

THE FRAGMENTS OF HOMOLOGIES OF ENDOGENOUS RETROVIRUSES IN THE GENOMES OF PLANTS AND ANIMALS

M.A. ELKINA¹, V.I. GLAZKO^{1, 2}

¹*K.A. Timiryazev Russian State Agrarian University*, 49, ul. Timiryazeva, Moscow, 127550 Russia, e-mail vglazko@yahii.com;

²*Center for Experimental Embryology and Reproductive Biotechnology*, str. 4, 12, ul. Kostyakova, Moscow, 127422 Russia, e-mail vglazko@yahoo.com

Received March 31, 2014

Abstract

Using sequences of mobile genetic elements in order to mark polymorphic sites in their insertion might be the most effective approach to identify specific features of gene pools of different groups of organisms and control their dynamics, which is especially important in dealing with agricultural species. *Sireviruses* (*Pseudoviridae*) one of the oldest LTR-retrotransposons are widespread throughout the plant kingdom. SIRE-1 is one of the largest and most detailed studied retrotransposon. Its analysis showed that insertions of the retroelement (particularly in the genome of maize) have occurred recently. We studied the polymorphism of DNA fragments flanked by LTR-retrotransposon SIRE-1 (IRAP-PCR markers) in the genomes of different taxonomic groups. The objects of our study were *Triticum aestivum* (Moscovskaya 39, Mironovskaya 808, Omskaya 36 varieties), *Glycine soja* (five wild populations of Primorskii region of Russia) and *G. max* (China), as well as representatives of the factory and indigenous breeds of cattle — Black-and-White cattle improved by Holstein cattle, Ayrshire, Yakut and Red Estonian cattle (97 animals in total). A terminal site of SIRE-1 was chosen as primer in IRAP-PCR (GCA-GTT-ATG-CAA-GTG-GGA-TCA-GCA). The data indicate that the multiloci genotyping by IRAP-PCR using retrotransposon LTR SIRE-1 as a marker reliably differentiates not only representatives of the monocotyledonous and dicotyledonous plants, but also their varieties. Spectrum of DNA fragments (13 to 16 depending on the breed) obtained in studies different breeds of cattle using IRAP-PCR markers was in the length range of 330-1470 bp. The highest polymorphism of DNA fragments was observed in the middle part of the spectrum (760-980 bp) in Ayrshire and Black-and-White cows. According to the obtained dendrogram one of the groups of Black-and-White cows and Ayrshire cattle were closer to each other, while the other with Red Estonian and Yakut cattle became isolated in a separate subcluster. Identified differences in heterogeneity in the two studied groups of Black-and-White from different farms possibly could be explained by the peculiarities of breeding work carried out or with other factors of artificial and/or natural selection. The possibility of using LTR-retrotransposons as molecular genetic markers for polyloci genotyping plants and animals is discussed.

Keywords: inverted terminal sites of LTR-retrotransposons, *Glycine*, *Triticum*, *Bos taurus*, polymorphic information content, the share of polymorphic loci.

A lot of molecular methods developed recently to mark different polymorphic region of genomes (e.g. RFLP, a restriction fragment length polymorphism, SNP, a single nucleotide polymorphism, AFLP, an amplified fragment length polymorphism, SSR-PCR, simple sequence repeats PCR, etc.) contribute to successful genotyping in both cultivated plant varieties (1-4), animal breeds (5-7). Of these markers, the mobile genetic elements have special status comparing to other structural and functional genetic elements due to their capability of transposition, and the high rate of transposition allows suggesting their essential role in generation of genetic variability (8).

In plants the LTR (long terminal repeat) retrotransposons make up a significant portion of genomes, particularly slightly more than 7 % in *Arabidopsis thaliana* (9), 50 and 90 % in rice and wheat, respectively (10), 75 % in corn (11).

Sireviruses is one of the oldest LTR-retrotransposons, which became

widespread in the plant kingdom and the only member of *Pseudoviridae* family that supposed to have gene similar to that encoding coat proteins of viral particles (i.e. *env*-like gene). SIRE-1 of about 11 000 bp is one of the largest and well studied member of *Sireviruses*. SIRE-1 sequence analysis evidences that SIRE-1 insertions of the retroelement (particularly in the genome of maize) have occurred recently.

