GENOMICS OF DOMESTICATION

T.T. GLAZKO

(Department of Genetics and Enimal breeding of Russian State Agrarian University - MTAA named after K.A. Timiryazev)

Abstract. The analysis of literary and own data on genomic "signature" of domestication and results of researches of the genes — artificial selection targets was carried out. Stored data testified that the genes involved in artificial selection were connected with the systems of hypothalamus — pituitary — adrenal axes, however breeds and species essentially differentiated on a set of such genes, participating in formation of phenotype traits, connected with domestication. The data allowed to conclude that the basic difference between domestic and wild species related with the raised genetic variability in domestic ones, apparently, caused by rather hyperactivity of mobile genetic elements.

Key words: genomic scanning, domestication "signature", landscape genomics, Single Nucleotide Polymorphism (SNP), Inter-Simple Sequence Repeats (ISSR), retrotransposons.

Genetic bases of domestication are the most important thing in fakm animals' genetic resources management. Revealing universal domestication traits on phenotypic level and analysis of their genomic determination remain central problems to be solved.

The importance of learning of the domestication process was noted by N.I. Vavilov who longed for its using in selection: "We want to seize animals, learn to create new forms based on existing ones and suitable for agriculture requirements. But to solve actual selection tasks it's necessary to know the history and apply historical method to the whole problem. Domestic animals selection itself so as plant selection is very similar to experimental evolution, but to understand and rule it we need historical insight of the evolutionary process".

The main difference of domestic species genofonds from close wild ones is in their unique phenotypic variability range which turns out to lots of breeds clearly distinguished from each other by morpho-physiological traits. Interbreed differences are often higher than those observed between close wild species. They say there are about 4,500 mammal species now (including twin species). Only 5 main agricultural mammal species (goats, sheep, cattle, horses and pigs) make 4920 breeds well distinguished by their phenotypes according to FAO data (180 countries, www.fao.org) — that's more than all mammal species diversity. 1484 sheep breeds, 1452 cattle, 803 horse, 627 swine ones and 560 goat breeds are discribed.

Human populations are part of agrarian civilization which began to build itself about 13,000 years ago starting from plant and animals domestication experiments. This process gave growth and development to our civilization and determined global demography. For example, about 88% people speak languages belonging to one of the seven groups ap-

* Author for correspondence: E-mail: vglazko@yahoo.com

peared in two Eurasia regions which are the first domestication centers (Mesopotamia and part of China) [8].

Many questions arise during domestication process studying. Thus, it's not clear why it began only in several regions on the Earth (Mesopotamia, China, South, Central and East America, tropical Africa, Ethiopia, Seychelles and New Guinea). Why were only 14 species domesticated from 148 species of big mammals, weighing more than 45 kilos, and why did only 100 plants give useful domesticants from 200,000 wild ones [10]? What is the mechanism and what's it connected with? Why are common domestication traits found among taxonomically remote animals [1] and plants [19] and distinguish them from close wild ones, breaking Vavilov's law of homological rows of close species variability?

Answers on these questions may promote switching to another level of solution of humanity foodstuff supplying shortage which appeared as result of demographic burst. To imagine originality of domesticated species it's necessary to observe traits which supported or prohibited species involving into domestication. Such traits preventing from successful animal domestication include:

1. Specificity of animal nutrition, difficulties for man to provide food under artificial conditions.

2. Slow growth and long life cycle compared to human (e.g. Elephants still remain half-domesticated).

3. Too high movement speed compared to human.

4. Impossibility of reproduction in captivity.

5. Absence of social hierarchy (no leaders)

6. Tendency to panic and stresses during contact with human (for example some gazelle and deer species)

7. Low adaptation ability to new environmental conditions, low colonization abilities cas ompared to human.

It's easy to see that all these traits prohibiting animal domestication are caused by impossibility to live and reproduce in human environment and by differences in number of nutritional, adaptive, social properties. Perhaps, similarity properties of agricultural species and man in abilities to adapt to artificial environment, movement speed, life cycle and social hierarchy may be considered as additional domestication traits.

So, limited number of domesticated animals is caused by several biological reasons. Revealing them, understanding laws of their heredity and variability become the necessary condition of speed and precision increase in selection.

Ways of origin and expansion of domesticated animals

M. Zeder [22] suggested that animal domestication appeared 11,000-9,500 B.C. beginning from goats and sheep, that happened 1,000 years after plant domestication in the south of Eastern Mediterranean countries (Levant). It appeared to the north and east of plant domestication center, but some suggest another independent domestication event in the southern part of Levant. Archaeological data prove that most bones preserved which were related to decreasing size during domestication hypothesis belong to females or young males. That makes available the following domestication scenario: hunters kept females letting young males involve in their populations, destroying, at the same time, older males on surrounding territories. That means, domestication appeared as hunting strategy in hunters' civilization — keeping female herds to lure big males.

Archaeology says that pig domestication occurred in the south-east of Anatolia 10,500-10,000 B.C. Their geographical expansion, as well as domestication, was very sim-

ilar to sheep but it went slower. Domesticated pigs with morphological changes have never been found in southern Levant or in Iranian plains before 8,500-8,000 B.C. Recent demographical findings are evidence that European cattle were domesticated in Euphrates valley between 11,000 and 10,000 B.C. But asboth pigs and sheep they migrated relatively slow to Fertile Crescent region. Domesticated pigs and cattle having morphological changes in Anatolia before were not found 8,500 B.C.

