AGRICULTURAL BIOLOGY,
2021, V. 56, Iss. 2, pp. 292-303
(SEL'SKOKHOZYAISTVENNAYA BIOLOGIYA)

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

UDC 636.7:575.174.015.3

doi: 10.15389/agrobiology.2021.2.292eng
doi: 10.15389/agrobiology.2021.2.292rus

SOCIALIZATION AND GENETIC VARIABILITY AS A DRIVER OF DOMESTICATION (BY THE EXAMPLE OF DOG BREEDS)

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

[1]*Timiryazev Russian State Agrarian University—Moscow Agrarian Academy,* 49, ul. Timiryazevskaya, Moscow, 127550 Russia, e-mail vigvalery@gmail.com, aida-cat@yandex.ru, an.zhirkowa@yandex.ru, tglazko@rambler.ru (✉ corresponding author);
[2]*Afanas'ev Research Institute of Fur-Bearing Animal Breeding and Rabbit Breeding*, 6, ul. Trudovaya, pos. Rodniki, Ramenskii Region, Moscow Province, 140143 Russia, e-mail gkosovsky@mail.ru
ORCID:
Glazko V.I. orcid.org/0000-0002-8566-8717
Kosovskii G.Yu. orcid.org/0000-0003-3808-3086
Blokhina T.V. orcid.org/0000-0002-1126-2781
Zhirkova A.A. orcid.org/0000-0001-5985-9395
Glazko T.T. orcid.org/0000-0002-3879-6935
The authors declare no conflict of interests
Acknowledgements:
The authors are sincerely grateful to Larissa M. Fedorova PhD for her interest in the problem of domestication, fruitful discussion and helpful advice in preparing this article for publication.
Received December 2, 2020

Abstract

Social activity is the basis of interaction between different species in the process of common niche forming, including animal domestication. The increased social activity is universal characteristic of the "domestication syndrome" for different species (M.A. Zeder, 2017). It is assumed that some elements of this increase are due to a certain neotenization of a number of brain metabolic pathways (M. Somel et al., 2009). This is in good agreement with the data on the association of "domestication syndrome" with the slowing of neural crest cell proliferation (M.A. Zeder, 2015). The syndrome of hypersocialization (Williams-Behren Syndrome — WBS) in humans has been described, associated with hemideletion/hemiduplication of the 7q11.23 region, which includes 25-28 genes whose products are critical for the activity of various aspects of the central nervous system (A. Antonell et al., 2010). It was found that the complex of such genes is located on chromosome 6 in canids, and the domestic dog, considered in recent years as the main model object for studying the genetic mechanisms of domestication (E.A. Ostrander et al., 2019), differs from wolves in the presence of transposon insertions, increased methylation, and reduced gene expression in this region (B.M. von Holdt et al., 2017, 2018; D. Tandon et al., 2019). The aim of this work was to analyze such insertions in the region of the key gene for increased social activity in dogs *WBSCR17* (Cfa6.6 and Cfa6.7) in representatives of different breeds and interspecific hybrids with jackals, as well as finding out the presence of mobile genetic elements in these areas. The detected sequences have high homology to the non-autonomous dispersed nuclear element SINEC2A1_CF (94 % homology) and to two regions of endogenous retrovirus 3 the sequences of which are described in humans and cattle (approximately 80 % homology). Data were obtained on the increased variability of the presence and number of insertions into these areas in dogs of different breeds and hybrids, on the presence of homology sites to endogenous human and bovine retroviruses, as well as a short dispersed nuclear element, species-specific for domestic dogs, SINEC2A1_CF, carrying the hexanucleotide AATAAA which contributes to the completion of transcription. These finding suggest the involvement of retroviruses in the formation of an aggregate niche in the domestication process, which leads to increased variability that contributes to the selection of animals with hypersocialization.

Keywords: domestication, hypersocialization, Williams-Behren syndrome, retrotransposons, dog breeds and hybrids, aggregate niche

Domestication is mainly considered in relation to humans, artificial selection, and as the isolation of an object from the wild with complete control of its life cycle by humans. Also, domestication is an example of mutualism which appears not only between humans and domestic species of plants and animals. This type of relationship is evolutionarily much older and very widespread in nature,

especially among socialized species, for example, among leaf-cutting ants. Another striking example is the mutualism of fungi and plants, algae, bacteria. Mutualistic relationships are developing based on coevolutionary selection for mutant genotypes and lead to behavioral, physiological, morphological changes in both partners during the formation of an aggregate ecological niche [1, 2].

Predisposition to socialization, including reduced aggressiveness, social tolerance, and research tendencies, is a core driver for animal domestication. There is evidence that the socio-cultural factors of human-created niches make a relatively greater contribution to the intraspecific differentiation of cultivated plants and domesticated animals than environmental factors [3]. Interestingly, the supposed ancestral human species bonobos (*Pan paniscus*) significantly differ from chimpanzees in tolerance and altruistic behavior, namely in sharing food with representatives of other species [4].

Since environmental conditions are changing and biodiversity of domesticated animals are narrowing due to various reasons, the elucidation of the mechanisms of domestication becomes important for the management of genetic resources of agricultural species [5].

The oldest domesticated species is the dog (*Canis lupus familiaris*) whose domestication is associated with the hunter-gatherer civilization that preceded the agrarian civilization. It is generally accepted that the longest time of coevolution of humans and dogs in a common niche will allow the identification of key genomic targets of domestication. Therefore, in recent years, the domestic dog has become the leading research model for identifying genes and gene complexes associated with domestication [6, 7].

