RNA*, 2021, 7(4): 78 (doi: 10.3390/ncrna7040078).
39. Ren W., Zhang Y., Dingkao R., Huang C., Ma X., Wu X., La Y., Chu M., Bao P., Guo X., Pei J., Yan P., Liang C. Comparative study of the expression profiles of miRNAs of milk-derived exosomes of yak and Jeryak. *Animals (Basel)*, 2022, 12(22): 3189 (doi: 10.3390/ani12223189).
40. Deb R., Sengar G.S. Comparative miRNA signatures among Sahiwal and Frieswal cattle breeds during summer stress. *Biotech.,* 2021, 11(2): 79 (doi: 10.1007/s13205-020-02608-4).
41. 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).
42. Liu Y., Xu J., Chen M., Wang C., Li S. A unified STR profiling system across multiple species with whole genome sequencing data. *BMC Bioinformatics*, 2019, 20(Suppl. 24): 671 (doi: 10.1186/s12859-019-3246-y).
43. Cui W., Jin X., Guo Y., Chen C., Zhang W., Wang Y., Lan J., Zhu B. Development and validation of a novel five-dye short tandem repeat panel for forensic identification of 11 species. *Front. Genet.*, 2020, 11: 1005 (doi: 10.3389/fgene.2020.01005).
44. Forutan M., Ansari Mahyari S., Baes C., Melzer N., Schenkel F.S., Sargolzaei M. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. *BMC Genomics*, 2018, 19(1): 98 (doi: 10.1186/s12864-018-4453-z).
45. Häfliger I.M., Spengeler M., Seefried F.R., Drögemüller C. Four novel candidate causal variants for deficient homozygous haplotypes in Holstein cattle. *Sci. Rep.*, 2022, 12(1): 5435 (doi: 10.1038/s41598-022-09403-6).
46. Liu X., Tian D., Li C., Tang B., Wang Z., Zhang R., Pan Y., Wang Y., Zou D., Zhang Z., Song S. GWAS Atlas: an updated knowledgebase integrating more curated associations in plants and animals. *Nucleic Acids Res.*, 2023, 51(D1): D969-D976 (doi: 10.1093/nar/gkac924).
47. CNCB-NGDC Members and Partners. Database resources of the National Genomics Data Center, China National Center for Bioinformation in 2023. *Nucleic Acids Res.*, 2023, 51(D1): D18-D28 (doi: 10.1093/nar/gkac1073).
48. Rigden D.J., Fernández X.M. The 2023 Nucleic Acids Research Database Issue and the online molecular biology database collection. *Nucleic Acids Res.*, 2023, 51(D1): D1-D8 (doi: 10.1093/nar/gkac1186).
49. Flori L., Fritz S., Jaffrézic F., Boussaha M., Gut I., Heath S., Foulley J.L., Gautier M. The genome response to artificial selection: a case study in dairy cattle. *PLoS ONE*, 2009, 4(8): e6595 (doi: 10.1371/journal.pone.0006595).
50. Prakapenka D., Liang Z., Jiang J., Ma L., Da Y. A large-scale genome-wide association study of epistasis effects of production traits and daughter pregnancy rate in U.S. Holstein cattle. *Genes (Basel)*, 2021, 12(7): 1089 (doi: 10.3390/genes12071089).
51. Dachs N., Upadhyay M., Hannemann E., Hauser A., Krebs S., Seichter D., Russ I., Gehrke L.J., Thaller G., Medugorac I. Quantitative trait locus for calving traits on *Bos taurus* autosome 18 in Holstein cattle is embedded in a complex genomic region. *J. Dairy Sci.*, 2023, 106(3): 1925-1941 (doi: 10.3168/jds.2021-21625).
52. Huang Y., Li Y., Wang X., Yu J., Cai Y., Zheng Z., Li R., Zhang S., Chen N., Asadollahpour Nanaei H., Hanif Q., Chen Q., Fu W., Li C., Cao X., Zhou G., Liu S., He S., Li W., Chen Y., Chen H., Lei C., Liu M., Jiang Y. An atlas of CNV maps in cattle, goat and sheep. *Sci. China Life Sci.*, 2021, 64(10): 1747-1764 (doi: 10.1007/s11427-020-1850-x).
