International Journal of Genetics and Genomics
2014; 2(4): 47-56
Published online August 10, 2014 (http://www.sciencepublishinggroup.com/j/ijgg)
doi: 10.11648/j.ijgg.20140204.11

Domestication and genome evolution

Glazko, Valeriy[1], Zybaylov, Boris[2, *], Glazko, Tatiana[1]

[1]Department of Zoo engineering, Russian State Agrarian University, Moscow agricultural Academy named after K.A. Timiryazev, Moscow, Russia
[2]Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, AR 72205

Email address:
blzybaylov@uams.edu (B. Zybaylov)

To cite this article:
Glazko, Valeriy, Zybaylov, Boris, Glazko, Tatiana. Domestication and Genome Evolution. *International Journal of Genetics and Genomics*. Vol. 2, No. 4, 2014, pp. 47-56. doi: 10.11648/j.ijgg.20140204.11

Abstract: Background. Pressure on modern agriculture to increase production is rising with the increase in human population. To meet this demand it is important to effectively manage domesticated species. However, genetic mechanisms and genomic targets of domestication are still poorly understood. It is well known that phenotypic variability in domesticated animals is higher compared to the variability in the closest wild relatives. Indeed, there are many breeds clearly distinguishable from each other by their morphological and physiological traits. In this report we review some of available literature and present original data to define genomic targets of domestication. Results. Using both publically available data and results of our own research we demonstrate the existence of a well-defined genomic signature (also called "sub-genome"), which consists of the molecular targets of artificial selection. The genetic signatures of domestication are revealed by comparison of different mammalian species and breeds. As a result, we found that a wide repertoire of genes is involved in the domestication process. The vast majority of these genes either plays a role in the neuroendocrine regulation, immune response, or encodes the milk proteins. Comparison of cattle genome to wild relatives reveals higher degree of polymorphism within retrotransposons, enzymes of the exogenous substrate metabolism, and in the genetic elements associated with the immune system. Conclusions. Our data for first time challenges the current explanation of phenotypic variation in domesticated species as a consequence of inbreeding and concomitant increase in homozygosity. Instead, we clearly show that there is no difference in the bulk genetic variability, and other explanation for difference in phenotypic variability is needed. We discover different targets of natural and artificial selection: in the case of domesticated species systems that are responsible for exogenous substrate metabolism are the targets, while in the case of wild species, genetic systems that are responsible for energy metabolism are targeted. We further speculate that the hyperactivity of mobile genetic elements – as evident from the higher polymorphisms within retro transposons – could be the source of increased genetic variability in domesticated species.

Keywords: Allele Distribution, Genetic Variability, Inverted Repeats, PCR Amplification Spectra, Mobile Genetic Elements

1. Introduction

Domesticated animals have enhanced phenotypic variability compared to the closely related wild animals. This increase in variability is evident from the existence of many different breeds, clearly distinguishable through their morphological and physiological traits. Furthermore, the inter-breed differences are quite often higher than the differences observed between closely related wild species. According to FAO data (180 countries, www.fao.org), only five of the key agricultural mammalian species (goats, sheep, cattle, horses and pigs) comprise 4,920 breeds, which are well distinguishable by their phenotypes, – the number that exceeds the diversity of all extant mammal species (4,500 including twin species).

A successful theory of domestication needs to answer the following five questions:

1. Why did domestication practices arise only in a few geographical areas (Mesopotamia, China, South, Central and East America, tropical Africa, Ethiopia, Seychelles and New Guinea)?
2. Why there were only 14 out of 148 large (heavier than 45kg) mammals domesticated?
3. Why did only 100 out of 200,000 wild plants yield useful domesticants [7]?
4. What is the underlying mechanism of domestication?

5. Why there are common domestication traits [4, 29] found among taxonomically remote animals [4] and plants [29] and yet, there are many cases when domesticants drastically differ from the genetically close wild species (in apparent violation of the Vavilov's law of the homologous series)?

Our starting assumption, which will lead to the answers of the above five questions is that all domesticated species must have some traits in common. Once we define these traits, once we understand laws of their heredity and variability, we will be able to easily determine what species can be domesticated, and also, we will be able to improve efficacy of artificial selection.

How did humans start to domesticate wild animals? Archeological evidence suggests that the majority of bones of domesticated animals belong to either females or young males. This makes the following domestication scenario plausible: hunters kept females and young males but hunted older males on surrounding territories, i.e. domestication appeared as a hunting strategy in hunters' civilization: keeping female herds to lure big males [37].

Pig domestication originated in south-east Anatolia in 10,500–10,000 B.C. The routes of geographical expansion and domestication of pigs are very similar to those of sheep but slower. European cattle was domesticated in the Euphrates Valley between 11,000 and 10,000 B.C. Pigs as well as sheep migrated relatively slowly to the Fertile Crescent (FC) region [37].

This domestication scheme is supported by genetic data. Recent studies showed that sheep and goat ancestors belong to species that existed in the FC (Ovis orientalis and Capra aegagrus respectively) [37]. These domesticated species have at least four genetically different domesticated lines or haplotypes (the goat has six). It is still not absolutely clear whether these lines correspond to a single or to several independent domestication events in space and time. For example, the high levels of intra-population diversity in Chinese sheep and the weak phylogeographic structuring indicate three geographically independent domestication events [5].

Cattle (*Bos taurus*) genetic data suggest the presence of five different haplotypes, with three, maybe four of them originated in the FC. Similarly, at least four of many pig lines appeared in the Middle East. Animal domestication in the FC occurred after prolonged interaction between humans and the ancestors of the main domesticated species [37].

At about this time the general hunting strategy, which was focusing on maximizing local availability of wild Bovidae, transformed into active management of herds of the four main species (11,000–10,000 B.C.). Even species such as gazelle, whose behavior is incompatible with domestication, were subject of taming in the southern and northern Levant where they were the largest wild Bovidae group, but the domestication results were poor.

Some scientists [7, 37] suggest that domestication was enabled by the phenotypic and genetic properties of species, but according to the others the main factor is the combination of climatic and soil properties, at least during the agrarian civilization expansion. The simple null model was proposed by Beck and Saber [3] and postulates that only climate and soil quality determine the four purposes of land usage: agriculture, settled animal husbandry, nomadism and hunting/gathering. This model correlates well with the real historical events (documented conflicts and population density changes), which took place in the Old World and Australia.