In humans, Williams-Beuren syndrome (WBS, hypersociability) has been described, which is caused by hemideletion or hemiduplication of 28 genes in the 7q11.23 region [8]. WBS is an autosomal dominant disorder caused by genomic rearrangements due to large region-specific changes and the presence of Alu transposons (short non-autonomous dispersed nuclear elements), which can lead to non-allelic homologous recombination in meiosis [9-11]. The incidence of WBS in the population is approximately $^{1}/_{10000}$ for WBS hemizygosity and $^{1}/20000$ for WBS hemoduplications. Deletion or duplication in WBSCR (Williams-Beuren syndrome critical region) leads to hemizygosity or hemoduplication of 25-28 genes, which explains the seen phenotypic manifestations [12]. Among others, the WBSCR region contains genes encoding transcription regulation factors, for example, GTF2I, GTF2IRD1, BAZ1B, and MLXIPL, as well as signaling molecules FZD9, TBL2, and LIMK1 [13]. It is assumed that the dosage of GTF2I can change the balance of excitation and inhibition [14], which is consistent with numerous evidences indicating an imbalance in the ratio between excitation and inhibition in cortical neurons as the main substrate for the development of the communication network [15, 16].

A block of genes corresponding to the human 7q11.23 region is found on the chromosome 6 of the domestic dog (CFA6). Comparative studies on the mechanisms that determine the increased tendency of the dog to initiate social contacts in comparison with socialized gray wolves have explained this behavior as a type of behavioral neoteny, the preservation of juvenile traits [17], which in itself is potentially the result of transcriptional neoteny in the brain [18]. It was found that structural variants of the WBS genes, particularly for *GTF2I* and its paralogs, underlie stereotyped hypersociability in domestic dogs and foxes [17, 19].

It is known that *WBSCR17* transcripts are predominantly expressed in the cerebellum, hippocampus, thalamus, and cerebral cortex of rats [20], and studies confirm the effect of WBSCR17 on cell morphology and traffic across

cell membranes [21]. WBSCR17 (in humans, GALNT17, N-acetylgalactosaminyl-transferase) is highly expressed in the cerebral cortex, participates in the function of lysosomes, cell adhesion, and the formation of the extracellular matrix [22].

Identification of the genetic basis of increased predisposition to social activity and learning ability in dogs is of particular importance in the selection of assistance dogs. For example, when training guide dogs for the blind, more than 60 % of individuals are rejected, mainly due to behavioral problems [23]. It was found that four insertions of retrotransposons (transposable elements, TEs) in genes *WBSCR17* (Cfa6.6 and Cfa6.7), *GTF2I* (Cfa6.66), and *POM121* (Cfa6.83) are associated with variability in the Canine Behavioral Assessment and Research Questionnaire test metrics (C-BARQ, http://www.cbarq.org). One of such insertions in *WBSCR17* turned out to be most closely associated with an increased predisposition to communicate with humans. The identified TEs with localization in introns of the *WBSCR17* and *GTF2I* genes and in the intergenic space are associated with increased methylation of the complex of genes of this chromosomal region, with their decreased expression and variability in the number of TE copies [24, 25].

To date, data have been accumulated indicating significant differences in the distribution of TEs in the domestic dog (*Canis lupus familiaris*), the gray wolf (*Canis lupus*), and the red wolf (*Cuon alpinus*), which are associated with domestication [26]. Thus, TE accounts for 41.75 % of the nucleotide sequences of the dog genome, which is higher than in the gray wolf (39.26 %) and in the red wolf (38.51 %). The most distinct TE components in these genomes are the long interspersed nuclear element LINE1 (L1) and microsatellites both making up 86.1 % of the differentiation between dog and gray wolf and 83.2 % of the differentiation between dog and red wolf. The content of canine-specific L1 Canis1 and L1 Cf in the dog's genome is significantly higher, particularly, almost twice as high as in gray and red wolves. It is assumed that the subfamilies L1 Canis1 and L1 Cf could have been accumulated in the dog's genome during domestication. In the dog genome, the canine-specific short interspersed element nuclear element SINEC Cf occurs in 27.3 million copies which is 1.16 times as much as in the genome of gray wolf and 1.23 times as much as in the genome of red wolf. It can be expected that the copy number of TEs also contributes to the differences in genome methylation profiles in dogs and wolves, as well as to the interbreed characteristics of dogs [27].

In several dog breeds, insertions of the retrotransposon were previously identified, without studies of its origin and homology with other transposons, in the 5 Mb region of chromosome 6 (in humans, deletions in this region are related to WBS syndrome manifestation), generating length variants of genes associated with behavioral responses [17]. However, it should be noted that the genetic basis of behavioral characteristics and molecular mechanisms of structural changes in genes associated with domesticated are poorly studied and require deeper insight.

In this work, as in the studies of von Holdt et al. [17], we used the primers [17] flanking the intron sequences of *WBSCR17*, the key gene determining the increased social activity of dogs and found variability in the presence and number of insertions in the Cfa6.6 and Cfa6.7 regions in representatives of different dog breeds and hybrids with a jackal. The obtained PCR products corresponded to the expected size of fragments in the presence or absence of insertions (555 or 357 bp for Cfa6.6 and 504 or 269 bp for Cfa6.7). Using bioinformatics methods, we isolated the corresponding region in a reference genome of Labrador Retriever and assessed the homology to retrotransposons in this region.