53. Revay T., Quach A.T., Maignel L., Sullivan B., King W.A. Copy number variations in high and low fertility breeding boars. *BMC Genomics*, 2015, 16(1): 280 (doi: 10.1186/s12864-015-1473-9).
54. Fontanesi L., Martelli P.L., Scotti E., Russo V., Rogel-Gaillard C., Casadio R., Vernesi C. Exploring copy number variation in the rabbit (*Oryctolagus cuniculus*) genome by array comparative genome hybridization. *Genomics*, 2012, 100(4): 245-251 (doi: 10.1016/j.ygeno.2012.07.001).
55. Davoudi P., Do D.N., Rathgeber B., Colombo S.M., Sargolzaei M., Plastow G., Wang Z., Karimi K., Hu G., Valipour S., Miar Y. Genome-wide detection of copy number variation in American mink using whole-genome sequencing. *BMC Genomics*, 2022, 23(1): 649 (doi: 10.1186/s12864-022-08874-1).
56. Khlopova N.S., Glazko T.T., Glazko V.I. Constitutive and variable components of gene expression profiles in pig liver. *Russ. J. Genet. Appl. Res.*, 2011, 1: 302-307 (doi: 10.1134/S2079059711040046).
57. Avvaru A.K., Sharma D., Verma A., Mishra R.K., Sowpati D.T. MSDB: a comprehensive, annotated database of microsatellites. *Nucleic Acids Res.*, 2020, 48(D1): D155-D159 (doi: 10.1093/nar/gkz886).
58. Zattera M.L., Bruschi D.P. Transposable elements as a source of novel repetitive DNA in the eukaryote genome. *Cells*, 2022, 11(21): 3373 (doi: 10.3390/cells11213373).
59. Gharesouran J., Hosseinzadeh H., Ghafouri-Fard S., Taheri M., Rezazadeh M. STRs: ancient architectures of the genome beyond the sequence. *J. Mol. Neurosci.*, 2021, 71(12): 2441-2455 (doi: 10.1007/s12031-021-01850-6).

60. Verbiest M., Maksimov M., Jin Y., Anisimova M., Gymrek M., Bilgin Sonay T. Mutation and selection processes regulating short tandem repeats give rise to genetic and phenotypic diversity across species. *J. Evol. Biol.*, 2023, 36(2): 321-336 (doi: 10.1111/jeb.14106).
61. Astolfi P., Bellizzi D., Sgaramella V. Frequency and coverage of trinucleotide repeats in eukaryotes. *Gene*, 2003, 317(1-2): 117-125 (doi: 10.1016/s0378-1119(03)00659-0).
62. Arslan A. Compendious survey of protein tandem repeats in inbred mouse strains. *BMC Genom. Data*, 2022, 23(1): 62 (doi: 10.1186/s12863-022-01079-1).
63. Wang X., Budowle B., Ge J. USAT: a bioinformatic toolkit to facilitate interpretation and comparative visualization of tandem repeat sequences. *BMC Bioinformatics*, 2022, 23(1): 497 (doi: 10.1186/s12859-022-05021-1).
64. Jafarian Z., Khamse S., Afshar H., Khorshid H.R.K., Delbari A., Ohadi M. Natural selection at the RASGEF1C (GGC) repeat in human and divergent genotypes in late-onset neurocognitive disorder. *Sci. Rep.*, 2021, 11(1): 19235 (doi: 10.1038/s41598-021-98725-y).
65. Khamse S., Alizadeh S., Bernhart S.H., Afshar H., Delbari A., Ohadi M. A (GCC) repeat in *SBF1* reveals a novel biological phenomenon in human and links to late onset neurocognitive disorder. *Sci. Rep.*, 2022, 12(1): 15480 (doi: 10.1038/s41598-022-19878-y).