Thus, the success of the agrarian civilization expansion probably was dependent on the balance between global soil and climate quality gradients and the adaptive potential of people along with domesticated animals and plants, constituting local agro-ecosystems. Within interspecies communities, humans and domesticated species gene pools are in complicated relations, which are determined by artificial selection and the agrarian, ecological and landscape background.

The objective of our study is to explore connections between the genomic targets of variability in domestic animals and plants, and to suggest possible sources of this variability.

2. Results and Discussion

2.1. High Level of Genetic Variability in Domesticated Animals and Plants

Environment-induced adaptation almost always involves various phenotypic changes which have complex genetic determination. This complexity increases if the landscape level of ecological changes is added [6, 26].Genomic screening is the main trend in modern population genomics and may vary from several hundreds of markers used to true screening by full sequencing [26].

Cattle. In 2009 the full cattle genome sequence was obtained by the international Bovine HapMap Consortium. The bovine genome was found to contain at least 22,000 genes with a set of 14,345 orthologs common to 7 mammal species; 1217 of these orthologs are absent or have not yet been identified in nonplacental genomes. The obtained data revealed genomic regions with high density of segmental duplications rich in repeats and species-specific gene variants connected to the immune response. Genes involved in metabolism are highly conserved, however there are 5 metabolic genes that are deleted or considerably changed compared to their human orthologs [30, 31].

Also, a number of cattle immune genes differ from other mammals in that they are represented by a higher number of copies. These include beta-defensins, involved not only in antibacterial defense (unspecific immunity), but also in cellulose digestion in the rumen. Similar to rodents and dogs, cattle has about 1000 genes that are not found in the human genome. These genes have many variants in promoters and binding motifs of transcription factors, which adds to the unique characteristics of cattle and to the differences in mammalian development and physiology.

A high density of segmental duplications, retrotransposon and retroviral long terminal repeats were found in chromosomal regions that have undergone rearrangements during the last 80 million years of *Bos taurus* karyotype formation. A conclusion is made that such repeat elements and segmental duplications directly provide chromosomal rearrangements connected to the species origin in many mammalian lineages. There is a high level of genetic variability in all cattle breeds. It is even higher than variability among dogs or humans. The maximum genetic diversity is found in zebu with a single nucleotide polymorphism (SNP) at every 285 base pairs (bp).

Genetic evidence shows that after domestication cattle breeds underwent selection bottlenecks with a limited number of progenitors and/or intensive selection on production traits. Taking into account comparisons of variability distribution based on thousands of SNPs, the international consortium revealed that many genomic regions differ between beef and dairy breeds and most of them contain genes responsible for quantitative variability of beef and dairy productivity and are located in cattle chromosomes 2, 6 and 14 [22, 31].

For example, we compared the ratio of synonymous/nonsynonimous (dS and dN) substitutions in kappa casein exon IV determining the size of micelles in milk (chromosome 6). The Nei-Gojobori method applied to different parts of the kappa casein gene revealed that there is a positive selection supporting high variability of amino acid substitutions in the kappa casein C-terminal domain only in Bovinae (Table 1, [21]).

Table 1. *Average values of interspecies differences based on the ratio of nonsynonymous/synonymous substitutions (dN/dS) scored in different parts of the kappa casein gene (Nei–Gojobori method).*

Polymorphism	Exon IV		RKS Protein		C domain	
Species	dN	dS	dN	dS	dN	dS
Bovinae	0.045	0.036	0.020	0.022	0.109	0.103
Caprinae	0.018	0.022	0.010	0.024	0.025	0.048
Odocoileinae	0.018	0.030	0.018	0.027	0.031	0.024
Cervinae	0.014	0.017	0.016	0.040	0.017	0.000

It is interesting that the highest rate of evolution of the kappa casein gene among Bovinae species is observed in the C-terminal domain, since this domain contains all the sites of posttranslational casein modifications (phosphorylation and glycosylation), which affect the physical properties (size, solubility) and reactivity of casein micelles.

The total number of threonine and serine residues (sites of the phosphorylation and glycosylation) remains the same only in the kappa casein of the Bovinae family, while their positions are changed. In other families both the number and positions of these amino acids remain the same [21]. It is known that different glycosylation distribution in the kappa casein C-domain correlate with different levels of inhibition of the gastrointestinal pathogen Helicobacter priori [20, 28]. Thus, it could be expected that the high evolutionary rate of the amino acid sequence of this kappa casein domain may be related to the adaptation to various pathogens of closely related Bovinae species. This phenomenon could be a result of nutritional differences introduced after the divergence of Bovinae species due to domestication which imposed the need for adaptation to various gastrointestinal pathogens.

Hence, the selection of allele variants among milk protein genes in particular may have resulted not from the instinctive work of a breeder trying to obtain high milk productivity, but rather as a by-product of natural selection directed towards higher resistance to environmental pathogens that domesticated animals faced along with human contact and colonization of new niches.

Artiodactyla and Perissodactyla. In addition, we carried out a comparative analysis of the polymorphism of 30 loci of various functional protein groups in the gene pools of 12 domesticated and close wild species of two animal orders: Artiodactyla and Perissodactyla [10, 12, 13, 21]. The study included wild "zoo" species bred in the biosphere reserve Askaniya-Nova, Ukraine, and some cattle and horse breeds bred in different Russian and Ukrainian farms (26 breeds and intra-breed groups).

The average polymorphism level for the studied loci was slightly higher among domesticated species compared to that among wild forms. In domestic Bovidae the share of polymorphic loci was found to be lower for the intracellular energy metabolism enzymes but higher for exogenous substrate metabolism enzymes and transport proteins, as compared to wild relatives (Table 2) [21].

Table 2. *Polymorphic loci share of various functional groups of genetic and biochemical systems in wild and domesticated mammal species.*

Species	Protein functional groups		
	Ia	IIb	IIIc
Wild	0.629	0.193	0.178
Domesitcated	0.179	0.464	0.357

a intracellular energy metabolism enzymes; b exogenous substrate metabolism enzymes;c transport proteins.