This work aimed to detect insertions in *WBSCR17* locus in dogs of various groups, i.e., the Vietnamese aboriginal breeds and specialized hunting greyhounds and the offspring of an interspecific hybrid of a dog and a jackal (Sulimov dogs —

shalaika breed group), using two pairs of primers to the regions of Cfa6.6 and Cfa6.7, and to analyze nucleotide sequences of the Cfa6.6 and Cfa6.7 regions of the *WBSCR17* locus in the Labrador Retriever reference genome for the presence of retrotransposons of various origins.

Materials and methods. Genetic material from 9 dogs of different breeds of greyhounds, 9 dogs of Vietnamese breeds, and 5 shalaikas (blood and tissue samples) were provided by the Department of Zoology, Timiryazev Russian State Agrarian University—Moscow Agricultural Academy. DNA extraction was performed using the Extran-1 and Extran-2 kits (Syntol LLC, Russia), following the manufacturer's instructions.

The employed primers were forward 5′-CCCCTTCAGCCAGCATATAA-3′, reverse 3′-TTCTCTGGGCTGTCTGGACT-5′ for Cfa6.6 and forward 5′-TGGA-GCCATGATTA-GGAAGG-3′, reverse 5′-TAAGGAAGGACCCCATTTCC-3′ for Cfa6.7 [17, 23]. In polymerase chain reaction (PCR), a PCR mixture (Syntol LLC, Russia) was used according to the manufacturer's recommendations. The PCR cycling parameters were i) initial denaturation at 94 °C for 2 min; ii) denaturation at 94 °C for 30 s, annealing at 58 °C for 30 s, elongation at 72 °C for 2 min (40 cycles); iii) final elongation at 72 °C for 10 min (a Tertsik amplifier, DNA-technology LLC, Russia). DNA amplification fragments were separated by horizontal electrophoresis in 1.5 % agarose gel stained with ethidium bromide. A DNA marker 100 bp + 1.5 Kb + 3 Kb (12 fragments from 100 to 3000 bp) (NPO SibEnzyme, Russia) was used to determine the size of the amplified fragments.

Using BLASTn software (BLAST: Basic Local Alignment Search Tool, https://blast.ncbi.nlm.nih.gov/Blast.cgi), nucleotide sequences of the *WBSCR17* gene fragments with insertions, a 555 bp fragment for Cfa6.6 and a 504 bp fragment for Cfa6.7, were extracted from a Labrador Retriever reference genome (ROS_Cfam_1.0 Genome Assembly, https://www.ncbi.nlm.nih.gov/assembly/GCF_014441545.1/ and UNSW_CanFamBas_1.0 Genome Assembly, https://www.ncbi.nlm.nih.gov/assembly/GCA_013276365.1) in searching for homology to flanking primers for Cfa6.6 and Cfa6.7 loci on chromosome 6 to determine in these loci the regions homologous to TEs of different origins. The search was performed using the Giri Repbase software (https://www.girinst.org/repbase/) and the CENSOR software (http://www.girinst.org/censor/index.php) [28].

Results. Electrophoretic separation of amplicons derived from the genomic DNA regions flanked by a pair of primers to Cfa6.6 detected fragments of 555-575 bp and 350-375 bp in the Vietnamese dogs and shalaikas (Fig. 1) along with longer fragments up to 850-875 bp in size which were more common in the greyhounds.

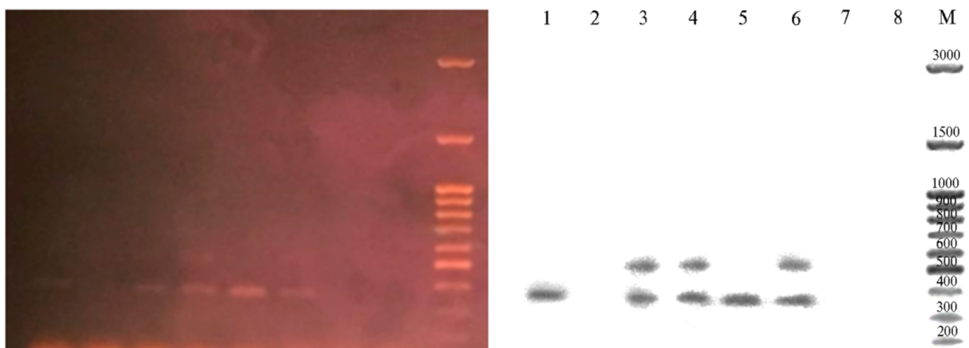


Fig. 1. Electrophoretic spectrum of PCR amplification products of genomic DNA regions flanked by primers to the Cfa6.6 locus in dogs: 1 — Vietnamese Phu Quoc, 2 — Vietnamese wolf-like dog, 3 — Vietnamese Hmong, 4-6 — shalaikas; M — molecular weight marker 100 bp + 1.5 Kb +3 Kb (NPO SibEnzyme, Russia). Fragments of 350-375 and 550-575 bp in length.