Genetic data support such scheme. Recent studies showed that ancestral species for sheep and goat belong to species existing in FC *(Ovis orientalis* и *Capra aegagrus* respectively). Moreover, these domesticated species have at least 4 genetically different domesticated lines or haplotypes (goat has 6). It's still not clear if these lines refer to space and time domestication events. Cattle genetic data are evidence of the presence of 5 different haplotypes, while 3 or maybe 4 of them originated in FC. Similarly at least 4 of many pig lines appeared in the Middle East. Animal domestication in the Middle East occurred after long interaction between man and ancestors of main domesticated species in FC. At that time the general hunting strategy concentrated on maximum local availability of wild *Bovidae* transformed into active management of herds of main 4 species from 11,000 till 10,000 B.C. Even such species as gazelle which behavior was incompatible with domestication, were subjects to be tamed in south and north Levant, where they were the largest wild *Bovidae* group, but domestication results were poor.

Some scientists (M. Zeder, G. Diamond) suppose that the basis of domestication was built by phenotypic and genetic properties of species more adapted to domestication as compared to others. But others think the main factors were combinations of climatic and soil properties, at least, for agrarian civilization expansion.

Ibus, G. Bek and A. Saber [7] supposed as simple null model that only climate and soil quality determine purpose of land use — agriculture, settled animal husbandry, nomadism and hunting-collecting. Applying ecological niches methods (ENM) they have received spatial forecasts for these 4 types and then they combined them with maps of Europe, Asia and Australia. Two aspects were observed — potential conflict situations and density of population. Regions of potential conflict in different land using were marked on map. Results correlate well with many real existing or historical conflicts between ethnic groups and traditions in land using. Moreover, it has been found that model of agricultural land using capability explains most part of density variability. Thus, it correlates with wealth indices of local people (GDP, PPP). Authors conclude that such a simple model provides good forecasts of complex properties in human expansion. It's supposed that results acquired and methods will be very useful, for example, the face of climatic changes.

Then we may expect that success of agrarian civilization expansion is caused by a balance between global soil and climate quality gradients and adaptive potential of man and domesticated animals and plants which create local agrosystems. In interspecific communities, humanity and domesticated species genofonds are in complicate relations, which features are determined by artificial selection and agrarian, ecological and landscape background.

Results of genetic scanning of domesticated animals' genomes

Development of gene and genomic investigation methods led to a possibility to identify genes involved in adaptation, participating in biological systems creation. Next step in evolutional biology will be determination of how geographic landscapes and ecological properties are involved in distribution of this functional adaptive genetic potential [15]. Recent surveys of population genetic of human allow to estimatie what can be found by

using of landscape evolutional genomics. Coop et al. [9] studied global allele frequencies in populations by hundreds of thousands single nucleotide polymorphism. It turned out that few genes in human genome had essential differences in allele frequencies between populations. This may be evidence of selection acts only over a few loci or adaptive phenotypic changes may occur through little changes of allele frequencies in many loci.

Adaptation formed by environment almost always involves various phenotypic changes provided each has complicated genetic reasons. This complexity increases if the landscape level of ecological changes is added. Understanding of adaptation on the natural landscape level may especially be difficult for polygene traits where adaptation goes through little allele shifts of many loci.

Genomic screening is the main trend of the modem population genomics. A set of biological systems in which surveys of genome diversity in different environment have carried out and expanded essentially over the last few years [18]. The logic of this approach is that alleles in neutral genome regions will freely move from one population to another through gene flow, while targets of natural selection will have higher diversity in different environments. Genome screening may vaiy from several hundreds of markers used to true screening by full sequencing.

In 2009 a full sequence of cattle was obtained as a result of international work inside BovineHapMap. 7 times sequencing revealed that it contains at least about 22,000 genes with a common set of 14,345 orthologs similar to 7 mammal species from which 1217 absent and have not been found in nonplacental genomes. Evolutional regions and crucial chromosome parts with high density of segment duplications en riched by repeats and speciespecific gene variants connected with immune response were found for cattle. Genes involved in metabolism are highly conserved, though 5 metabolic genes are deleted or changed essentially as compared to their human orthologs. [20, 21]

A number of cattle immune genes differ in increased amount of copies. These include beta-defensines involved not only in antibacterial defense (unspecific immunity) but also in cellulose digestion in four-chamber stomach. Similar to both rodent and dogs cattle possess of about 1000 genes not found in human. These genes have many variants in promoters and binding motifs of regulation factors. That makes its own contribution in mammal development and physiology differences.

A high density of segment duplications, retrotransposone and retrovirus long terminal repeats have been found in chromosome regions which were rearranged last 80 million years of *Bos taurus* karyotype forming. Then a conclusion is drawn that these repeat elements and segment duplications directly provide chromosomal rearrangements, connected with specie origin in many mammal lines. A high level of genetic variability has been found in all cattle breeds. This level was even higher than the same one among dogs or humans. The maximum genetic diversity is found in zebu with a single nucleotide polymorphism (SNP) at every 285 base pairs (bp).