The analysis of the sequence of SIRE-1 shows that the insertion of this element in the genome of corn occurred recently (12). The using of the sequences of mobile genetic elements in order to mark polymorphic sites of their insertion might be the most effective approach to identify specific features of gene pools of different groups of organisms and control their dynamics, which is especially important in dealing with agricultural species.

Herein, we studied the possibility to use terminal regions of endogenous retroviruses to characterize specific parameters of cultivated plant varieties and genetic features of cattle breeds.

Technique. Monocots, the *Triticum aestivum* plants, and dicots, the *Glycine soja* and *G. max* plants, were investigated. The examined wheat varieties were Moskovskaya 39 (soft winter wheat), Mironovskaya 808 (soft winter wheat derived from a spring wheat), and Omskaya 36 (soft spring wheat). Soybean was represented by 5 wild populations of *G. soja* derived from Primorskii krai, and a weedy, or semi-wild, form of *G. max* from China. Of tested animals, there were Black-and-White holsteinized cattle from commercial herds (Moscow Province) and the vivarium of the Russian State Agrarian University (Moscow) (groups I and II, respectively), Ayrshire cattle (Moscow Province), Yakut cattle (Sakha Republic) and Red Estonian cattle (Pskov Province) (a total of 97 animals).

The IRAP-PCR markers were applied in polyloci genotyping (13). Genome DNA was extracted by means of DNA-extran 1 Commercial Kit (Syntol, Russia). The Tercik amplifier (Russia) and PCR-RT mixture (Syntol, Russia) were used for PCR. The PCR protocol was as follows: initial denaturation at 94 °C for 2 min; denaturation at 94 °C for 30 sec; annealing (hybridization) at 55 °C for 30 sec; elongation at 72 °C for 2 min; final elongation at 72 °C for 10 min (35 cycles). The terminal regions of LTR-retrotransposon SIRE-1 (GCA-GTT-ATG-CAA-GTG-GGA-TCA-GCA) served as primers. Amplification products were analyzed by electrophoresis in 1.5 % agarose with DNA molecular marker GeneRuler™ 100 bp DNA Ladder Plus (MBI Fermentas, USA). The bands were stained by ethidium bromide and visualized in UV light. Only the fragments less than 1500 bp in size were examined to avoid inaccuracy.

TFPGA program was applied for calculation of the genetic distances. PIC (polymorphic information content) index was calculated according to the equation for diallel loci: $PIC = 2f(1 - f)$ where f is a frequency of one of two alleles (14).

The GenBank database (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) was used to find out the homologous sequences.

Results. Because of prevalence of LTR-retrotransposons, including SIRE-1, in plant genomes, we used its terminal regions as primers in IRAP-PCR. As a result, both in soybean and wheat the clearly reproducible spectra of DNA fragments of the same size rang were obtained, particularly up to 22 loci of 350-1240 bp and 26 loci of 220-1450 bp in total (Fig. 1, 2).

The examined *Glycine* groups were high polymorphic ($PIC_{\text{avarage}} = 0.414$, $P = 91\%$) comparing to *T. aestivum* ($PIC_{\text{avarage}} = 0.120$, $P = 65\%$). The PIC_{avarage} was calculated as the average value of the PIC index across the spectrum

of amplicons.