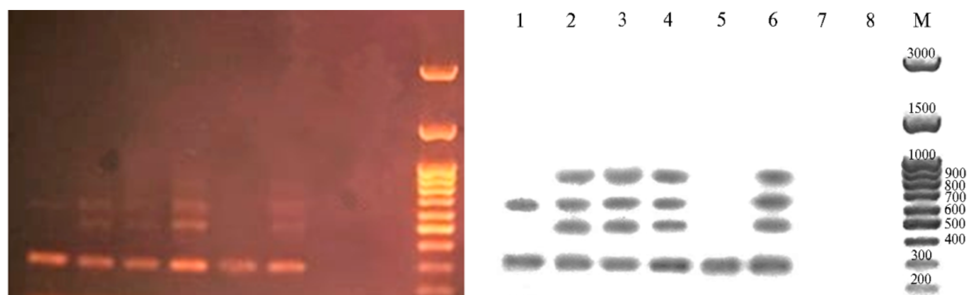


Рис. 2. Electrophoretic spectrum of PCR amplification products of genomic DNA regions flanked by primers to the Cfa6.7 locus in dogs: 1 — Vietnamese Phu Quoc, 2 — Vietnamese wolf-like dog, 3 — Vietnamese Hmong, 4-6 — shalaikas; M — molecular weight marker 100 bp + 1.5 Kb +3 Kb (NPO SibEnzyme, Russia). Fragments of 290-300, 500-510, 650-670, and 850-870 bp in length.

Amplification with a pair of primers to Cfa6.7 resulted in both fragments of the expected values (500-525 bp with insertions and 280-300 bp without insertions) and longer sequences which ranged from 600 to 800 bp in size (Fig. 2).

These data draw us to the conclusion that in the DNA of dogs with the domestication histories which extremely differ in the timing and the degree of influence of artificial selection, there is a pronounced polymorphism of the studied regions in the lengths of genomic fragments flanked by primers to the Cfa6.6 and Cfa6.7 loci. For Cfa6.6 and Cfa6.7, there are predicted regions with the insertions of retrotransposon fragments (555 and 504 bp) and without insertions (357 and 269 bp) and, in some cases, longer regions the presence and the size of which, as we assume, may indicate the appearance of unique insertions and duplications within the genome region under investigation. Our findings are consistent with the results of genotyping reported by other researchers for various dog breeds [23].

As noted above, TE insertions into the *WBSCR17* locus at the Cfa6.6 and Cfa6.7 regions of the domestic dog chromosome 6 of the significantly affect the increase in methylation and a decrease in the expression of a number of genes [23]. To find out which TEs are typical for this region, we compared the sequences of the 555 bp fragment amplified by primers to Cfa6.6, and the 504 bp fragment flanked by primers to Cfa6.7 in sequenced genome of a domestic dog (ROS_Cfam_1.0 Genome Assembly, https://www.ncbi.nlm.nih.gov/assembly/GCF_014441545.1/ and UNSW_CanFamBas_1.0 Genome Assembly, https://www.ncbi.nlm.nih.gov/assembly/GCA_013276365.1). As a result, the nucleotide sequences shown in Figures 3 and 4 were identified.

Then, using the Giri Repbase and the CENSOR software [28], a search was performed for homology to different transposons in Cfa6.6 and Cfa6.7 loci of chromosome 6 (see Fig. 3, 4). It turned out that Cfa6.6 from different sources (ROS_Cfam_1.0 and UNSW_CanFamBas_1.0) contains a sequence (nucleotides 183-370, Table 1) 94 % homologous to a 188 bp fragment of SINEC2A1_CF, a short non-autonomous interspersed nuclear element typically found in dogs.

1. Analysis of the presence absence of dispersed repeats in the 555 bp fragment of Cfa6.6 region flanked by the primers (these data were obtained with the use of the Giri Repbase, https://www.girinst.org/repbase/ and the CENSOR software http://www.girinst.org/censor/index.php)

/tmp/censor.57641.tmp/data.ori (SVG Plot; Alignments; Masked)

Name	From	To	Name	From	To	Class	Dir	Sim	Pos/Mm:Ts	Score
/tmp/censor.57641.tmp/data.ori	183	370	SINEC2A1_CF	1	192	NonLTR/SINE/SINE2	d	0.9841	2.0000	1512

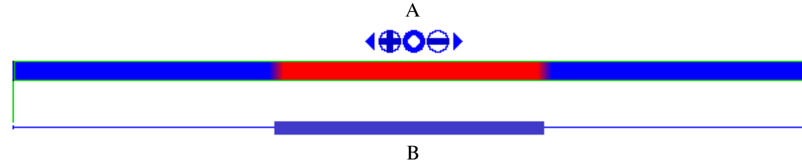
N o t e. Figures 3 and 5 show the primer sequences.

5′-(forward)**CCCCTTCAGCCAGCATATAAATTAAGGGAACTACACACTTAGTTTTGACCCAGATTTCTTTTTTAGGATGGCAGAAGCACACGCACGCACACACACACACACACACACACACACACACAA-AGTTTCTCTGATGGAAAACACAGAGACCTTCCAGATGAAATGCAGATATAAAAATTCTTTCTGGGATCCCTGGGTGGCGCAGCGGTTTGGCGCCTGCCTTTGGCCCAGGGCGCGATCCTGGAGAC-CCGGGATCGAGTCCCACGTCGGGCTCCCGGTGCATGGAGCCTGCTTCTCCCTCTGCCTGTGTCTCTGCCTCTCTCTCTCTCTCTCTGTGACTATCATAAATAAATAAAAAATTAAAAAAAAAAAATT-CTTTCTACTGAAATTTCCAGCCACCCCACCTTCAGACACCAAGACTAAAATGGCTTCTTGGCTAATGTGAGGTACCCACCTTCCTGAGTTATGTAAGCTAGAGTACGGGTAATCGTGAGGTTCTTCC-TTCTTAGTGTTCTGTGTTGGCATCGGTTGTCTACCACTAGTCCAGACAGCCCAGAGAA**(reverse)-3′

Fig. 3. CFA6.6 loci. The sequence identified in the whole genome reference sequence *Canis lupus familiaris* isolate:SID07034| breed: Labrador retriever (chromosome 6; whole genome sequence databases ROS_Cfam_1.0 https://www.ncbi.nlm.nih.gov/assembly/GCF_014441545.1/ and ROS_Cfam_1.0 https://www.ncbi.nlm.nih.gov/assembly/GCF_014441545.1/) **flanked by primers to the CFA6.6 region.** The regions of homology to primers are highlighted in yellow.