That's more than 2 times higher than in European Holstein and Angus breeds, thus, there were more progenitors of zebu before domestication. Genetic data are evidence of the fact that cattle breeds after domestication passed through selection bottlenecks with limited number of progenitors or intensive selection by production traits. Moreover, taking into account comparisons of variability distribution based on thousands of SNP the international consortium came to a conclusion that a number of genomic regions differ between beef and diary breeds and most part of them contain genes responsible for quantitative variability of beef and diary productivity and located in cattle chromosomes 2, 6 and 14 [14, 21].

At the same time selection of allele variants in particular among milk protein genes may be resulted not from instinctive work of a selectionist trying to reach high milk pro-

ductivity, but by side product of a natural selection directed to higher resistance to environmental pathogens which domesticated animals faced along with human contacts and colonization of new niches. Thus, in our own investigations a comparison of ratio of synonymous/nonsynonimous (dS and dN) substitutions in kappa casein exon IV determining the size of micelles in milk (chromosome 6) was carried out. Kappa casein is subdivided into two peptides: a signal coded by exons 2 and 3; and a ripe coded by exon 4 where C-terminal domain is located. We made comparisons of specie properties of dN and dS ratios by using Nei-Gojobori method in different parts of kappa casein gene. They revealed that a positive selection supporting high variability of amino acid substitutions in kappa casein C-terminal domain takes place only in *Bovinae* (table 1.) [5].

T a b l e   1

Average values of interspecie differences by ratio of nonsynonymous *I* synonymous substitution (dN/dS) counted by using of Nei-Gojobori method in different parts of kappa casein gene

	Exon IV		RKS Protein		C domain	
	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

C-terminal domain contains all sites of posttranslational casein modifications (phosphorylation and glycosylation). Physical properties (size, solubility) and reactivity of micelle casein may essentially depend on its phosphorylation and glycosylation. It is the C-terminal whuch domain has the highest rate of this gene evolution among *Bovinae* species.

It is found that only in *Bovinae* family the total number of threonine and serine residues remains the same while their positions are changed. In other families both number and positions of these residues remained the same. It is known that various glycosylation distribution in kappa casein C-domain may be accompanied by differences in inhibition of *Helicobacter pilori* pathogen which cause gastrointestinal diseases. We may expect that a high evolution rate of amino acid sequence of this part of kappa casein observed may be caused by an adaptation to various pathogens of closely related *Bovinae* species.

Hence, only a small part of kappa casein (C-domain) and only in *Bovinae* family is characterized by presence of positive selection (high rate of accumulation of nonsynonymous substitutions) during the relatively short time of their divergence. We may expect that this phenomenon resulted from nutritional differences appeared after divergence of these species because of domestication of most part of investigated individuals from this family which led to a necessity of adaptation to various gastrointestinal pathogens.

We also carried out a comparative analysis of polymorphism of 30 loci of various functional protein groups in genofonds of domesticated and close wild species of two animal orders — *Artiodactyla* and *Perissodactyla* among which there were wild "zoo" species reproduced on various biosphere reserve "Askaniya-Nova" and some cattle and horse breeds reproduced in different Russian and Ukrainian farms (26 breeds and intrabreed groups). In total, 12 species were observed. Also 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 kindly provided by Doctor of Biological Science V.V. Sherepitko (Ukraine) and Ph.D. I.V. Seferova (VIGG, Russia). The following data were received.

The average polymorphism level for loci studied was slightly higher among domesticated species compared to domesticated forms. Among domesticated specie it ranged from 0.036 (for pigs) to 0.171 (for cattle); among close wild species — from 0.017 (Grant Zebra) to 0.135 (Kanna antilope). At the same time there are clear differences in contributions in polymorphism of different functional groups of genetic and biochemical systems between these species. Polymorphic loci shares were among intracellular energetic metabolism enzymes (average for number of species studied) 0.179 for domestic *Bovidae* and 0.629 for wild; among exogenous substrate metabolism enzymes 0.464 and 0.193; transport proteins: 0.357 and 0.178 respectively (table 2) [5].

T a b l e 2

Polymorphic loci share of various functional groups of genetic and biochemical systems in wild (A) and domesticated (B) mammal species

Specie groups	Protein functional groups		
	I	II	III
A	0.629	0.193	0.178
B	0.179	0.464	0.357

I — enzymes of intracellular energetic metabolism;
II — enzymes of exogenous substrate metabolism;
III— transport proteins.

Differences in contribution in total polymorphism of different functional groups found between wild and domestic mammal species correlate well with suggestions about connection between species origin and reorganization of energy supplying mechanisms and about   fact thatusually artificial selection doesn't lead to appearance of new species, excludingcases of artificial interspecific hybridization. We may expect that natural selection supports origin of new species and polymorphism of enzymes of intracellular energetic metabolism while artificial selection supports origin of forms adapted to unsteady flow of exogenous substrates. Perhaps, wide range of phenotypic variability among domesticated species is connected with diversity of exogenous substrate metabolism rates. The latter allows suggestion presence of "subgenome" whose variability is essential for wide phenotypical diversity typical to domestic animals that correlates well with "balanced theory" of polymorphism supporting. Obviously, variability of such subgenome may act as necessary condition in directed selection performing.