The differences in the total polymorphism of different functional groups observed between wild and domestic mammal species are in agreement with the suggested connection between species origin and reorganization of energy supplying mechanisms and with the fact that artificial selection does not lead to the emergence of new species, except for artificial interspecies hybridization. It could be speculated that natural selection is directed towards the emergence of new species (i.e. ability to occupy new habitats and/or niches) and hence, favors the polymorphism of intracellular energy metabolism enzymes (glycolysis, citric acid cycle). Artificial selection, on the other hand, aims at forms adapted to an unsteady flow of exogenous substrates (i.e. ability to eat whatever is being fed) and hence, favors the polymorphism of exogenous substrate metabolism enzymes and transport proteins.

Soybean. Similarly to animals, in domesticated soybean cultivars polymorphism seems to involve mainly enzymes participating in pathways other than the glucose metabolism. This suggestion is based on our results of a comparative

analysis of the polymorphic loci share in 18 soybean (Glycine max) cultivars and 3 populations of the putative progenitor wild soybean (Glycine soja, previously G. ussuriensis) collected in various regions of the Far East [11, 25]. Seeds were kindly provided by V.V. Sherepitko, DSc (Ukraine) and I.V. Seferova, PhD (VIGG, Russia).

We found 21 polymorphic loci (out of 42) for all analyzed plants. Genetically and biochemically they fell into two groups: enzymes participating in intracellular processes of ATP accumulation, i.e. glucose metabolism (G) – glycolysis, citric acid cycle; and all others (NG). In total 21 G and 21 NG enzymes were analyzed; 7 polymorphic loci of wild populations included 1 NG (ESTD-1) and 6 G loci. Out of 19 loci of soybean cultivars we found polymorphism in 11 G and 8 NG.

In summary, in wild soybean polymorphism was observed predominately in G loci (86% of all polymorphic loci). This percent was lower (58%) in the studied domesticants, where the share of NG polymorphic loci was 3 times higher (42%) than in the wild relatives.

Moreover, the range of genetic variability, i.e. polymorphic loci share P, was found to be higher in G. max (45%) than in G. soja (17%), indicating that the domesticated species is more polymorphic than its close wild relative.

These results support the idea of a "subgenome" containing loci whose products participate in the regulation of interactions between the intra- and exracellular medium (enzymes of the exogenous substrate metabolism, transport proteins). The higher variability of these loci acts as a necessary condition for domestication both for plants and animals.

2.2. Evidence for the Existence of Domestication Genomic Signature

Short DNA fragments flanked by inverted repeats predominate in the amplification spectra of domesticated species. Interspecies differences between domesticated and closely related wild animals were also revealed based on amplification spectra (RAPD - PCR, ISSR - PCR) obtained by using two decanucleotide primers: UBC - 85 (5'-GTGCTCGTGC-3') and UBC- 126 (5'- CTTTCGTGCT-3') [16, 21]. The amplification spectra of domesticated species consisted predominantly of short DNA fragments flanked by inverted repeats of these primers (Table 3).

Table 3. *Comparative analysis of frequencies of occurrence of amplicons with different length in amplification spectra (RAPD-PCR) obtained from domesticated and wild Ungulata species by using decanucleotide primers UBC-85 and UBC-126.*

Species	Amplicon length		
	Short (0.4–1.0 kb, %)	Average (1.1–1.9 kb, %)	Long (2.0–2.5 kb, %)
Domesticated	36.3	50.9	12.8
Wild	29.8	49.0	21.2

ISSR-PCR analysis was carried out to estimate the similarities and differences in the distribution of fragments of different length (amplicons) for the studied mammal species (Glazko, 2004). Three di- and 12 trinucleotide primers were used: (AGC)6T, (TGC)6A, (AGC)6G, (ACC)6G, (GCT)6A, (GAG)6C, (TCG)6G, (CTC)6A, (CAC)7A, (CTC)6C, (GTG)7C, (CAC)7T, and 310 amplicons were identified. It was obtained that short issr-pcr amplicons are found significantly ($P < 0.05$, t-test) more often in domesticated species than in close wild relatives (Table 4).

Table 4. *Amplicons of different length (as percent in the total amplicon spectra) obtained from wild and domesticated mammal species by using di- and trinucleotide microsatellite loci fragments as primers.*

Amplicon length (kb)	Domesticated species (%)	Wild species (%)
1.1–0.4	50	39
1.8–1.1	38	44
2.5–1.8	12	17

There is similarity in the dendrograms based on genetic distances estimated by both types of markers (proteins, RAPD-PCR, ISSR-PCR).

Taken together, obtained data indicate that domesticated species differ from closely related wild animals mainly in the polymorphism of protein-coding genes whose products participate in the regulation of the interactions between the intracelluar and extracelluar medium, and in the higher frequency of occurrence of short DNA fragments flanked by inverted repeats.

The analysis of nucleotide sequences of DNA fragments, flanked by the inverted repeat (AG)9C, which presence distinguished the ISSR-PCR spectra of the house breed from those in genomes of cattle and sheep breeds with the same flanking was carried out. It was revealed that the analyzed fragment appeared as a result of recombination between ancient mobile elements (DNA transposon of fish, endogenous mammalian retrovirus ERV3) and sequence which was specific to horse endogenous retrovirus ERV1. The obtained data point to the direct participation of mobile genetic elements in differentiation of gene pools not only between species, but also, apparently, between breeds of farm animal species [17, 18].

Interestingly, the highest number of amplicons (32) was obtained with (CTC)6A – 33, (CTC)6C – 33, (GAG)6C, i.e. with primers belonging to purine-pyrimidine tracks. These tracks take part in secondary structures and perhaps are involved in gene expression regulation mechanisms [21]. A relatively large number of amplicons was obtained with 2 other motifs: (ACC)6G (36) and (AGC)6G (31). At the same time according to data of the international Bovine HapMap consortium [30, 31] the frequency of occurrence of microsatellites with a core motif AGC in Artiodactyla (cattle, sheep, pigs) is 90 and 142 times higher than in dogs or humans, respectively. Moreover, in 39% of the cases this microsatellite goes together with retrotransposon Bov-A2 SINE, which is evolutionarily young and specific for the cattle genome.