66. Khamse S., Jafarian Z., Bozorgmehr A., Tavakoli M., Afshar H., Keshavarz M., Moayedi R., Ohadi M. Novel implications of a strictly monomorphic (GCC) repeat in the human *PRKACB* gene. *Sci. Rep.*, 2021, 11(1): 20629 (doi: 10.1038/s41598-021-99932-3).
67. Khamse S., Arabfard M., Salesi M., Behmard E., Jafarian Z., Afshar H., Khazaei M., Ohadi M. Predominant monomorphism of the RIT2 and GPM6B exceptionally long GA blocks in human and enriched divergent alleles in the disease compartment. *Genetica*, 2022, 150(1): 27-40 (doi: 10.1007/s10709-021-00143-5).
68. Cheng M., Xie D., Price M., Zhou C., Zhang X.. Comparative analysis of microsatellites in coding regions provides insights into the adaptability of the giant panda, polar bear and brown bear. *Genetica*, 2022, 150(6): 355-366 (doi: 10.1007/s10709-022-00173-7).
69. Song X., Yang T., Zhang X., Yuan Y., Yan X., Wei Y., Zhang J., Zhou C.. Comparison of the microsatellite distribution patterns in the genomes of *Euarchontoglires* at the taxonomic level. *Front. Genet.*, 2021, 12: 622724 (doi: 10.3389/fgene.2021.622724).
70. Verbiest M.A., Delucchi M., Bilgin Sonay T., Anisimova M. Beyond microsatellite instability: intrinsic disorder as a potential link between protein short tandem repeats and cancer. *Front. Bioinform.*, 2021, 1: 685844 (doi: 10.3389/fbinf.2021.685844).
71. Gymrek M., Willems T., Guilmatre A., Zeng H., Markus B., Georgiev S., Daly M.J., Price A.L., Pritchard J.K., Sharp A.J., Erlich Y. Abundant contribution of short tandem repeats to gene expression variation in humans. *Nat. Genet.*, 2016, 48(1): 22-29 (doi: 10.1038/ng.3461).
72. Kinney N., Kang L., Bains H., Lawson E., Husain M., Husain K., Sandhu I., Shin Y., Carter J.K., Anandakrishnan R., Michalak P., Garner H. Ethnically biased microsatellites contribute to differential gene expression and glutathione metabolism in Africans and Europeans. *PLoS ONE*, 2021, 16(3): e0249148 (doi: 10.1371/journal.pone.0249148).
73. Basu S., Mackowiak S.D., Niskanen H., Knezevic D., Asimi V., Grosswendt S., Geertsema H., Ali S., Jerkovi I., Ewers H., Mundlos S., Meissner A., Ibrahim D.M., Hnisz D. Unblending of transcriptional condensates in human repeat expansion disease. *Cell*, 2020, 181(5): 1062-1079.e30 (doi: 10.1016/j.cell.2020.04.018).
74. Paul S., Dansithong W., Figueroa K.P., Scoles D.R., Pulst S.M. Staufen1 links RNA stress granules and autophagy in a model of neurodegeneration. *Nat. Commun.*, 2018, 9(1): 3648 (doi: 10.1038/s41467-018-06041-3).
75. Johnson S.L., Tsou W.L., Prifti M.V., Harris A.L., Todi S.V. A survey of protein interactions and posttranslational modifications that influence the polyglutamine diseases. *Front. Mol. Neurosci.*, 2022, 15: 974167 (doi: 10.3389/fnmol.2022.974167).
76. Loureiro J.R., Oliveira C.L., Silveira I. Unstable repeat expansions in neurodegenerative diseases: nucleocytoplasmic transport emerges on the scene. *Neurobiol. Aging*, 2016, 39: 174-183 (doi: 10.1016/j.neurobiolaging.2015.12.007).
77. Loureiro J.R., Castro A.F., Figueiredo A.S., Silveira I. Molecular mechanisms in pentanucleotide repeat diseases. *Cells*, 2022, 11(2): 205 (doi: 10.3390/cells11020205).