Further we carried out a comparative analysis of polymorphic loci share between soy cultivars and populations of wild Ussurian soy (suggested progenitor). We observed 21 polymorphic loci (from 42) for all plants studied. Genetic and biochemical systems were divided into two groups: enzymes participating in intracellular processes of ATP accumulation (glycolysis, Krebs cycle — defined as enzymes participating in glucose metabolism — G); and all others (NG). In total 21 G and 21 NG were analyzed. 7 polymorphic loci ofwild populations included 1 NG (ESTD-1) and 6 G loci. From 19 loci of soy cultivars we found polymorphism in 11 G and 8 NG.

Hence, while wild soy had polymorphism in loci, whose products control intracellular energetic metabolism processes (86% from all), among polymorphic loci in cultivars they were less presented (58%) but the share of NG loci was 3 times higher than in wild relatives. This situation is similar to animals: polymorphism involves enzymes which don't participate in glucose metabolism.

Moreover range of genetic variability in G.max turned out to be higher than in G.soja. Polymorphic loci share P: was 45% for G.max and 17% for G.soja. That means domesticated species is more polymorphic than close wild one.

Data acquired evidence in favor of presence of "subgenome" which is marked by loci which products participate in regulation of interactions between inter and outer biochemical medium (enzymes of exogenous substrate metabolism, transport proteins) which higher variability acts as a necessary condition of domestication both for plants and animals.

An analysis of amplification spectra (RAPD-PCR, ISSR-PCR) was carried out for domesticated and close wild species. It showed interspecific differences by using both of them: decanucleotides UBC-85: 5'-GTGCTCGTGC-3' and UBC-126: 5'CTTTCGT-GCT-3'). It turned out that domesticated species had prevalence of short DNA fragments, flanked by inverted repeats of these sequences (table 3) [5].

T a b l e 3

Comparative analysis of frequencies of occurrence of amplicons differing by length in amplification products spectra (RAPD-PCR) received from domesticated and wild Ungulata species by using of decanucleotides UBC-85 and UBC-126 as primers

Species	Amplicon lengths		
	short (0.4.-1.0 kbp,%)	average (1.1-1.9 kbp, %)	long (2.0-2.5 kbp,%)
Domesticated	36,3	50,9	12,8
Wild	29,8	49,0	21,2

ISSR-PCR analysis allowed toestimate both similarities and differences in distribution of such fragments of different length (amplicons) for mammal species investigated. In summary, by using of 3 di- a d 12 trinucleotide (($AGC)_6T$, $(TGC)_6A$, $(AGC)_6G$, $(ACC)_6G$, $(GCT)_6A$, $(GAG)_6C$, $(TCG)_6G$, $(CTC)_6A$, $(CAC)_7A$, $(CTC)_6C$, $(GTG)_7C$, $(CAC)_7T$) there were 310 amplicons identified. The distribuiton was: domesticated species: 2.0-2.5 kbp — 12%, 1.1-1.9 kbp — 38%, 0.4-1.0 kbp — 50%; wild species: 17%, 44% and 39% respectively, (table 4) [5] It's clear that short amplicons are found statistically reliable (P<0.05) more often in domesticated species in ISSR-PCR.

T a b l e 4

A contribution (%) of amplicons of different length received from wild and domesticate mammal species Into summary amplicon spectra collected by using of di- and trinucleotide microsatellite loci fragments as primers

Specie	Wild species	Domesticated species
Amplicon lengths (kbp)	%	%
2.5-1.8 kbp	17	12
1.8-1.1 kbp	44	38
1.1- 0.4 kbp	39	50

Some similarity of estimations of genetic interactions is revealed when constructing dendrograms based on genetic distances counted by markers of both types (proteins, RAPD-PCR, ISSR-PCR).

Data acquired prove that domesticated species differ from close wild ones mainly by structure genes polymorphism which products participate in regulations of interactions between inter and outer biochemical medium and by higher frequency of occurrence of short DNA fragments flanked by inverted repeats.

Higher frequency of occurrence of such fragments correlates well with our suggestions about active role of processes connected with events of horizontal interchange of genetic information, close interactions between human, domesticated spefies genofonds and that microbiota which was their symbionts and pathogens in domestication.

We must mention that maximum amplicon quantity was revealed by using of $(CTC)_6A$ — 33, $(CTC)_6C$ — 33, $(GAG)_6C$ — 32 amplicons, thus, by using of primers be-

longing to purine/pyrimidine tracks. These tracks participate in secondary structures building and, perhaps, are involved into gene expression regulation mechanisms. Relatively large number of amplicons was received by using of 2 other motifs: $(ACC)_6G$ (36) and $(AGC)_6G$ (31). At the same time according to data of the international consortium Bovine HapMap [20,21] the frequency of occurrence of microsatellites with core motif AGC in *Artiodactyla* (cattle, sheep, pigs) is 90 and 142 times higher than in dogs or human respectively. It's interesting that in 39% of cases this microsatellite goes together with retrotransposon Bov-A2 SINE, which is evolutionarily young and specific for cattle genome.

Thus, cattle genome screening proves higher polymorphism of genetic elements connected with immunity system, retrotransposones and enzymes of exogenous substrate metabolism compared to wild species.