Thus, the cattle genome screening reveals higher polymorphism of genetic elements connected with the immune system, retrotransposons and enzymes of the

exogenous substrate metabolism compared to wild species.

2.3. Domestication Signature in the Genomes of Domesticated Animals

Using F-statistic (Fst) analysis for SNP in different genomic regions, Barendse et al. [2] attempted to identify selection targets in the cattle genome. The region with maximum Fst value in cattle chromosome 2 contains a number of genes connected with human selection pressure [2]. R3HDM1 and ZRANB3 genes are related with these cattle SNPs. Most breeds are homozygotes but Hereford, Santa Gertrudis and Belmont Red breeds differ by moderate frequency of occurrence of an alternative allele. This region is well known to be associated with positive human selection owing to the lactase gene (LCT) localization and human adaptation to milk consumption during adulthood. It is unlikely, however, that cattle have been selected by lactase activity in adulthood, since all animals are weaned at the same age. Recently it has been shown [2] that the R3HDM1 locus is under positive selection in the European human population and does not diverge from the LCT gene by the "hitchhiking" principle. These results may possibly become a starting point for the discovery of traits associated with selection among Homo sapiens.

If a mutation in a coding region is not significantly adaptive, the dN/dS ratio should be approximately equal among closely related species. When a neutral variability takes place, the dN/dS ratio in interspecies divergence should be close to the dN/dS ratio in intraspecies polymorphism. Checking this hypothesis is especially interesting for closely related Bovinae species because domestication led to very quick phenotypic differentiation as a result of intensive artificial selection.

The comparison of genes involved in the dairy productivity of cattle and close wild species revealed that domestication led to dN/dS increasing in European cattle breeds (*Bos taurus*) and southeastern gayal (*Bos frontalis*) [23, 24]. The authors conclude that the selection result depends on effective population size and on the selection coefficient. Generally, during domestication the selection pressure on traits important to adaptation for wild species decreases. This may have led to the rapid evolution of domesticated species, especially of B. taurus and B. frontalis, which have the highest dN/dS ratio among Bovinae [23, 24]. Surprisingly, significant differences in supposedly neutral substitution levels between synonymous and noncoding regions in the cattle genome were found: they were 30% higher in synonymous sites. This may be associated in part with an excess of highly variable CpG dinucleotides in synonymous sites, which in turn will affect the time estimations of species divergence based on molecular data [23, 24].

An important contribution to the idea about the specificity of domestication "signatures" is made by data indicating absence of linear relations between the variability of some genes and selection pressure features. A number of coincidences of polymorphism of noncoding sites in Hereford-yaks, Herefords-Holsteins and yaks-bisons imply that they may be descendants of different lines formed in a common ancestor species [23, 24].

2.4. The Hypothalamo-Pituitary-Adrenal System as the Main Initial Domestication Target

A unique experiment of wild fox domestication was begun by D.K. Belyaev and continued by L. Trut [32, 36]. They identified the hypothalamo-pituitary-adrenal system as the main initial domestication target. Selection for behavior was shown to weaken the activity of this system both on the phenotypic level and on the level of gene expression (levels of corticotrophin releasing factor, propiomelanocortin, and glucocorticoid receptor).

SNP genomic screening of specialized dairy breeds in France indicated that many genes involved in the neuroendocrine system formation respond to selection of dairy productivity traits [9]. However, the gene networks including different genes of the somatotropic axis known as dairy productivity selection targets did not coincide for the three investigated breeds [9].

2.5. Allele Distribution of Structural Genes as a Possible Additional Breed Characteristic

To evaluate the allele distribution of structural genes involved in desirable genotypes of 5 autochtonous cattle breeds in Ukraine and Russia, we [21] analyzed the following structural genes by RFLP-PCR: coding dairy proteins (kappa casein – CSN3 and beta lactoglobulin –BLG), myostatin participating in control of muscle mass growth rate (MSTN), hormone of lipid metabolism (LEP), growth hormone (GH) and transcription regulation factor – a locus of somatotropic hormone (PIT-1).

CNS3 amplification product included part of exon 4 and intron 4; Hind III restriction gave two allele variants A and B. The presence of B essentially increases the quality of firm cheeses, while A is associated with high total yield and dominates in dairy cattle breeds.

BLG locus was 247 bp long and included part of exon 4 and intron 4. Allele variant BLG A is associated with high milk yield.

LEP amplification product was 1830 bp long and included part of exon 2, the whole intron 2 and part of exon 3. Sau3A enzyme restriction revealed 3 allele variants (A, B and C). LEP AA genotype (it has 2 digestion sites for Sau3A) is associated with decreased fodder efficiency compared to BB (with an additional digestion site); AC genotype is associated with high butter-fat and protein content in milk and also with the best lactation dynamics [38].

In the GH locus a 223 bp fragment of exon 5 was amplified. AluI digestion revealed 2 variants: L (leucine in site 127) and V (valine in the same position). The milk of cows with LL genotype contains more fat and protein but has a bit lower total yield than that of VV genotype cows [38]. Amplification of a fragment of intron 6 of Pit-1 gene (1355 bp long) and further restriction by Hinf I yielded two

allele variants (A and B). A is associated with higher protein yield but lower fat yield [38].

Table 5. *Genotype and allele distribution in loci involved in diary and beef yield properties in Grey Ukrainian cattle reproducing in different ecological and geographical conditions.*

Genes	Grey Ukrainian, Kherson, Ukraine			Grey Ukrainian, Cherga, Russia		
	Genotypes	Number of animals	Allele frequencies	Genotypes	Number of animals	Allele frequencies
CSN3	AA	5	A-0.692	AA	4	A-0.612
	AB	8	B-0.307	AB	9	B-0.398
	BB	0		BB	1	
BLG	AA	4	A-0.5	AA	5	A-0.400
	AB	7	B-0.5	AB	8	B-0.600
	BB	4		BB	6	
GH	LL	8	L-0.733	LL	10	L-0.674
	LV	6	V-0.267	LV	5	V-0.336
	VV	1		VV	3	
Pit-I	AA	7	A-0.714	AA	10	A-0.702
	AB	1	B-0.268	AB	4	B-0.208
	BB	6		BB	4	
LEP	AA	7	A-0.888	AA	8	A-0.711
	AB	2	B-0.111	AB	1	B-0.299
	BB	0		BB	1	
MSTN	nt812(del11)/N	0	–	nt812(del11)/N	0	–

To estimate if the distribution of allele variants could be considered as an additional breed characteristic, two Grey Ukrainian breed groups reared in different ecological and geographical conditions for generations (Kherson region, "Askaniya-Nova" and "Cherga", Altai region) were compared [21]. In these two groups the allele frequencies coincided almost for all the investigated genes, suggesting that allele occurrence of these genes does not depend on the ecological and geographical conditions and may serve as an additional breed characteristic (Table 5) [21].