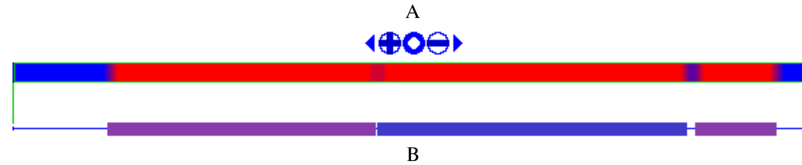
5′-(forward)**TGGAGCCATGATTAGGAAGGAGTCCTTGGAAACTCTGGCTTGTACCCTACATCTGAATTGTGGGGACAGTCTTGTGGGCTAACCCTTAACCTGTGGCATCTGATGCTATCCCTCAGGG-AGGTAATATCAGGATGGAGTTGAATTACAGGACCTCCAAGCTGGACACACAGAGAATCACTTGGTGTGGGAAACAACATTTGGTGGCCAGGAGCATCAGAAATGGGGTGTTCTCGGGATCCCTGG-GTGGCGCAGCGGTTTAGCGCCTGCCTTTGGCCCAGGGCGCGATCCTGGAGATCCGGGATCGAATCCCACGTCGGGCTCCCGGTGCATGGAGCCTGCTTCTCCCTCTGCCTGTGTCTCTGCCT-CTGTCTCTCTCTCTCTGTGACTATCATAAATAAATGAAAAAAAAAATTTAAAAAAAAAAAAAAAAAAAAAAAGAAATGGGGTGTTCTCTGTGAGTAATGAAGGAGACACATAGGAGAGAAATAGGAAGGAA-ATGGGGTCCTTCCTT**(reverse)-3′

Fig. 4. CFA6.7 loci. The sequence identified in the whole genome reference sequence *Canis lupus familiaris* isolate:SID07034| breed: Labrador retriever (chromosome 6; whole genome sequence databases ROS_Cfam_1.0 https://www.ncbi.nlm.nih.gov/assembly/GCF_014441545.1/ and ROS_Cfam_1.0 https://www.ncbi.nlm.nih.gov/assembly/GCF_014441545.1/) **flanked by primers to the CFA6.7 region.** The regions of homology to primers are highlighted in yellow.



5′-(forward)**CCCCTTCAGCCAGCATATAAATTAAGGGAACTACACACTTAGTTTGACCCAGATTTCTTTTTTAGGATGGCAGAAGCACACGCACGCACACACACACACACACACACACACACACACAA-**
AGTTTCTCTGATGGAAAACACAGAGACCTTCCAGATGAAATGCAGATATAAAAATTCTTTCTGGGATCCCTGGGTGGCGCAGCGGTTTGGCGCCTGCCTTTGGCCCAGGGCGCGATCCTGGAGAC-
CCGGGATCGAGTCCCACGTCGGGCTCCCGGTGCATGGAGCCTGCTTCTCCCTCTGCCTGTGTCTCTGCCTCTCTCTCTCTCTCTCTGTGACTATCATAAATAAATAAAAAATTAAAAAAAAAAAATT-
CTTTCTACTGAAATTTCCAGCCACCCCACCTTCAGACACCAAGACTAAAATGGCTTCTTGGCTAATGTGAGGTACCCACCTTCCTGAGTTATGTAAGCTAGAGTACGGGTAATCGTGAGGTTCTTCC-
TTCTTAGTGTTCTGTGTTGGCATCGGTTGTCTACCACTAGTCCAGACAGCCCAGAGAA(reverse)-3′

Fig. 5. The SINEC2A1_CF insertion position (in red) (A) **and the nucleotide sequence of a 555 bp fragment of the *WBSCR17* locus flanked by the primers to CFA6.6** (highlighted in yellow) **with a region of 94 % homology to the non-autonomous retrotransposon SINEC2A1_CF** (in red) (B). The AATAAA hexanucleotide (transcription termination signal) is marked in blue.



5′-(forward)**TGGAGCCATGATTAGGAAGGAGTCCTTGGAAACTCTGGCTTGTACCCTACATCTGAATTGTGGGGACAGTCTTGTGGGCTAACCCTTAACCTGTGGCATCTGATGCTATCCCTCAGGGAG-**
GTAATATCAGGATGGAGTTGAATTACAGGACCTCCAAGCTGGACACACAGAGAATCACTTGGTGTGGGAAACAACATTTGGTGGCCAGGAGCATCAGAAATGGGGTGTTCTCGGGATCCCTGGGT-
GGCGCAGCGGTTTAGCGCCTGCCTTTGGCCCAGGGCGCGATCCTGGAGATCCGGGATCGAATCCCACGTCGGGCTCCCGGTGCATGGAGCCTGCTTCTCCCTCTGCCTGTGTCTCTGCCTCTGT-
CTCTCTCTCTCTGTGACTATCATAAATAAATGAAAAAAAAATTTAAAAAAAAAAAAAAAAAAAAAAGAAATGGGGTGTTCTCTGTGAGTAATGAAGGAGACACATAGGAGAGAAATGGAAGGAAATGGGG-
TCCTTCCTT(reverse)-3′