Search of domestication "signature" in genomes of domesticated animals

Australian investigators [6] carried out a comparison of their own data with results of cattle genome screening by SNPs made by the international consortium Bovine HapMap in order to reveal selection targets by using of F statistics methods (Fst) which allowed to estimate statistical reliability of allele linkage of flanking SNPs for different genome regions.

That resulted in correlation between data about population differentiation by Fst. But 40% of variability in Fst correspond to differences in breed structure while 66% — to differences by SNPs and their density when the same breeds are studied. By using of 129 SNPs which provided significant differences by Fst in both databases 12 regions were identified. They had an additive influence on fodder assimilation traits, total beef productivity or intramuscular fat as it was found in Australian samples. Four of these regions influenced more than one trait. One of them included R3HDM gene which was under selection in European human population.

The range of maximum Fst value in cattle chromosome 2 contain a number of genes connected with human selection pressure. R3HDM1 and ZRANB3 genes are related with these cattle SNPs. Most breeds are homozygotes with the same alleles but Hereford, Santa Gertrudis and Belmont Red breeds differ in moderate frequency of occurrence of an alternative allele. This region is, for a rather long time, known as associated with positive human selection owing to lactase gene presence (LCT) and human adaptation to milk consumption by adults. But it's unlikely that cattle was selected by lactase activity in adult condition because all animals finish breast feeding equally. Recently it has been shown that R3HDM1 locus is under positive selection in European human population and doesn't diverge by "autostop" principle with LCT gene. This information allows to consider cattle research results as a starting point to find out traits, associated with selection among *Homo sapiens*.

To compare evolution rates of regions in which domestication "signature" has been observed McEachem et al. [16, 17] chose for sequencing genomic regions where genes corresponding to milk QTL in chromosomes 1, 2, 4 and 9 according to prior investigations. Researchers compared a large amount of genes and transcripts and found essentially higher evolution rate there compared to average genomic values for *Bos taurus, Homo sapiens,* and *Bos indicus.* Hence, authors arrived at a conclusion that such a high evolution rate in these regions may be directly relevant to selection. So, 15 candidate genes were selected, assuming their QTL control over diary productivity and with known chromosome localization, 84 full amplicons were also selected which covered their exons including noncoding flanks which were sequenced to find out ratios between synonymous and nonsynonymous substitutions in different *Bovinae* species.

If a mutation in coding genome region is not significantly adaptive dN/dS ratio must be approximately equal among closely related species. Furthermore, when a neutral variability takes place dN/dS ratio in interspecie divergence must be close to dN/dS in intraspecie polymorphism. Checking up this hypothesis is especially interesting for closely related *Bovini* species because domestication led to very quick phenotypic differentiation as a result of intensive artificial selection among several domesticated species as compared to close wild ones.

Authors made comparisons of genes involved into cattle and close wild species' diary productivity to estimate influence of domestication on molecular differentiation between them. dN/dS data received were further analyzed to find out if they had been a result of positive selection, of effective number decreasing or selection intensity decreasing. It was found that domestication led to dN/dS increasing in European cattle breeds *(Bos taurus)* and southeastern gayal *(Bos frontalis)*. But high dN/dS values inside *Bos taurus* as compared to divergence between species do not point at presence of positive selection of these genes. Similarly lower dN/dS ratio among bisons, which recently had a sufficient decrease in population, is evidence of only population number decrease doesn't lead to essential change of this parameter.

Authors conclude that selection result depends on effective population number and selection coefficient. Usually during the domestication selection pressure on traits important to adaptation decreases for wild species. That may lead to rapid evolution observed among domesticated species especially for *B. taurus* and *B. frontalis* which have the highest dN/dS ratio among Bovini. Surprisingly significant differences in assumed neutral substitution levels between synonymous and noncoding sites in cattle genome were found: it was 30% higher in synonymous sites. This may be connected at least partially with an excess of CpG dinucleotides in synonymous sites which will make its own contribution into time estimations of specie divergence counted by molecular data.

An important contribution into development of ideas about specificity of such domestication "signatures" is made by data on absence of linear relations between certain genes variability and selection pressure features. Thus, in surveys of the same authors [16, 17] they have found a number of coincidences of polymorphism of noncoding sites in Hereford yaks, Herefords — Holsteins and yaks — bisons that allowed to conclude that they are descendants of different lines appeared in common ancestor specie. It turned out that Herefords are more similar to yaks than to Holstein by nucleotide substitutions in gene receptor to serotonin 6 (5HT6) and protein binding glucocorticoid modulator (CMTB1) these are genes directly involved into neuroendocrine regulation.

We must mention that in the unique experiment of wild specie domestication, began by academician D.K. Belyaev and continued by L.N. Trut [4], data were obtained that main and starting domestication target is the key regulatory system of organism level — hypothalamo-pituitary-adrenal system. It was clearly shown that selection by behavior weakens activity of that system not only on the level of phenotypic properties of this activity but also on the level of gene expression (corticotrophin releasing factor, propiomelanocortin, glucocorticoid receptor).

At the same time results of genomic screening by SNP of specialized diary breeds in France say that many genes related with neuroendocrine system formation participate in answer on selection by diary productivity traits [12].