The allele frequency distribution corresponded to different production trends: dairy breeds had higher frequency of alleles associated with high total yield as compared to beef or multiple purpose breeds. On the individual level, however, the occurrence of such "dairy" alleles was not associated with the dairy yield of individual cows. Since, the 5 studied loci are located on different chromosomes (PIT 1 is on chr. 1; LEP on chr. 4; CSN3 on 6; BLG on 11; and GH on 19), the milk productivity of different cows with equally high yields might be decisively determined by different loci, including those analyzed by us.

This suggestion is in agreement with the idea about the gene networks among three dairy breeds in France [9].

In short, the domestication "signatures" revealed by comparisons between different mammal species and breeds from different production trends demonstrate that a wide repertoire of genes is involved in the domestication process. Most of them play a role in the neuroendocrine regulation or immune response, or encode milk proteins. The same spectrum of phenotypic traits, however, could be caused by the genotypes of different genes involved in genes networks. Hence, the main factor which determines the possibility for domestication must be the high genetic variability of genetic systems, related with the neuroendocrine and immune functions.

One intriguing hypothesis, which could potentially explain the phenotypic variability among domesticated species is that this increased variability is the result of interactions with various pathogens, as evident from is the observed density of retrotransposons, retroviral LTRs in regions of segmental duplication in the cattle genome, and from the higher number of copies of genes related to the immune system [30, 31].

2.6. Possible Sources of Genetic Variability in Domesticated Species

To identify possible sources of increased genetic variability and to determine if there are elements in the cattle genome that are connected to the fodder resources, we carried out an analysis of polymorphism of DNA fragments flanked by LTR of the soybean transposon SIRE-1 (GenBank: AF053008) in the Lebedinskaya cattle breed [21]. The amplification spectra contained 14 fragments; 11 did not have individual variability and were also observed in the spectra of other breeds.

BLASTn search in GenBank found fragments with partial homology to SIRE-1 (11–23 nucleotides) in 20 of 29 cattle autosome sequences and in the X and Y chromosomes. The cattle EST database contains fragments homological to mRNAs of genes participating in thyroxin folding (GenBank: SH3BGRL), coding transcription factors (GenBank: LOC782608, LOC781021, FOXJ1), synthesis of telomere-associated proteins (tankyrase, TNKS), plasma- and nuclear-membrane-associated proteins (laminin alpha 1, LAMA1, attractin-like protein 1 – ATRNL1; spectrin containing protein of nuclear membrane 1 – SYNE1), proteins involved in the defense from infectious diseases (T-cell receptor alpha, TCRA; one of the early inflammatory proteins – TLR3). Short homological sites are also revealed in a number of miRNAs: bta-mir-2303 (chromosome 12); bta-mir-2356 (chromosome 2); bta-mir-2480 (chromosome 9); bta-mir-2441 (chromosome 5). It is known that miRNAs are widely represented in various genomes, participate in gene expression regulation and are likely to be associated

with virus infection [19]. A homology search in the genomes of other taxa has shown many homologous sites in the human genome, in the mRNAs of bromodomain chromatin remodeling factor (PBRM1), intercellular skin protein filaggrin (FLG) and membrane bound receptor of neurotropic tyrosine kinase (NTRK2) participating in the processes of cell division and differentiation regulation. A homologous region was also found in hens (chromosome Z); and in several prokaryote genomes.

We note that the search for homology using short sequences is questionable at best, and the results are spurious, with necessarily high false positive rate. However, we feel that the apparent homology to short miRNAs is interesting in the context of hypothesis that the interaction with pathogens is one of the important conditions of domestication.

DNA fragments flanked by inverted repeats of these sequences vary greatly between rice cultivars [14], wheat cultivars and even between plants of a common cultivar origin [15].

BLASTn search in GenBank found a large amount of sequences with partial homology to these sequences in mammalian species which were usually localized in the P450 polygenic family, immune system genes and transcription factors. Homologous sites have wider taxonomical representation than those with flanking soybean retrotransposons and are also found in prokaryotes.

A pattern of rapid evolution of target sites under artificial selection in the domesticated species can be seen in cattle and horse genome sequences. The data are based on transposable elements involved in segmental duplication evolution, their localization in actively transcribed genome regions, correlations between the number of integrated provirus-like transposable elements and resistance of, for example, rice cultivars, to retrovirus infections [30, 31, 35]. This pattern is consistent with the recent data that demonstrate a link between the epigenetic control of transposable elements in species/populations and their genome evolution, high speed of their transpositions within their host and between different genomes [27, 33, 34].

It is possible that natural habitat expansion after human migration routes may have increased the number of contacts of domesticated species with new retroviruses and thus favored the integration of new transposable elements. Such sequences remained conservative because of natural selection (they prevented reinfection), but increased the genetic variability in their integration sites (insertion mutagenesis, recombination acts), which might have generated new mutations essential for artificial selection.

The participation of transposable elements in the divergence of domesticated and close wild species could explain some empirical data, e.g., the relatively high evolutionary rate of several genetic elements in genes of domesticated species and our data [21] about higher frequency of occurrence of short DNA fragments flanked by inverted repeats in the genomes of domesticated species compared to close wild forms.

This assumption is further supported by the finding that in the cattle and human genomes there are regions with homology to retrotransposon fragments typical for forage plants. Moreover, they are located in sites associated with genes involved in the immune response and signal transduction (structure elements of the plasma membrane, nuclear membrane, chromatin, transcription factors).

The hypothesis is also supported by some recently revealed mechanisms of interaction between viruses and metaphase chromosomes: special virus proteins interact directly with chromatin proteins to allow viruses to preserve themselves during cell division [8].