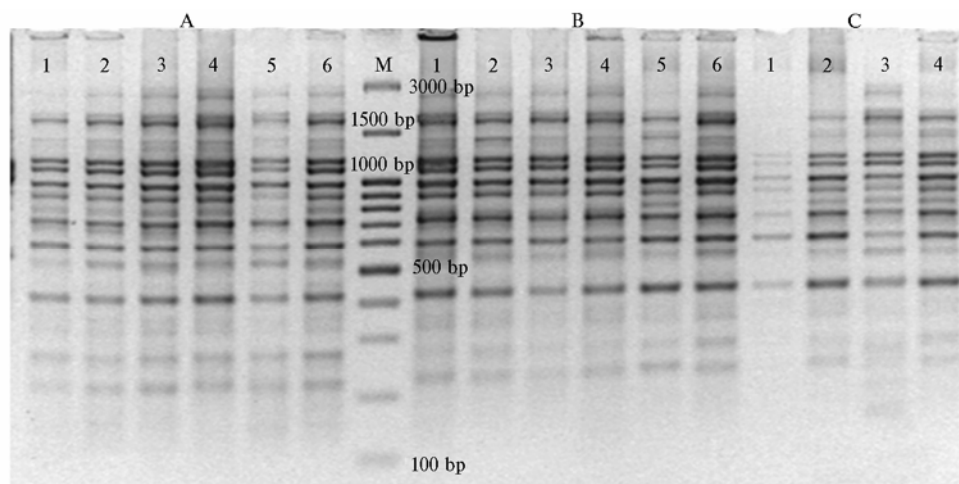


Fig. 1. Electrophoretic spectra of fragments obtained in IRAP-PCR on DNA from winter (A — Moscovskaya 39, B — Mironovskaya 808) and spring (Omskaya 36) varieties of wheat *Triticum aestivum*: M — molecular marker GeneRuler™ 100 bp DNA Ladder Plus (MBI Fermentas, USA). Terminal region of retrotransposon LTR SIRE-1 was used as primers.

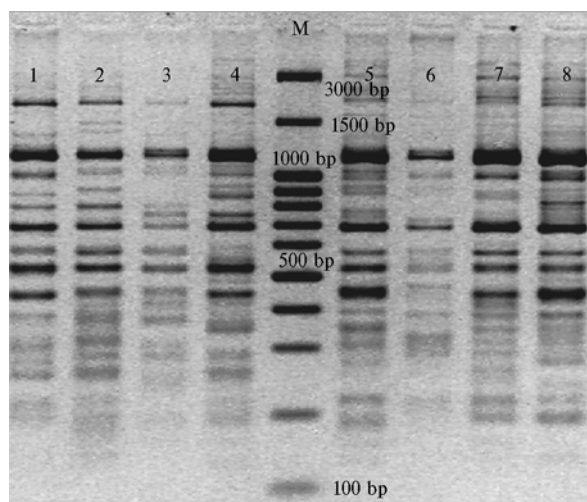


Fig. 2. Electrophoretic spectra of fragments obtained in IRAP-PCR on DNA from different *Glycine soja* groups: 1-4 — group I, 5-8 — group II; M — molecular marker GeneRuler™ 100 bp DNA Ladder Plus (MBI Fermentas, USA). Terminal region of retrotransposon LTR SIRE-1 was used as primers.

The sequences homologous to the fragment of LTR SIRE-1 were found using databases of the GenBank, so 122 and 102 fragments were found in the genomes of *T. aestivum* and *G. max*, respectively (Table 1). In *Glycine*, 350 to 490 bp fragments and 1010 to 1240 bp fragments appeared the most polymorphic. For them, PIC_{average} values were 0.449 and 0.364, respectively. Conversely, in *T. aestivum* middle-sized fragments were high polymorphic, in particular, for 520 до 720 bp fragments and 760 to 990 bp fragments the PIC_{average} values were 0.196 and 0.155, respectively. All large fragments were monomorphic and found in

100 % of tested samples of wheat (Table 2).

1. Number of fragments homologous to terminal sites of retrotransposon LTR SIRE-1 in genomes of soybean *Glycine max* and wheat *Triticum aestivum* and some genetic parameters based on IRAP-PCR data

Parameter	<i>G. max</i>	<i>T. aestivum</i>
Genome length, ×10 ⁹ bp ¹	1.10	1.78
Number of:		
homologous region found ¹	122	102
DNA fragments in amplified spectrum	15-22	26
PIC averaged on the primer	0.414	0.120
P, %	91	65

Comments. 1 — according to <http://blast.ncbi.nlm.nih.gov/Blast.cgi>; IRAP-PCR — inter-retrotransposon amplified polymorphism PCR, PIC — polymorphic information content, P — share of polymorphic loci.

2. Frequency of amplicons (FA) in patterns obtained in IRAP-PCR with terminal region of retrotransposon LTR SIRE-1 as primers, and PIC per locus in wheat *Triticum aestivum* varieties