More than 95% of lactating cows in France belong to Holsteins (HOL), Normande (NOR) or Montbeliard (MON). Herd-books about these three breeds were begun in 1922, 1883 and 1872 respectively, created by individuals originating from remote places (North Europe, Northwestern France and Middle Eastern France). From the middle of the XXth

century these three breeds were subjects to an intensive selection to diary productivity improvement. At the same time high yielding cows have negative correlations between the quality of milk and successfulness of the artificial insemination (AI) according to investigations done for HOL, NOW and MON.

Full genome screening of 42,486 SNPs provided identification of 13 highly reliable regions which are under strong pressure of positive selection. A number of them contain genes, which allele variants influence essentially either diary productivity traits (GHR) or color (MC1R). MC1R is a gene which product determines the ratio between eumelanin and pheomelanin and corresponds to extension locus involved into cattle color control. Because of determinative importance of color properties for herd-book corresponding genes are the first selection targets immediately after breed standard definition. Gene analysis in these 13 regions allowed to conclude they participate in gene networks which allow suppose that in answer on selection by diary productivity the central role belongs to somatotropic and gonadotropic axes. Authors suppose that data acquired explain, on the genomic level, antagonism between diary productivity and reproduction characteristics for high yielder diary breeds [12].

But gene networks including different genes of somatotropic axis and revealed as selection targets didn't coincide for three investigated breeds.

In our own investigations an analysis of allele and structure gene involving into desirable genotypes distributions was carried out for 5 autochtonous cattle breeds reproducing in both Ukraine and Russia. 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).

CNS3 amplification product included part of 4th exon and 4th intron, which produced 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, the whole 2nd intron and part 3th exon. By using 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 as efficiency compared to BB (with additional digestion site); AC genotype is associated with high butter-fat and protein content in milk and also with the best lactation dynamics [23].

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 W 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). A is associated with higher protein yield but lower fat yield [23].

In order to estimate if it's possible to consider distribution of allele variants as an additional breed characteristic or it has an intrabreed variability comparisons of two groups of Grey Ukrainian breed reproduced under different ecological and geographical conditions in generations (Kherson region, "Askaniya-Nova" and "Cherga", Altai region). It turned out that frequencies of occurrence of allele variants among these two groups coincide almost in all genes investigated. This allows to suggest that allele occurrence of these genes

may serve as an additional breed characteristic and doesn't depend on either ecological or geographical conditions in which animals are reproducing (table 5) [5].

T a b l e 5

Genotype and allele distribution in loci involved into 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	–

Investigations into allele frequencies distribution of named genes carried out coincide with breed, belonging to different production directions: for diary breeds higher frequency of occurrence of allele variants associated with high total yield was typical as compared to beef or multiple purposes. But at the same time we found no associations between presence of such "diary" alleles and diary yield, of certain cows. All 5 loci surveyed are located in different chromosomes (PIT 1 is at chr. 1; LEP is at chr. 4; CSN3 — 6; BLG — 11 and GH — 19). It is likely to mean that in cows with equally high yields determinative contribution might be made by genotypes of different loci including those researched by us.

This result coincides with gene networks investigations data among three diary breeds in France [12]. In conclusion, domestication "signatures" revealed this time in comparison with different mammal species and breeds differing by production direction evidence about wide range of genes involved into this process. Most of them take part in neuroendocrine regulation systems, immune system, and in milk proteins coding. But the same spectrum of phenotypic traits is caused by genotypes of different genes involved into gene networks of named systems. Hence, the main component providing domestication possibility must be high genetic variability of such systems.

High density of retrotransposons, retrovirus long terminal repeats found in regions of segment duplication in cattle genome, higher number of copies of genes related with immunity system [20, 21] allows suggestion of possible participation of interactions between domesticated species and various pathogens in increase of variability among domesticated species.

Possible sources of genetic variability in domesticated species

To find out possible sources of high genetic variability and presence of elements connecting cattle with possible fodder resources we carried out an analysis of polymorphism of DNA fragments flanked by LTR of soy transposon SIRE-1 (cat. Number AF053008) for Lebedinskaya cattle breed. Amplification spectra had 14 fragments, 11 didn't have individual variability and were observed also in spectra of other breeds. To estimate possibility of localization in cattle genome DNA fragments homological to soy LTR transposon we used BLASTn program to search in GenBank. Fragments with partial homology (11-23 nucleotides) are found in 20 of 29 cattle autosome sequences and in X and Y chromosomes. Expressing sequences cattle database has homological fragments in mRNA of genes participating in thyroxin folding (SH3BGRL) coding transcription regulation factors (LOC782608, LOC781021, FOXJ1), synthesis of proteins associated with telomeres (tankyrase, TNKS), with plasmatic and nuclear membranes (laminin alpha 1, LAMA1, attractin-like protein 1 —ATRNL1; spectrin containing protein of nuclear membrane 1 — SYNE1), with defense from infectious diseases (T cell receptor alpha, TCRA; one of early inflammation proteins — TLR3). That means in expressed sequences homological regions are revealed mainly in mRNA of transcription regulation factors, structural elements of membranes and receptor proteins of immune system. Moreover, short homological sites are revealed in a number of microRNA which are expressed in cattle genome: bta-mir-2303 (chromosome 12); bta-mir-2356 (chromosome 2); bta-mir-2480 (chromosome 9); bta-mir-2441 (chromosome 5). It is known that microRNA is widely represented in various genomes it participates in gene expression regulation and it is likely to be associated with virus infection [13]. A search of homology in genomes of other taxons showed large amount of homological sites in glycine family and in human in mRNA of a transcription regulation factor participating in chromatin remodeling, containing bromine domain (PBRM1), intercellular skin protein filaggrin (FLG) and membrane bound receptor of neurotropic tyrosine kinase (NTRK2), participating in processes of cell division and differentiation regulation. A homologous region was found also in hens (chromosome Z). Homologous sites are also found in prokaryote genomes.