3. Conclusions

The current evidence shows that domesticated species differ from closely related wild animals in several specific targets of the selection pressure, related to the necessary interaction with humans (neuroendocrine factor), adaptation to a wide spectrum of food sources, pathogens (including “crowd diseases”), ecological and geographical factors.

The results of our analyses support the idea of a "subgenome" containing loci whose products participate in the regulation of interactions between the intra- and exracellular medium (enzymes of the exogenous substrate metabolism, transport proteins). The higher variability of these loci acts as a necessary condition for domestication both for plants and animals. The cattle genome screening reveals higher polymorphism of genetic elements connected with the immune system, retrotransposons and enzymes of the exogenous substrate metabolism compared to wild species.

What is the source of this selective genetic variability, which provides adaptive potential for domesticated species? One possible answer is the relatively higher pathogen repertoire faced during colonization. Exposure to pathogens could lead to increase integration of retrotransposons into domesticated genomes, which in turn are activated during inbreeding, leading to the consequent increase in phenotypic variation. This hypothesis, of course, would further require a convincing demonstration of the transfer of genetic material between pathogenic microbiota and domesticated species. We believe that our finding related to polymorphisms within retrotransposons is one of the first steps in this direction.

To summarize, the domestication “signatures” revealed by comparisons between different mammalian species and breeds from different production trends demonstrate that a wide repertoire of genes is involved in the domestication process. Most of them play a role in the neuroendocrine regulation or immune response, or encode milk proteins. The same spectrum of phenotypic traits, however, could be caused by the genotypes of different genes involved in genes networks. Hence, the main factor which determines the possibility for domestication must be the high genetic variability of genetic systems, related with the neuroendocrine and immune functions.

4. Methods

Domestic and wild mammalian species analyzed. We studied species belonging to the following two genera – Artiodactyla and Perissodactyla, that lived in the Biosphere Reserve "Askaniya-Nova" (Ukraine):

Artiodactyla, family Antilopinae: *Saiga tatarica (saiga), Taurotragus oryx (eland), Boselaphus tragocamelus (bluebuck), Connochaetes gnu (gnu); family Capra: domesticus goat (Orenburg's breed); family Ovis: Ovis domesticus (Carphatian breed), Ovis canadensis (snow buck); family Bovinae: Bos taurus (cattle); Bos taurus macroceros (watussi), Bison bison (bison), Bison bonasus (european bison), Bibos gaurus frontalis (mithan).*

Perissodactyla. Family Equidae: Eguus Przewalsky's (Przewalsky's Horse), Eguus caballus (Arabian breed, Orlov's trotter), Eguus asinus (donkey), Equus hemionus hemionus (kulan), Equus burchelli chapmanix (zebra Chapman's), Equus burchelli granti (zebra Grant's), Equus (Dolichohippus) grevyi (zebra Grevy's)

Additionally, we investigated cattle, sheep and horse breeds reproduced in different farms across farms in Russia and Ukraine (26 breeds and intrabreed groups).

Plant species analyzed. The population and genetic estimates of differentiation of 18 soy cultivars (*Glycine max*) and 3 populations of wild *Ussurian soy* collected in various regions of the Far East (*Soja Glycine ussuriensis* Moench, a suggested ancestor) were added to the analysis. Seeds were pleasantly provided by Dr. Sherepitko (Ukraine) and Dr. Seferova (VIGG, Russia).

Protein polymorphism analysis in blood plasma and cells of animals. The proteins in question were separated using acrylamide and, sometimes, starch gels for the analysis, suing standard protocols. The following proteins, representative of different genetic and biochemical systems were analyzed: proteins of blood plasma: albumin, ceruloplasmin, transferrin, vitamin D receptor, alpha1-1 beta-glycoprotein (A1B), esteraseamylase-1, alkaline phosphatase; enzymes of blood cells: sorbitol dehydrogenase, lactate dehydrogenase, malat dehydrogenase, malic-enzyme, 6-phosphogluconat dehydrogenase, glucose-6-phosphat dehydrogenase, diaphorase, superoxid dismutases 1 and 2, purin nucleoside phosphorylase, glutamate oxaloacetate transaminase, hexokinase, kreatin kinase, adenylate kinase, phosphoglucomutase, carboanhydrase, leucin arylaminopeptidase, peptidases A and B, adenosine desaminase, fumarate hydratase, glucose phosphate isomerase, mannose phosphate isomerase.

Comparative analysis of polymorphic loci shared between soy cultivars and populations of wild Ussurian soy. The comparison between domestic soy and its wild relaive was carried out on 42 loci, encoding principal enzymatic systems of general intracellular metabolism, using protein separation via starch gel followed by immuno-staining. We observed 21 polymorphic loci (from 42) for all plants studied. The evaluated biochemical systems were further divided into two groups: enzymes participating in intracellular processes of ATP accumulation (glycoliysis, Krebs cycle – defined as enzymes participating in glucose metabolism – G); and all others (NG). In total 21 G and 21 NG were analyzed.

Animal DNA markers. We extracted nuclear DNA from blood cells using standard protocol. The method RAPD-PCR was applied with the use of two decanucleotides, which primary sequence was described in the work of Bailey and Lear (1994): UBC-85: 5 '-GTGCTCGTGC-3 ' and UBC-126: 5'CTTTCGTGCT-3 '. The authors have chosen these primers from 212 tested as most convenient for inter- and intra-species analysis of the genetic structure of Equidae species representatives These primers were also used for the analysis of plant species. The reaction mix of volume 20 µl contained: 50 mM KCL, 10 mM TRIS-HCl (pH 9.0), 0.01 % triton X-100, 0.3 mM of each of the dNTPs, 2 mM MgCl2, 0.2 mM primers, 1 unit of polymerase Thermus aquaticus (" Dialat LTD ", Moscow), 20-50 ng DNA. PCR was carried out on thermocycler "Biocon" (Moscow). At use of a method RAPD-PCR the temperature mode was the following: 5 cycles - 1min at 92 C, 1 min. at 35 C, 2.5 min. at 72 C; 35 cycles - 1 min at 92 C, 1 min at 42 C, 2.5 mines at 72C (in summary 40 cycles). Inter-Simple Sequences Repeats (ISSR-PCR) DNA markers [39] allowed estimation of similarities and differences in genome distribution of DNA fragments, flanking by invert repeats of microsatellite locus. In summary, by using of 3 di- a d 12 trinucleotide ((AGC)6T, (TGC)6A, (AGC)6G, (ACC)6G, (GCT)6A, (GAG)6C, (TCG)6G, (CTC)6A, (CAC)7A, (CTC)6C, (GTG)7C, (CAC)7T) as primers in polymerase cycle reaction (PCR) there were 310 amplicons identified. In the case ISSR-PCR of reaction carried out in such temperature mode: initial denaturation - 2 min. at 94 °C; 30 cycles: 30 sec at 94 °C, 30 sec at 55 °C, 2 minutes at 72 °C; terminal elongation - 10 min at 72 °C; cooling down to 4 °C. Amplicons were identified by electrophoresis in 1.5% agarose gel with ethidium bromide visualized in UV-light. Only those amplicons were taken into account which were reproduced in 3-5 independently repeated PCRs from the same DNA. To identify amplicon lengths a marker of molecular weights was used (100bp DNA Ladder Gibco BRL).