Fig. 6. Positions of regions of homology to the region of homology to the endogenous human retrovirus EVR3 (in lilac in the lower part of the figure) **and to the non-autonomous retrotransposon SINEC2A1_CF** (in blue in the lower part of the figure) (A) **and the nucleotide sequence of a 509 bp fragment of the *WBSCR17* locus flanked by the primers** (highlighted in yellow) **to CFA6.7 with regions of homology to the endogenous retrovirus EVR3** (nucleotides 62-231, 435-486, in green) **and to the non-autonomous retrotransposon SINEC2A1_CF** (nucleotides 233-429, in red) (B). The AATAAA hexanucleotide (transcription termination signal) is marked in blue.

Figure 5 shows the positioning of the SINEC2A1_CF transposon (Fig. 5, A) in the revealed Cfa6.6 fragment and its nucleotide sequence (B, in red).

2. **Analysis of the presence or absence of dispersed repeats in the 509 bp fragment of Cfa6.7 region flanked by the primers** (these data were obtained with the use of the Giri Repbase, https://www.girinst.org/repbase/ and the CENSOR software http://www.girinst.org/censor/index.php)

/tmp/censor.58600.tmp/data.ori (SVG Plot; Alignments; Masked)

Name	From	To	Name	From	To	Class	Dir	Sim	Pos/Mm:Ts	Score
/tmp/censor.58600.tmp/data.ori	62	231	MER21C_BT	741	920	ERV/ERV3	d	0.7771	1.3333	750
/tmp/censor.58600.tmp/data.ori	233	429	SINEC2A1_CF	1	196	NonLTR/SINE/SINE2	d	0.9742	1.5000	1497
/tmp/censor.58600.tmp/data.ori	435	486	MER21C_BT	905	958	ERV/ERV3	d	0.8302	1.1429	312

Note. Figures 4 and 6 show the primer sequences.

A 504 bp fragment of the *WBSCR17* locus flanked by primers to CFA6.7 contains sequences homologous to the endogenous retrovirus EVR3 (nucleotides 62-231 and 435-486) and to a 197 bp fragment of SINEC2A1_CF retrotransposon (94 % homology, nucleotides 233-429) (Table 2, Fig. 6).

It should be noted that both fragments homologous to SINEC2A1_CF contain hexanucleotides AATAAA corresponding to the polyadenylation signal which serves as a key regulator of the end of the transcript. This hexamer or a similar sequence is very often found within 30 bp from the 3′UTR-ends [29]. Retrotransposons of the SINE class are found in all genomes in high copy number. Built into genes, they can disrupt expression, alter splicing, or stop transcription. The genome of the domestic dog harbors hundreds of thousands of such insertions, which can significantly affect the transcription of nearby genes or the genes at introns of which such insertions are located [29]. Therefore, a decrease in the expression of genes carrying insertions of these transposons may be due not only to the induced change in the methylation pattern, as suggested by some researchers [23], but also to the presence of the AATAAA hexanucleotide.

The data of Repbase indicate that EVR3 has sequences of homology to the mobile human genetic element MER21C and to a long terminal repeat of one of the endogenous bovine retroviruses [30].

Therefore, the endogenous human retrovirus EVR3 is directly involved in the genomic instability of *WBSCR17*, and its introduction into this locus is associated with a change in the transcription of a group of genes. This may confirm not only the idea of a long-term and directed selection of dogs for interest in contact with humans from the earliest stages of domestication, but also indicate the involvement of retroviruses in their horizontal transfer in and between various species (in particular, among bovine livestock) of a single aggregate ecological niche.

Our findings draw to the following conclusion. A prerequisite for domestication is the formation of a specific niche created by man. The key factor is selection of a plant or an animal to be domesticated. For animals, the key traits are increased social activity and the ability to adapt to conditions modified by humans. To ensure the success of such selection over a relatively short period of domestication, it is necessary to have an increased genetic variability which makes it possible to select the desired phenotypes [31]. Mobile genetic elements may be among tools providing such variability.

The totality of data indicates that the genome of the domestic dog, considered as a model of domestication, differs from the genomes of closely related

wild canines in a certain excess of retrotransposons [26] which make a significant contribution to genetic and genomic variability. The formation of a new niche presupposes a wide exchange between its "inhabitants", including exogenous retroviruses, the direct or altered descendants of which are autonomous retrotransposons which, in turn, involve non-autonomous retrotransposons in this variability. The increased variability of genome regions with a high density of factors which regulate transcription of genes involved in higher nervous activity (in particular, in socialization) with a direct participation of retrotransposons (see Fig. 5, 6) indicates that the generalization of genetic resources through the transfer of genetic material of initial pathogens between species can be the source of interspecific genetic relationships and intraspecific variability during the formation of a common niche. This may also explain the homology to the endogenous human virus ERV3 that we revealed in dogs in the key genome segment the variability of which is associated with hypersociability.

Thus, our findings indicate that polymorphism in the length of the *WBSCR17* gene associated with increased sociability is characteristic of dogs of different origins, as it was also reported by other researchers, and of interspecific hybrids such as shalaikas (Sulimov dogs), a jackal-dog hybrid. The detected polymorphism is due to insertions of transposed sequences. Analysis of such insertions in a reference genome of the domestic dog indicates that they contain nucleotide sequences of a non-autonomous transposon which occurs in the genome of dogs with high frequency, as well as a region of homology to the retrotransposon first described in humans and having homology to the cattle retrotransposon. It can be expected that it was the virome community that aggregated the members of one niche at the genomic level to generate variability enough to select individuals with an increased predisposition to interspecific interactions as a basis for the development of an agrarian civilization.