We also carried out genotyping of main domesticated species of food plants, rice and wheat by DNA fragments flanked by inverted repeats of retrotransposon-like elements of the same family R173 in particular PawS5 and PawS5. This family was described initially for rye *(Secale cereale)*. It has been shown that it includes about 15 thousands of copies in diploid genome and these copies are dispersed over all 7 chromosomes. The size of individual elements ranges from 3000 bp to 6000 bp and these regions are very similar to one of wheat retrotransposons.

DNA fragments flanked by inverted repeats of such sequences vary greatly between rice cultivars [3], wheat cultivars and even between reclaims with common cultivar origin [2].

Further by using of BLASTn program a search in GenBank has been made. We found a large amount of sequences with partial homology to these sequences which were usually localized in P450 polygenic family and immune system genes, transcription regulation factors. Homologous sites have wider taxonomical representation than soy retrotransposon flank and are also found in prokaryotes.

So, analysis of data received about participation of transposable elements in segment duplication evolution, their localization in genomic regions of active transcription, correlations between quantity of built-in provirus transposing elements and resistance of, for example, rice cultivars to retrovirus infections allow to suggest the following

scheme of mechanisms of rapid evolution of sites- artificial selection targets in domesticated species.

We may expect that natural habitat expansion after human migration routes increased the number of contacts of domesticated species with new variants of retrovirus infections and, thus, supported appearance of new transposable elements in genomes. On the one hand such sequences were conserved because of the natural selection (they prohibited recurring infection), but on the other hand they increased genetic variability in sites of their integration (insertion mutagenesis, recombinational processes) which might lead to appearance of new mutations essential for the artificial selection.

Participation of transposable elements in divergence of domesticated and close wild species could explain some empirical data. In particular, the relatively high evolution rate of several genetic elements in genes of domesticated species and our recently received data about higher frequency of occurrence of short DNA fragments flanked by inverted repeats in genomes of domesticated species compared to close wild forms.

A point in favour of this assumption is made by data on presence of sites which are homologous to retrotransposon fragments typical for forage plants in cattle and human genome. They are presented in sites associated with immune system genes and with genes participating in signal transfer to cell genome (structure elements of plasmatic membrane, nuclear membrane, chromatin, transcription regulation factors).

Another evidence in favor of hypothesis presented is given by recently revealed mechanisms of interaction between viruses and metaphase chromosomes: presence of special virus proteins interacting directly with chromatin proteins allow viruses to keep themselves during cell division (schemes 1, 2) [11].
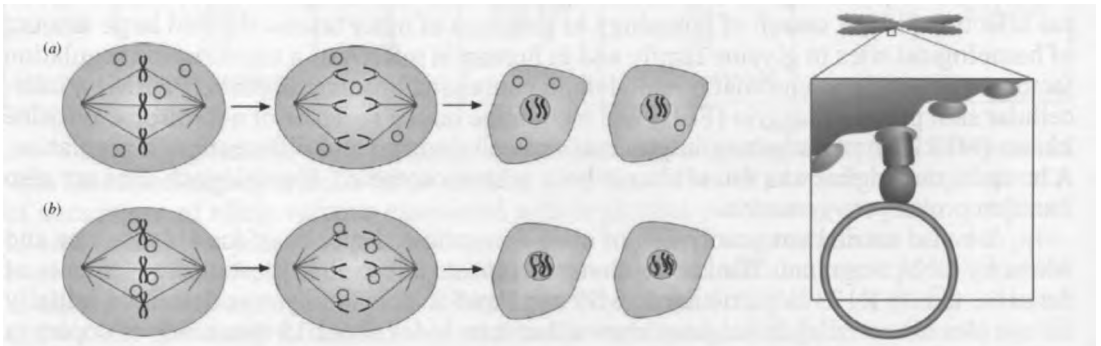
Scheme 1. Possible variants of heredity of nongenomic elements:
a) without contact; b) with conctact with chromatin proteins [11]

Scheme 2. Coding virus DNA— binding protein (shown in red) associates with specific sequence of virus DNA (ring DNA) and immediately binds cellular protein (shown in green) which binds chromatin (shown in blue) [11]

Conclusion

Domesticated species differ from wild ones in specific properties of selection pressure which includes necessity of their interaction with human (neuroendocrine factor), adaptation to a wide spectrum of food sources, pathogens (including "crowd" diseases),

ecological and geographical factors. Genomic screening of cattle genome allowed to reveal high density of segment duplication connected with repeatsm retrotransposons, retroviruses and increased nucleotide variability (higher than in dogs and human). Higher polymorphism of genes coding transport proteins and enzymes of exogenous substrate metabolism was observed as compared to close wild species. Cattle differ also in higher amount of gene duplication related with immune system (beta defensin, interferon, lysozyme). Cattle have a positive selection (supporting variability) in regions of localization of genes related to somatotropic axis — the key neuroendocrine target of the artificial selection — but there are interbreed differences in variability of various genes of genetic networks of this axis. Prevalence of regions homologous to retrotransposon elements of forage plants allows to suggest that a source of genetic variability providing adaptive potential of domesticated species is relatively higher pathogenic repertoire which are faced during colonization compared to close wild ones.