Fragment length, bp	Moskovskaya 39		Mironovskaya 808		Omskaya 36	
	FA	PIC _{locus}	FA	PIC _{locus}	FA	PIC _{locus}
1450	0	0	0.2	0.159	1	0
1420	1	0	1	0	1	0
1370	1	0	1	0	0.4	0.349
1340	0.8	0.483	0	0	0.4	0.349
1210	1	0	1	0	1	0
1130	1	0	1	0	1	0
990	1	0	1	0	1	0
900	1	0	1	0	1	0
810	1	0	0.3	0.300	1	0
790	1	0	0	0	0	0
760	0	0	0.8	0.483	0	0
720	1	0	1	0	1	0
680	1	0	1	0	1	0
630	0.500	0.414	0.800	0.483	1	0
590	1	0	1	0	1	0
570	0	0	0.3	0.3	0	0
550	0	0	1	0	1	0
540	1	0	0	0	0.2	0.189
520	0.300	0.300	0	0	0	0
460	1	0	0.700	0.488	0	0
420	1	0	1	0	1	0
360	1	0	1	0	1	0
320	1	0	1	0	1	0
270	1	0	1	0	1	0
250	0.500	0.414	0.300	0.300	0.200	0.189
220	1	0	1	0	1	0
P, %	15		27		15	

Comments. IRAP-PCR — inter-retrotransposon amplified polymorphism PCR, PIC — polymorphic information content, P — share of polymorphic loci.

In *Glycine* there was the only monomorphic locus of 700 bp, at 93 % as an average level of polymorphic loci for all amplicons and $PIC_{average} = 0.414$. That evidences about relatively high genetic diversity in the investigated groups on both levels (i.e. within the genus *Glycine*, and within the species *G. soja*). The locus of 680 bp was not observed in *G. max* but found in *G. soja*.

A specific polymorphic spectrum was obtained for each wheat variety in IRAP-PCR with LTR SIRE-1 as primers. Thus, in Moskovskaya 39 wheat the DNA fragment of 790 bp was observed in tested samples while in Mironovskaya 808 and Omskaya 36 varieties it was not found. Moreover, the locus corresponding to DNA fragment of 550 bp was not found in Moskovskaya 39 wheat but in other varieties this locus was detected with $PIC = 1$.

It was shown that inbreeding in plants leads to physiological and phenotypical changes, presumably due to epigenetic effects resulted from a pairwise interaction of chromosomes. Such epigenetic changes can cause activation of mobile genetic elements and, as a result, the intensification of karyotype modification, the increase of genetic diversity and epigenetic rearrangements, being an important source of phenotypic variability (15).

Basing on this suggestion, we can partly explain the high polymorphism within loci in *T. aestivum*, especially considering the wheat as a self pollinating plant. It is also known that the number of plants with cross-pollination in this autogamous culture reaches to 3 % and, sometimes, up to 10 % of the population. Besides, there are some forms for which an increased frequency of cross pollination, from 2.6 to 5.0 % depending on the variety, is peculiar (in particular, the line of soft wheat Mironovskaya winter varieties) (16). Hence, the high intervarietal genetic heterogeneity, e.g. the large share of polymorphic loci and an increase of PIC value (see Table 2), in the Mironovskaya

808 if compared with two other varieties is probably due to this fact.

The loci of 790 and 550 bp fragments in wheat, and the loci of 790, 680 and 550 bp fragments in soybean seem to be important because of their unequal prevalence in different varieties.

The genetic distances (DN) were calculated for the plants according to M. Nej (1972) (Fig. 3) basing on the frequency of amplicons with different sizes flanked with inverted regions of LTR SIRE-1. There are two distinct big clusters of monocots and dicots with a specific cluster of wheat varieties Mironovskaya 808 and Omskaya 36 within monocots in the obtained dendrogram (see Fig. 3).

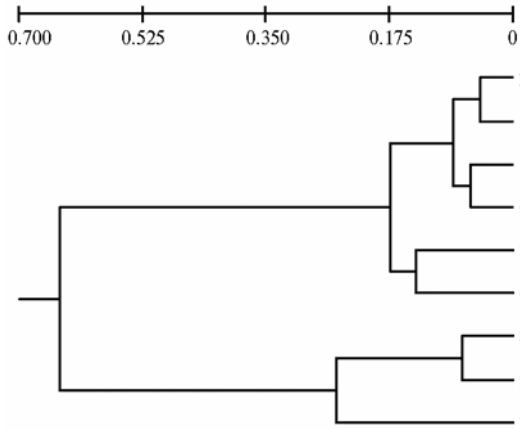


Fig. 3. Dendrogram of studied plants based on the frequency of amplicons flanked by inverted regions of LTR SIRE-1: 1 — *Glycine max* (China); 2-6 — *G. soja* (Primorskii krai); 7-9 — *Triticum aestivum*, varieties Moskovskaya 39, Mironovskaya 808 and Omskaya 36, respectively.