REFERENCES

1. Zeder M.A. Core questions in domestication research. *Proceedings of the National Academy of Sciences*, 2015, 112(11): 3191-3198 (doi: 10.1073/pnas.1501711112).
2. Zeder M.A. Domestication as a model system for the extended evolutionary synthesis. *Interface Focus*, 2017, 7(5): 20160133 (doi: 10.1098/rsfs.2016.0133).
3. Colino-Rabanal V.J., Rodríguez-Díaz R., Blanco-Villegas M.J., Peris SJ, Lizana M. Human and ecological determinants of the spatial structure of local breed diversity. *Sci Rep.*, 2018, 8(1): 6452 (doi: 10.1038/s41598-018-24641-3).
4. Sánchez-Villagra M.R., van Schaik C.P. Evaluating the self-domestication hypothesis of human evolution. *Evolutionary Anthropology*, 2019, 28(3): 133-143 (doi: 10.1002/evan.21777).
5. FAO. *The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture.* B.D. Scherf, D. Pilling (eds.). FAO Commission on Genetic Resources for Food and Agriculture Assessments, Rome, 2015. Available: http://www.fao.org/3/a-i4787e/index.html. No date.
6. Ostrander E.A., Wang G.D., Larson G., von Holdt B.M., Davis B.W., Jagannathan V., Hitte C., Wayne R.K., Zhang Y.P., Dog10K Consortium. Dog10K: an international sequencing effort to advance studies of canine domestication, phenotypes and health. *National Science Review*, 2019, 6(4): 810-824 (doi: 10.1093/nsr/nwz049).
7. Sykes N., Beirne P., Horowitz A., Jones I., Kalof L., Karlsson E., King T., Litwak H., McDonald R.A., Murphy L.J., Pemberton N., Promislow D., Rowan A., Stahl P.W., Tehrani J., Tourigny E., Wynne C.D.L., Strauss E., Larson G. Humanity's best friend: a dog-centric approach to addressing global challenges. *Animals (Basel)*, 2020, 10(3): 502 (doi: 10.3390/ani10030502).
8. Zanella M., Vitriolo A., Andirko A., Martins P.T., Sturm S., O'Rourke T., Laugsch M., Malerba N., Skaros A., Trattaro S., Germain P.L., Mihailovic M., Merla G., Rada-Iglesias A., Boeckx C., Testa G. Dosage analysis of the 7q11.23 Williams region identifies *BAZ1B* as a major human gene patterning the modern human face and underlying self-domestication. *Science Advances*, 2019, 5(12): eaaw7908 (doi: 10.1126/sciadv.aaw7908).
9. Antonell A., Del Campo M., Magano L.F., Kaufmann L., de la Iglesia J.M., Gallastegui F.,

Flores R., Schweigmann U., Fauth C., Kotzot D., Pérez-Jurado L.A. Partial 7q11.23 deletions further implicate *GTF2I* and *GTF2IRD1* as the main genes responsible for the Williams-Beuren syndrome neurocognitive profile. *Journal of Medical Genetics*, 2010, 47(5): 312-320 (doi: 10.1136/jmg.2009.071712).