References

1. Bogolubsky S.N. Origin and transformation of farm animals. Moscow: Sov. Nauka, 1959. 593 p.

2. Glazko V.I., TsvetkovI.L., Sozinova L.F., Glazko T.T. Molecular Genetic Markers of DNA Polymorphism and Their Genomic Positioning// Russian Agricultural Sciences, 2009. Vol. 35. No. 3. Pp. 135-139.

3. Glazko V.I., Tsvetkov I.L., Ivanov A.N. Genetic differentiation of rice varieties on IRAP markers // Izvestia of Timiryazev Academy, 2006. Iss.4. P. 55-60.

4. Trut L.N. Domestication in historical processes and in experiment// VOGiS Herald, 2007. Vol. 11. N. 2. P. 273-289.

5. Khartchenko P.N., Glazko VI. DNA-technologies in agrobiology development. Moscow: Voskresenje, 2006. 480 p.

6. Barendse W., Harrison B. E., Bunch R. J. et al. Genome wide signatures of positive selection: The comparison of independent samples and the identification of regions associated to traits// BMC Genomics, 2009. Vol. 10. P. 178.

7. BeckJ., Sieber A. Is the Spatial Distribution of Mankind's Most Basic Economic Traits Determined by Climate and Soil Alone?// PLoS ONE, 2010. Vol. 5, N5. P. el0416.

8. Binford L.F. In New Perspectives in Archaeology//Aldine. Chicago, 1968. P. 313-341.

9. Coop G., Pickrell J.K, Novembre J. et al. 2009 The role of geography in human adaptation.//PLoS Genet., 2009. Vol. 5. P. el000500.

10. Diamond J. Evolution, consequences and future of plant and animal domestication//Nature, 2002. Vol. 418., N 6898. P. 700-707.

11. Feeney K.M., Parish J.L. Targeting mitotic chromosomes: a conserved mechanism to ensure viral genome persistence// Proc. R. Soc. B., 2009. Vol. 2. N 9. P. 1642.

12. Flori L., Fritz S., Jaffrezic F, Boussaha M. et al. The Genome Response to Artificial Selection: A Case Study in Dairy Cattle// PLoS ONE, 2009. Vol. 4. N.8. e6595.

13. Glazov E.A., Kongsuwan K., Assavalapsakul W. Repertoire of Bovine miRNA and miRNA-Like SmallR egulatory RNAs Expressed upon Viral Infection// PLoS ONE, 2009. Vol. 4. № 7. P. e6349.

14. Lewin H.A. It's a Bull's Market // Science, 2009. Vol. 324. P. 478-479.

15. Lowry D.B. Landscape evolutionary genomics // Biol. Lett., 2010. Vol. 6. P. 502-504.

16. MacEachern S., McEwan J., Goddard M. Phylogenetic reconstruction and the identification of ancient polymorphism in the Bovini tribe (Bovidae, Bovinae) // BMC Genomics, 2009. Vol. 10. P. 177.

17. *MacEachern S., McEwan J., McCulloch A. et al.* Molecular evolution of the Bovini tribe (Bovidae, Bovinae): Is there evidence of rapid evolution or reduced selective constraint in Domestic cattle? // BMC Genomics, 2009. Vol. 10. P. 179-193.

18. *Nosil P., Funk D.J., Ortiz-Barrientos D.* Divergent selection and heterogeneous genomic divergence.// Mol. Ecol., 2009. Vol. 18. P. 375-402.

19. *Tang H., Sezen U., Paterson A.H.* Domestication and plant genomes//Curr Opin Plant Biol., 2010. Vol. 13. N.2. P. 160-1666.

20. *Tellam R.L., Worley K.C.* The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution//Science, 2009. Vol. 324. P. 522-528.

21. The Bovine HapMap Consortium Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds//Science, 2009. Vol. 324. P. 528-532.

22. *Zeder M.A.* Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact//PNAS, 2008. Vol. 105. №. 33. P. 11597-11604.

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

Аннотация. Выполнен анализ литературных и собственных данных о геномных «росписях» доместикации в геномах млекопитающих, рассмотрены результаты исследований генов - мишеней искусственного отбора. Накопленные данные свидетельствуют о том, что основные гены, вовлекаемые в искусственный отбор, связаны с системами гипоталамус - гипофиз - надпочечниковой оси, однако породы и виды существенно отличаются по набору таких генов, участвующих в формировании фенотипических признаков, связанных с доместикацией. Делается вывод о том, что основное отличие обусловлено повышенной генетической изменчивостью доместицированных видов по сравнению с близкородственными видами, по-видимому, обусловленная относительно повышенной активностью мобильных генетических ээлементов.