The analysis of allele distribution in genes, involving desirable cattle phenotype traits. For this analysis we used genotype evaluation methods described by [38] was carried out. Allele variants were considered by the following structure genes: coding diary proteins (kappa casein – CSN3 and beta lactoglobulin –BLG), myostatin participating in control of muscle mass growth rate (MSTN), hormone of lipid metabolism (LEP), growth hormone (GH) and transcription regulation factor – a locus of somatotropic hormone (PIT-1) by PCR amplification with further restriction (RFLP-PCR). All 5 loci surveyed in cattle are located in different chromosomes (PIT 1 is at chr. 1; LEP is at chr. 4; CSN3 – 6; BLG – 11 and GH – 19).

CNS3 amplification product included part of 4th exon and 4th intron, which gave after restriction by Hind III two allele

variants A and B. The presence of B essentially increases the quality of firm cheeses while A is associated with high total yield and dominates in diary cattle breeds.

BLG locus was 247 bp long and included part of 4th exon and 4th intron. Allele variant BLG A is associated with high milk yield.

LEP amplification product was 1830 bp long and included part of 2nd exon, whole 2nd intron and part 3th exon. By using of Sau3A restriction 3 allele variants were revealed (A, B and C). It has been shown that in LEP AA genotype (it has 2 digestion sites for Sau3A) is associated with decreased fodder efficiency compared to BB (with additional digestion site); AC genotype is associated with high butter-fat and protein containment in milk and also with the best lactation dynamics.

In GH locus a fragment of 5th exon was amplified and had 223 bp. By presence of digestion site of Alu I we revealed 2 variants: L (leucine in 127) and V (valine in the same position). Several researchers found for GH gene that milk of cows with LL genotype contain more fat and protein than VV genotype but also has a bit lower total yield.

During amplification of fragment of 6th intron of Pit-1 gene (1355 bp long) and further restriction by Hinf I two allele variants were found (A and B). Allele A was associated with higher protein yield but lower fat yield in investigation of [38].

Ethical guidelines during sample collection. All of the blood samples analyzed in this work were collected during, and were part of, routine veterinary check-ups – for both wild animals in the biosphere reserves, and for their domesticated relatives.

Statistical Analysis. The allelic and genotype frequencies, Nei's distances, estimation of gene balance according to the Hardy-Weinberg's law, and cluster analyses were carried out with use of the standard computer programs "BIOSYS-I", "TFPGAPRG". The p-values were obtained using the Student's t-test.

List of abbreviations: bp, base pair(s); CpG, C/G phosphate linked; dN, nonsynonimous substitution; dS, synonymous substitution; ESTD-1, Esterase D; FC, Fertile Crescent; ISSR, inter-simple sequence repeat(s); kb, kilobase(s); LTR, long terminal repeat(s); miRNAs, microRNAs; PCR, polymerase chain reaction; RAPD, random amplified polymorphic DNA; SNP, single nucleotide polymorphism; VIGG, Vavilov Institute of General Genetics, Russia

Competing interests. The authors declare no competing interests.

Authors contribution. Valery Glazko formulated the research questions, designed the study, and performed the experiments. Tatyana Glazko and Boris Zybaylov performed data analysis, and wrote the manuscript.

Funding. The authors' personal funds were used to facilitate the research activities in this article.

Acknowledgements

The authors would like to thank Evgeniya Dimova, PhD, for proof-reading assistance.

References

[1] Bailey E, Lear TL. 1994. Comparison of thoroughbred and Arabian horses using RAPD markers. Anim. Genet. Suppl 1, p.105-108.

[2] Barendse, W., Harrison, B.E., Bunch R.J., Thomas, M.B., Turner, L.B., 2009. Genome wide signatures of positive selection: The comparison of independent samples and the identification of regions associated to traits. BMC Genomics. 10, 178.

[3] Beck, J., Sieber A., 2010. Is the spatial distribution of mankind's most basic economic traits determined by climate and soil alone? PLoS ONE. 5(5), e10416.

[4] Bogolyubskiy, S.N., 1959. Origin and Transformation of Domestic Animals. Soviet Science, Moscow.

[5] Chena, S.Y., Duan, Z.Y., Shaa, T., Xiangyub, J., Wub, S.F., Zhanga, Y.P., 2006. Origin, genetic diversity, and population structure of Chinese domestic sheep. Gene. 376(2), 216–223.

[6] Coop, G., Pickrell, J.K., Novembre, J., Kudaravalli, S., Li, J., Absher, D., Myers, R.M., Cavalli-Sforza, L.L., Feldman, M.W., Pritchard, J.K., 2009. The role of geography in human adaptation. PLoS Genet. 5(6), e1000500.

[7] Diamond, J., 2002. Evolution, consequences and future of plant and animal domestication. Nature. 418(6898), 700–707.

[8] Feeney, K.M., Parish, J.L., 2009.Targeting mitotic chromosomes: a conserved mechanism to ensure viral genome persistence. Proceedings of the Royal Society: B. 276(1662):1535–1544.

[9] Flori, L., Fritz, S., Jaffrezic, F., Boussaha M. et al., 2009. The Genome response to artificial selection: a case study in dairy cattle. PLoS ONE. 4(8), e6595.