78. Li Y., Lu Y., Polak U., Lin K., Shen J., Farmer J., Seyer L., Bhalla A.D., Rozwadowska N., Lynch D.R., Butler J.S., Napierala M. Expanded GAA repeats impede transcription elongation through the *FXN* gene and induce transcriptional silencing that is restricted to the *FXN* locus. *Hum. Mol. Genet.*, 2015, 24(24): 6932-6943 (doi: 10.1093/hmg/ddv397).
79. Jain A., Vale R.D. RNA phase transitions in repeat expansion disorders. *Nature*, 2017, 546(7657): 243-247 (doi: 10.1038/nature22386).
80. Isiktas A.U., Eshov A., Yang S., Guo J.U. Systematic generation and imaging of tandem repeats reveal base-pairing properties that promote RNA aggregation. *Cell Rep. Methods*, 2022, 2(11): 100334 (doi: 10.1016/j.crmeth.2022.100334).
81. Murmann A.E., Patel M., Jeong S.Y., Bartom E.T., Jennifer Morton A., Peter M.E. The length of uninterrupted CAG repeats in stem regions of repeat disease associated hairpins determines the

amount of short CAG oligonucleotides that are toxic to cells through RNA interference. *Cell Death & Disease*, 2022, 13(12): 1078 (doi: 10.1038/s41419-022-05494-1).
82. Nguyen L., Cleary J.D., Ranum L.P.W. Repeat-associated non-ATG translation: molecular mechanisms and contribution to neurological disease. *Annu. Rev. Neurosci.*, 2019, 42: 227-247 (doi: 10.1146/annurev-neuro-070918-050405).
83. Li Y.C., Korol A.B., Fahima T., Beiles A., Nevo E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol. Ecol.*, 2002, 11(12): 2453-2465 (doi: 10.1046/j.1365-294x.2002.01643.x).
84. Chistiakov D.A., Hellemans B., Volckaert F.A. Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fsh genetics. *Aquaculture*, 2006, 255(1-4): 1-29 (doi: 10.1016/j.aquaculture.2005.11.031).
85. Sun J.X., Helgason A., Masson G., Ebenesersdyttir S.S., Li H., Mallick S., Gnerre S., Patterson N., Kong A., Reich D., Stefansson K. A direct characterization of human mutation based on microsatellites. *Nat. Genet.,* 2012, 44(10): 1161-1165 (doi: 10.1038/ng.2398).
86. Eckert K.A., Hile S.E. Every microsatellite is different: intrinsic DNA features dictate mutagenesis of common microsatellites present in the human genome. *Mol. Carcinog.*, 2009, 48(4): 379-388 (doi: 10.1002/mc.20499).
87. Lynch M. Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci. USA*, 2010, 107(3): 961-968 (doi: 10.1073/pnas.0912629107).
88. Rajaei M., Saxena A.S., Johnson L.M., Snyder M.C., Crombie T.A., Tanny R.E., Andersen E.C., Joyner-Matos J., Baer C.F. Mutability of mononucleotide repeats, not oxidative stress, explains the discrepancy between laboratory-accumulated mutations and the natural allele-frequency spectrum in *C. elegans*. *Genome Res.*, 2021, 31(9): 1602-1613 (doi: 10.1101/gr.275372.121).
89. Brand C.L., Levine M.T. Functional diversification of chromatin on rapid evolutionary timescales. *Annu. Rev. Genet.*, 2021, 55: 401-425 (doi: 10.1146/annurev-genet-071719-020301).
90. Murat P., Guilbaud G., Sale J.E. DNA polymerase stalling at structured DNA constrains the expansion of short tandem repeats. *Genome Biol.*, 2020, 21(1): 209 (doi: 10.1186/s13059-020-02124-x)
91. Stein M., Hile S.E., Weissensteiner M.H., Lee M., Zhang S., Kejnovský E., Kejnovská I., Makova K.D., Eckert K.A. Variation in G-quadruplex sequence and topology differentially impacts human DNA polymerase fidelity. *DNA Repair (Amst),* 2022, 119: 103402 (doi: 10.1016/j.dnarep.2022.103402).