The differentiation reflects the phenotypic features and the origin of plants. Particularly, the white awnless and beardless ears are characteristic for both Mironovskaya 808 and Omskaya 36 varieties (17), while Moskovskaya

39 has the white awn and beardless ears (<http://www.nemchinovka.ru/sorta/yar>). It should be noted that the winter variety Mironovskaya 808 has been obtained from spring soft wheat by means of group and mass selection of morphologically homogenous plants (18).

The obtained data evidence that polyloci genotyping by means of IRAP-PCR markers, when LTR SIRE-1 fragment is used as a primer, allows to differentiate reliably not only the monocots and dicots, but also the varieties..

More than 165 homological regions have been found for cattle in the course of our screening data of GenBank by BLASTn algorithm to show the homology to LTR SIRE-1 among animals.

We investigated Ayrshire, Red Estonian, Yakut and Black-and-White cattle from different livestock farms and also from a vivarium with regard to the IRAP-PCR markers detected due to amplification with LTR SIRE-1 as a primer in order to estimate possible use of plant transposons in farm animal polyloci genotyping and gene pool control. The patterns of DNA amplicons, from 13 to 16 fragments within 330-1470 bp, were observed (Fig. 4).

The highest polymorphism occurs among the middle-sized DNA fragments of 760-980 bp in the Black-and-White cattle from group I and in Ayrshire cows. The locus of 1470 bp fragment was the most polymorphic in other breeds. In Black-and-White cattle from group II and Yakut cattle no loci of 380 bp and 760 bp fragments were revealed, while they were found in other breeds (Table 3).

The amplicon patterns generated in PCR with LTR SIRE-1 as a primer across all the loci demonstrated no significant differences in Yakut, Red Estonian and Black-and-White cattle of group II with respect to the level of polymorphic loci and the $PIC_{average}$ values, being 14, 19 and 23 %, and 0.062, 0.066 and 0.094, respectively. The Ayrshire cattle population was the most heterogeneous ($PIC_{average} = 0.212$, $P = 56$ %) (Table 4).

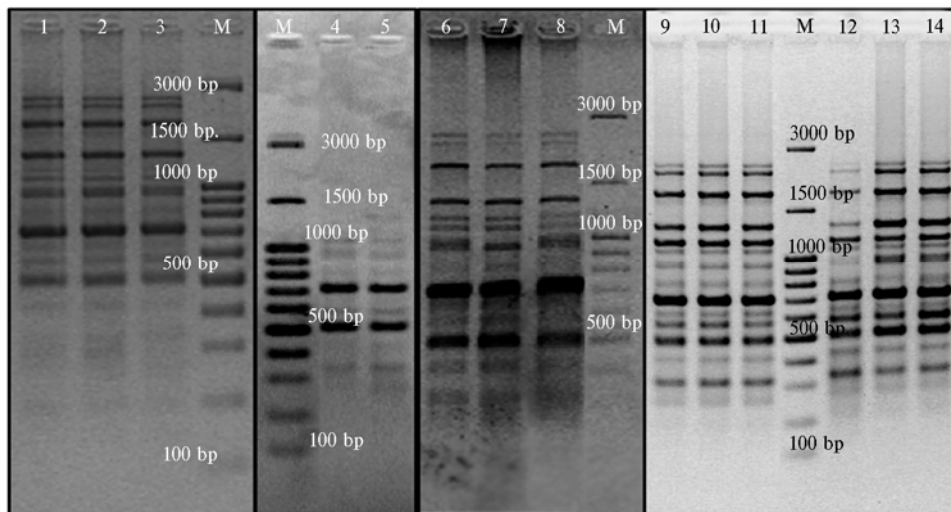


Fig. 4. Electrophoretic spectra of amplicons obtained in IRAP-PCR with DNA of *Bos taurus* cattle breeds using LTR SIRE-1 as a primer: 1-3 — Ayrshire cattle, 4, 5 — Red Estonian cattle, 6-8 — Black-and-White cattle (group I), 9-11 — Black-and-White cattle (group II), 12-14 — Yakut cattle; M — molecular marker GeneRuler™ 100 bp DNA Ladder Plus (MBI Fermentas, USA).