10. Etokebe G.E., Axelsson S., Svaerd N.H., Storhaug K., Dembi Z. Detection of hemizygous chromosomal copy number variants in Williams-Beuren Syndrome (WBS) by duplex quantitative PCR array: an unusual type of WBS genetic defect. *International Journal of Biomedical Science*, 2008; 4(3): 161-170.
11. Ferrero G.B., Howald C., Micale L., Biamino E., Augello B., Fusco C., Turturo M.G., Forzano S., Reymond A., Merla G. An atypical 7q11.23 deletion in a normal IQ Williams-Beuren syndrome patient. *Eur. J. Hum. Genet.*, 2010, 18(1): 33-38 (doi: 10.1038/ejhg.2009.108).
12. López-Tobón A., Trattaro S., Testa G. The sociability spectrum: evidence from reciprocal genetic copy number variations. *Molecular Autism*, 2020, 11(1): 50 (doi: 10.1186/s13229-020-00347-0).
13. Li H.H., Roy M., Kuscuoglu U., Spencer C.M., Halm B., Harrison K.C., Bayle J.H., Splendore A., Ding F., Meltzer L.A., Wright E., Paylor R., Deisseroth K., Francke U. Induced chromosome deletions cause hypersociability and other features of Williams-Beuren syndrome in mice. *EMBO Mol. Med.*, 2009, 1(1): 50-65 (doi: 10.1002/emmm.200900003).
14. Makeyev A.V., Bayarsaihan D. Molecular basis of Williams-Beuren syndrome: TFII-I regulated targets involved in craniofacial development. *The Cleft Palate-Craniofacial Journal,* 2011, 48(1): 109-116 (doi: 10.1597/09-093).
15. Lopatina O.L., Komleva Y.K., Gorina Y.V., Olovyannikova R.Y., Trufanova L.V., Hashimoto T., Takahashi T., Kikuchi M., Minabe Y., Higashida H., Salmina A.B. Oxytocin and excitation/inhibition balance in social recognition. *Neuropeptides*, 2018, 72: 1-11 (doi: 10.1016/j.npep.2018.09.003).
16. Sohal V.S., Rubenstein J.L.R. Excitation-inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders. *Mol. Psychiatry*, 2019, 24(9): 1248-1257 (doi: 10.1038/s41380-019-0426-0).
17. von Holdt B.M., Shuldiner E., Koch I.J., Kartzinel R.Y., Hogan A., Brubaker L., Wanser S., Stahler D., Wynne C.D.L., Ostrander E.A., Sinsheimer J.S., Udell M.A.R. Structural variants in genes associated with human Williams-Beuren syndrome underlie stereotypical hypersociability in domestic dogs. *Science Advances*, 2017, 3(7): e1700398 (doi: 10.1126/sciadv.1700398).
18. Somel M., Franz H., Yan Z., Lorenc A., Guo S., Giger T., Kelso J., Nickel B., Dannemann M., Bahn S., Webster M.J., Weickert C.S., Lachmann M., Pääbo S., Khaitovich P. Transcriptional neoteny in the human brain. *Proceedings of the National Academy of Sciences*, 2009, 106(14): 5743-5748 (doi: 10.1073/pnas.0900544106).
19. Kukekova A.V., Johnson J.L., Xiang X., Feng S., Liu S., Rando H.M., Kharlamova A.V., Herbeck Y., Serdyukova N.A., Xiong Z., Beklemischeva V., Koepfli K.P., Gulevich R.G., Vladimirova A.V., Hekman J.P., Perelman P.L., Graphodatsky A.S., O'Brien S.J., Wang X., Clark A.G., Acland G.M., Trut L.N., Zhang G. Red fox genome assembly identifies genomic regions associated with tame and aggressive behaviours. *Nat. Ecol. Evol.*, 2018, 2(9): 1479-1491 (doi: 10.1038/s41559-018-0611-6).
20. Nakamura N., Toba S., Hirai M., Morishita S., Mikami T., Konishi M., Itoh N., Kurosaka A. Cloning and expression of a brain-specific putative UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase gene. *Biological and Pharmaceutical Bulletin*, 2005, 28(3): 429-433 (doi: 10.1248/bpb.28.429).
21. Nakayama Y., Nakamura N., Oki S., Wakabayashi M., Ishihama Y., Miyake A., Itoh N., Kurosaka A. A putative polypeptide N-acetylgalactosaminyltransferase/Williams-Beuren syndrome chromosome region 17 (WBSCR17) regulates lamellipodium formation and macropinocytosis. *Journal of Biological Chemistry*, 2012, 287(38): 32222-32235 (doi: 10.1074/jbc.M112.370932).
22. Merla G., Ucla C., Guipponi M., Reymond A. Identification of additional transcripts in the Williams-Beuren syndrome critical region. *Hum. Genet.*, 2002, 110(5): 429-438 (doi: 10.1007/s00439-002-0710-x).
23. Tandon D., Ressler K., Petticord D., Papa A., Jiranek J., Wilkinson R., Kartzinel R.Y., Ostrander E.A., Burney N., Borden C., Udell M.A.R., Von Holdt B.M. Homozygosity for mobile element insertions associated with WBSCR17 could predict success in assistance dog training programs. *Genes (Basel)*, 2019, 10(6): 439 (doi: 10.3390/genes10060439).
24. Kajikawa M., Okada N. LINEs mobilize SINEs in the eel through a shared 3′ sequence. *Cell*, 2002, 111: 433-444 (doi: 10.1016/S0092-8674(02)01041-3).
25. von Holdt B.M., Ji S.S., Aardema M.L., Stahler D.R., Udell M.A.R., Sinsheimer J.S. Activity of genes with functions in human Williams-Beuren syndrome is impacted by mobile element insertions in the gray wolf genome. *Genome Biology and Evolution*, 2018, 10(6): 1546-1553 (doi: 10.1093/gbe/evy112).
26. Wang G.D., Shao X.J., Bai B., Wang J., Wang X., Cao X., Liu Y.H., Wang X., Yin T.T., Zhang S.J., Lu Y., Wang Z., Wang L., Zhao W., Zhang B., Ruan J., Zhang Y.P. Structural

variation during dog domestication: insights from grey wolf and dhole genomes. *National Science Review*, 2019, 6(1): 110-122 (doi: 10.1093/nsr/nwy076).

27. Sundman A.S., Pértille F., Lehmann Coutinho L., Jazin E., Guerrero-Bosagna C., Jensen P. DNA methylation in canine brains is related to domestication and dog-breed formation. *PLoS ONE*, 2020, 15(10): e0240787 (doi: 10.1371/journal.pone.0240787).
28. Kohany O., Gentles A.J., Hankus L., Jurka J. Annotation, submission and screening of repetitive elements in Repbase: Repbase Submitter and Censor. *BMC Bioinformatics*, 2006, 7: 474 (doi: 10.1186/1471-2105-7-474).
29. Choi J.D., Del Pinto L.A., Sutter N.B. SINE retrotransposons import polyadenylation signals to 3′UTRs in dog (*Canis familiaris*). bioRxiv preprint, 2020 (doi: 10.1101/2020.11.30.405357). December 1, 2020.
30. Jurka J. Long terminal repeats from domestic cow. *Repbase Reports*, 2008, 8(8): 847-847.
31. Glazko V., Zybailov B., Glazko T. Asking the Right question about the genetic basis of domestication: what is the source of genetic diversity of domesticated species? *Advancements in Genetic Engineering*, 2015, 4(2): 125 (doi: 10.4172/2169-0111.1000125).