92. Wheeler V.C., Dion V. Modifiers of CAG/CTG repeat instability: insights from mammalian models. *Journal of Huntington's Disease*, 2021, 10(1): 123-148 (doi: 10.3233/JHD-200426).
93. Bansal A., Kaushik S., Kukreti S. Non-canonical DNA structures: diversity and disease association. *Front. Genet.*, 2022, 13: 959258 (doi: 10.3389/fgene.2022.959258).
94. Georgakopoulos-Soares I., Chan C.S.Y., Ahituv N., Hemberg M. High-throughput techniques enable advances in the roles of DNA and RNA secondary structures in transcriptional and post-transcriptional gene regulation. *Genome Biol.*, 2022, 23(1): 159 (doi: 10.1186/s13059-022-02727-6).
95. Kunkler C.N., Schiefelbein G.E., O'Leary N.J., McCown P.J., Brown J.A. A single natural RNA modification can destabilize a U•A-T-rich RNA•DNA-DNA triple helix. *RNA*, 2022, 28(9): 1172-1184 (doi: 10.1261/rna.079244.122).
96. Brown J.A. Unraveling the structure and biological functions of RNA triple helices. *Wiley Interdiscip. Rev. RNA*, 2020, 11(6): e1598 (doi: 10.1002/wrna.1598).
97. Bowater R.P., Bohálová N., Brázda V. Interaction of proteins with inverted repeats and cruciform structures in nucleic acids. *Int. J. Mol. Sci.*, 2022, 23(11): 6171 (doi: 10.3390/ijms23116171).
98. Brázda V., Laister R.C., Jagelská E.B., Arrowsmith C. Cruciform structures are a common DNA feature important for regulating biological processes. *BMC Mol. Biol.*, 2011, 12: 33 (doi: 10.1186/1471-2199-12-33).
99. Gao Z., Moorjani P., Sasani T.A., Pedersen B.S., Quinlan A.R., Jorde L.B., Amster G., Przeworski M. Overlooked roles of DNA damage and maternal age in generating human germline mutations. *Proc. Natl. Acad. Sci. USA*, 2019, 116(19): 9491-9500 (doi: 10.1073/pnas.1901259116).
100. Mitra I., Huang B., Mousavi N., Ma N., Lamkin M., Yanicky R., Shleizer-Burko S., Lohmueller K.E., Gymrek M. Patterns of de novo tandem repeat mutations and their role in autism. *Nature*, 2021, 589(7841): 246-250 (doi: 10.1038/s41586-020-03078-7).
101. Hannan A.J. Tandem repeats mediating genetic plasticity in health and disease. *Nat. Rev. Genet.*, 2018, 19(5): 286-298 (doi: 10.1038/nrg.2017.115).
102. Steely C.J., Watkins W.S., Baird L., Jorde L.B. The mutational dynamics of short tandem repeats in large, multigenerational families. *Genome Biol.*, 2022, 23(1): 253 (doi: 10.1186/s13059-022-02818-4).
103. McGurk M.P., Barbash D.A. Double insertion of transposable elements provides a substrate for the evolution of satellite DNA. *Genome Res.*, 2018, 28(5): 714-725 (doi: 10.1101/gr.231472.117).
104. Nishiyama E., Ohshima K. Cross-kingdom commonality of a novel insertion signature of RTE-related short retroposons. *Genome Biol Evol.*, 2018, 10(6): 1471-1483 (doi: 10.1093/gbe/evy098).
105. Paço A., Freitas R., Vieira-da-Silva A. Conversion of DNA sequences: from a transposable element

to a tandem repeat or to a gene. *Genes (Basel),* 2019, 10(12): 1014 (doi: 10.3390/genes10121014).
106. Richard G.F., Kerrest A., Dujon B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol. Mol. Biol. Rev.*, 2008, 72(4): 686-727 (doi: 10.1128/MMBR.00011-08).