[10] Glazko, V.I., 1999. The polymorphism of proteins, RAPD-PCR and ISSR-PCR markers in European and American bison and cattle. Cytology and Genetics. (Tsitol. Genet. Russian). 33(6), 30–38.

[11] Glazko, V.I., 2000. Genetically determined enzyme polymorphism in soy varieties (Glycine max) and in wild soy (Glycine soja). Cytology and Genetics. (Tsitol. Genet. Russian). 34(2), 77–84.

[12] Glazko, V.I., 2003. An attempt at understanding the genetic basis of domestication. Animal Science Papers and Reports. 21(2), 109–120.

[13] Glazko, V.I., 2004. Polymorphism of proteins, RAPD-PCR and ISSR-PCR markers in Ungulata species and domestication problems. Proceeding of the Russian Academy of Agricultural Sciences. 4, 40–46.

[14] Glazko, V.I., Tsvetkov, I.L., Ivanov, A.N., 2006. Genetic differentiation of rice cultivars by IRAP markers. Izvestiya of Timiryazev Academy. 4, 155–159.

[15] Glazko, V.I., Tsvetkov, I.L., Sozinova, L.F., Glazko, T.T., 2009. Molecular and genetic markers of DNA polymorphism and their genome positioning. Proceeding of the Russian Academy of Agricultural Sciences (Russian). 3, 11–14.

[16] Glazko, V.I., 2011. Nano- and microscales in genetic material organization: On the issue of Lima-de-Faria "chromosome fields". Dokl. Biochem. Biophys. 436, 5–7.

[17] Glazko, V.I., Bardukov, N. V., Pheophilov, A.V., Sipko, T.P, Elkina, M.A., Glazko, T.T., 2012a. Polymorphism of ISSR and IRAP markers in genomes of musk-oxen (Ovibos moschatus) and horse (Equus caballus) of Altai breed. Izvestia of Timiryazev Agricultural Academy. Special Issue, 16–26.

[18] Glazko, V.I., Pheophilov, A.V., Bardukov, N.V., Glazko, T.T., 2012b. Species-specific ISSR-PCR markers and the ways of their appearing. Izvestia of Timiryazev Agricultural Academy. (1), 118–125.

[19] Glazov, E.A., Kongsuwan, K., Assavalapsakul, W., 2009. Repertoire of bovine miRNA and miRNA-like small regulatory RNAs expressed upon viral infection. PLoS ONE. 4(7), e6349.

[20] Khaldi, N., Shields, D.C., 2011. Shift in the isoelectric-point of milk proteins as a consequence of adaptive divergence between the milks of mammalian species. Biology Direct. 6(1), 40–49.

[21] Kharchenko, P.N., Glazko, V.I., 2006. DNA Technologies in Agrobiology Development. Voskresenie, Moscow.

[22] Lewin, H.A., 2009. It's a bull's market. Science. 324, 478–479.

[23] MacEachern, S., McEwan, J., Goddard, M., 2009a. Phylogenetic reconstruction and the identification of ancient polymorphism in the Bovini tribe (Bovidae, Bovinae). BMC Genomics. 10, 177.

[24] MacEachern, S., McEwan, J., McCulloch, A. et al., 2009b. Molecular evolution of the Bovini tribe (Bovidae, Bovinae): Is there evidence of rapid evolution or reduced selective constraint in domestic cattle? BMC Genomics. 10, 179–193.

[25] Nagornyk, T.A., Lopatina, N.V., Glazko, V. I., 2005. Genetic interrelation between varieties of cultural soja and wild species of Glycine. Proceedings of the National Academy of Science of Ukraine. 8, 163–172.

[26] Nosil, P., Funk, D.J., Ortiz-Barrientos, D., 2009. Divergent selection and heterogeneous genomic divergence. Mol. Ecol. 18, 375–402.

[27] Rebollo, R., Horard B., Hubert, B., Vieira C., 2010. Jumping genes and epigenetics: Towards new species. Gene. 454(1–2), 1–7.

[28] Strömqvist, M., Falk, P., Bergström, S. et al., 1995. Human milk kappa-casein and inhibition of Helicobacter pylori adhesion to human gastric mucosa. J. Pediatr. Gastroenterol. Nutr. 21(3), 288-296.

[29] Tang, H., Sezen, U., Paterson, A.H., 2010. Domestication and plant genomes. Curr. Opin. Plant Biol. 13(2), 160–166.

[30] Tellam, R.L., Worley, K.C., 2009. The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. Science. 324, 522–528.

[31] The bovine HapMap consortium genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. 2009. Science. 324, 528–532.

[32] Trut, L.N., 2007. Domestication in historical process and experiment. VOGiS Herald. 11(2), 273–289.

[33] Van der Kuyl, A.C., 2011. Characterization of a full-length endogenous beta-retrovirus, EqERV-Beta1, in the genome of the horse (Equus caballus). Viruses. 3, 620–628.

[34] Venner, S., Cédric Feschotte, C., Biémont, C., 2009. Transposable elements dynamics: toward a community ecology of the genome. Trends Genet. 25(7), 317–323.

[35] Wade, C.M., Giulotto, E., Sigurdsson, S. et al., 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science. 326(5954), 865-867.

[36] Wilkins A. S., Wrangham R. W., Fitch W. T. The "Domestication Syndrome" in Mammals: A Unified Explanation Based on Neural Crest Cell Behavior and Genetics // Genetics – 2014. - V. 197. – P. 795–808

[37] Zeder, M.A., 2008. Domestication and early agriculture in the Mediterranean basin: Origins, diffusion, and impact. PNAS. 105: (33), 11597–11604.

[38] Zwierzchowski, L., Oprzadek, J., Dymnicki, E., Dzierzbicki, P., 2001. An association of growth hormone, K-casein, β-lactoglobulin. leptin and Pit-1 loci polymorphism with growth rate and carcass traits in beef cattle. Animal Science Papers and Reports. 19, 65–78.

[39] Zietkiewicz E., Rafalski A., Labuda D. 1994. Genome fingerprinting by seguence repeat (SSR) anchored polymerase chain reaction amplification. Genomics. 20, 176–183.