107. López-Giráldez F., Andrés O., Domingo-Roura X., Bosch M. Analyses of carnivore microsatellites and their intimate association with tRNA-derived SINEs. *BMC Genomics*, 2006, 7: 269 (doi: 10.1186/1471-2164-7-269).
108. Duffy A.J., Coltman D.W., Wright J.M. Microsatellites at a common site in the second ORF of L1 elements in mammalian genomes. *Mamm. Genome*, 1996, 7(5): 386-387 (doi: 10.1007/s003359900111).
109. Lexa M., Jedlicka P., Vanat I., Cervenansky M., Kejnovsky E. TE-greedy-nester: structure-based detection of LTR retrotransposons and their nesting. *Bioinformatics*, 2020, 36(20): 4991-4999 (doi: 10.1093/bioinformatics/btaa632).
110. Glazko V.I., Kosovskiy G.Yu., Glazko T.T. **The sources of genome variability as domestication drivers (review)**. *Sel'skokhozyaistvennaya biologiya* [*Agricultural Biology*], 2022, 57(5): 832-851 (doi: 10.15389/agrobiology.2022.5.832eng).
111. Ivancevic A.M., Kortschak R.D., Bertozzi T., Adelson D.L. Horizontal transfer of BovB and L1 retrotransposons in eukaryotes. *Genome Biol.*, 2018, 19(1): 85 (doi: 10.1186/s13059-018-1456-7).
112. Tellam R.L., Lemay D.G., Van Tassell C.P., Lewin H.A., Worley K.C., Elsik C.G. Unlocking the bovine genome. *BMC Genomics*, 2009, 10: 193 (doi: 10.1186/1471-2164-10-193).
113. Kelly C.J., Chitko-McKown C.G., Chuong E.B. Ruminant-specific retrotransposons shape regulatory evolution of bovine immunity. *Genome Res.*, 2022, 32(8): 1474-1486 (doi: 10.1101/gr.276241.121).
114. Mirceta M., Shum N., Schmidt M.H.M., Pearson C.E. Fragile sites, chromosomal lesions, tandem repeats, and disease. *Front. Genet.,* 2022, 13: 985975 (doi: 10.3389/fgene.2022.985975).
115. de Oliveira T.D., Bertocchi N.A., Kretschmer R., de Oliveira E.H.C., Cioffi M.D.B., Liehr T., de Freitas T.R.O. Genomic organization of microsatellites and LINE-1-like retrotransposons: evolutionary implications for *Ctenomys minutus* (*Rodentia*: *Ctenomyidae*) cytotypes. *Animals*, 2022, 12(16): 2091 (doi: 10.3390/ani12162091).
116. Kryshchuk I.A., Orlov V.N., Cherepanova E.V., Borisov Y.M. Unusual chromosomal polymorphism of the common shrew, *Sorex araneus* L., in southern Belarus. *Comparative Cytogenetics*, 2021, 15(2): 159-169 (doi: 10.3897/CompCytogen.v15.i2.63084).
117. Vorontsov N.N. *Razvitie evolyutsionnykh idey v Mezhdunbiologii.* Moscow, 2004 (in Russ.).
118. Ramírez D., Rodríguez M.E., Cross I., Arias-Pérez A., Merlo M.A., Anaya M., Portela-Bens S., Martínez P., Robles F., Ruiz-Rejón C., Rebordinos L. Integration of maps enables a cytogenomics analysis of the complete karyotype in *Solea senegalensis. Int. J. Mol. Sci.*, 2022, 23(10): 5353 (doi: 10.3390/ijms23105353).
119. Mudd A.B., Bredeson J.V., Baum R., Hockemeyer D., Rokhsar D.S. Analysis of muntjac deer genome and chromatin architecture reveals rapid karyotype evolution. *Commun. Biol.*, 2020, 3(1): 480 (doi: 10.1038/s42003-020-1096-9).
120. Yin Y., Fan H., Zhou B., Hu Y., Fan G., Wang J., Zhou F., Nie W., Zhang C., Liu L., Zhong Z., Zhu W., Liu G., Lin Z., Liu C., Zhou J., Huang G., Li Z., Yu J., Zhang Y., Yang Y., Zhuo B., Zhang B., Chang J., Qian H., Peng Y., Chen X., Chen L., Li Z., Zhou Q., Wang W., Wei F. Molecular mechanisms and topological consequences of drastic chromosomal rearrangements of muntjac deer. *Nat. Commun.*, 2021, 12(1): 6858 (doi: 10.1038/s41467-021-27091-0).
121. Damas J., Corbo M., Kim J., Turner-Maier J., Farré M., Larkin D.M., Ryder O.A., Steiner C., Houck M.L., Hall S., Shiue L., Thomas S., Swale T., Daly M., Korlach J., Uliano-Silva M., Mazzoni C.J., Birren B.W., Genereux D.P., Johnson J., Lindblad-Toh K., Karlsson E.K., Nweeia M.T., Johnson R.N., Zoonomia Consortium, Lewin H.A. Evolution of the ancestral mammalian karyotype and syntenic regions. *Proc. Natl. Acad. Sci. USA*, 2022, 119(40): e2209139119 (doi: 10.1073/pnas.2209139119).
122. Waters P.D., Patel H.R., Ruiz-Herrera A., Álvarez-González L., Lister N.C., Simakov O., Ezaz T., Kaur P., Frere C., Grützner F., Georges A., Graves J.A.M. Microchromosomes are building blocks of bird, reptile, and mammal chromosomes. *Proc. Natl. Acad. Sci. USA*, 2021, 118(45): e2112494118 (doi: 10.1073/pnas.2112494118).
123. Fritz A.J., Sehgal N., Pliss A., Xu J., Berezney R. Chromosome territories and the global regulation of the genome. *Genes Chromosomes Cancer,* 2019, 58(7): 407-426 (doi: 10.1002/gcc.22732).
124. Deakin J.E., Potter S., O'Neill R., Ruiz-Herrera A., Cioffi M.B., Eldridge M.D.B., Fukui K., Marshall Graves J.A., Griffin D., Grutzner F., Kratochvнl L., Miura I., Rovatsos M., Srikulnath K., Wapstra E., Ezaz T. Chromosomics: bridging the gap between genomes and chromosomes. *Genes (Basel),* 2019, 10(8): 627 (doi: 10.3390/genes10080627).
125. 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).
126. Kumar R.P., Senthilkumar R., Singh V., Mishra R.K. Repeat performance: how do genome packaging and regulation depend on simple sequence repeats? *Bioessays*, 2010, 32(2): 165-174

(doi: 10.1002/bies.200900111).
127. van Kruistum H., Nijland R., Reznick D.N., Groenen M.A.M., Megens H.J., Pollux B.J.A. Parallel genomic changes drive repeated evolution of placentas in live-bearing fish. *Mol. Biol. Evol.,* 2021, 38(6): 2627-2638 (doi: 10.1093/molbev/msab057).
128. Tang X., Li T., Liu S., Wisniewski J., Zheng Q., Rong Y., Lavis L.D., Wu C. Kinetic principles underlying pioneer function of GAGA transcription factor in live cells. *Nat. Struct. Mol. Biol.*, 2022, 29(7): 665-676 (doi: 10.1038/s41594-022-00800-z).
129. Foster C.S.P., Van Dyke J.U., Thompson M.B., Smith N.M.A., Simpfendorfer C.A., Murphy C.R., Whittington C.M. Different genes are recruited during convergent evolution of pregnancy and the placenta. *Mol. Biol. Evol.*, 2022, 39(4): msac077 (doi: 10.1093/molbev/msac077).
130. Glazko V.I., Andreichenko I.N., Kovalchuk S.N., Glazko T.T., Kosovsky G.Yu. Candidate genes for control of cattle milk production traits. *Russ. Agricult. Sci.*, 2016, 42: 458-464 (doi: 10.3103/S1068367416060082).