3. PIC_{locus} for amplicons obtained in IRAP-PCR with DNA of *Bos taurus* cattle breeds using LTR SIRE-1 as a primer

Fragment length, bp	Ayrshire	Black-and-White cattle		Yakut	Red Estonian
		group II	group I		
1470	0.325	0.486	0	0.488	0.452
1300	0.500	0	0.465	0	0
1200	0	0	0	0	0
1100	0	0	0	0	0
980	0.325	0.416	0.494	0	0.470
950	0.325	0	0	0	0
820	0.483	0	0.349	0	0
800	0.496	—	0	0.380	0
760	0.121	—	0.349	—	0
720	0	0	0	0	0
640	0	0.330	0.432	0	0
590	0	0	0	0	0
510	0	0	0	0	0
440	0	0	0	0	0
380	0.361	—	0.097	—	0.137
330	0.457	0	0	0	0

C o m m e n t s. IRAP-PCR — inter-retrotransposon amplified polymorphism PCR, PIC — polymorphic information content. Dashes indicate absence of the locus.

4. Polymorphism of the loci identified in IRAP-PCR with LTR SIRE-1 as a primer in *Bos taurus* cattle breeds

Breed, group	PIC _{loci}	P, %
Ayrshire	0.212	56
Black-and-White:		
group I	0.137	38
group II	0.094	23
Yakut	0.062	14
Red Estonian	0.066	19

C o m m e n t s. IRAP-PCR — inter-retrotransposon amplified polymorphism PCR, PIC — polymorphic information content, P — share of polymorphic loci.

A phylogenetic tree between the studied groups of animals was constructed based on the estimation of amplicon polymorphism (Fig. 5). Black-and-White cattle from group I and Ayrshire cattle were clustered the most tightly, while Black-and-White cattle from group II, Red Estonian cattle and Yakut cattle formed a segregated subcluster. It should be noted that the genetic distance between Black-and-White and Yakut cows was extremely small ($DN = 0.0076$). Two groups of Black-and-White cattle clearly differed according to their genetic characteristics (see Table 4). The animals from group I were more heterogeneous ($P = 38\%$, $PIC_{average} = 0.137$), probable due to different strategy of selection or

other artificial and natural factors of selection.

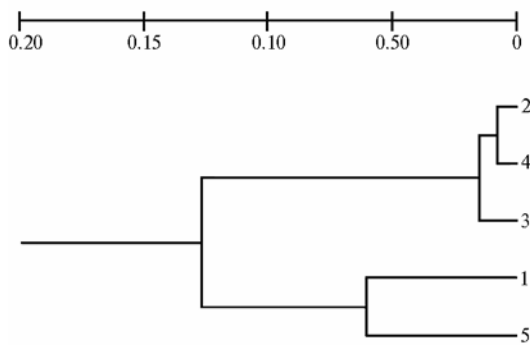


Fig. 5. Dendrogram of studied *Bos taurus* cattle based on the frequency of amplicons flanked by inverted regions of LTR SIRE-1: 1 — Ayrshire cattle, 2 — Black-and-White cattle from group I, 3 — Yakut cattle, 4 — Red Estonian cattle, 5 — Black-and-White cattle from group II.

Thus, the obtained data show that the fragments of mobile element, particularly LTR SIRE-1, could be used for investigations of genetic structure of the groups and

varieties of both dicotyledonous (*Glycine*) and monocotyledonous (*Triticum*) plants. Moreover, such polyloci genotyping of DNA fragments generated in IRAP-PCR with LTR SIRE-1 as a primer allows to identify the genetic characteristics of cattle breeds.

REFERENCES

1. Carvalho A., Guedes-Pinto H., Lima-Brito J. Intergenic spacer length variants in Old Portuguese bread wheat cultivar. *J. Genet.*, 2011, 90(2): 203-208 (doi: 10.1007/s12041-011-0051-4).
2. Corrado G., Piffanelli P., Caramante M., Coppola M., Rao R. SNP genotyping reveals genetic diversity between cultivated landraces and contemporary varieties of tomato. *BMC Genomics*, 2013, 14: 835-849 (doi: 10.1186/1471-2164-14-835).
3. Caballero A., García-Pereira M.J., Quesada H. Genomic distribution of AFLP markers relative to gene locations for different eukaryotic species. *BMC Genomics*, 2013, 14: 528-539 (doi: 10.1186/1471-2164-14-528).
4. Upadhyay P., Neeraja C.N., Kole C., Singh V.K. Population structure and genetic diversity in popular rice varieties of India as evidenced from SSR analysis. *Biochem. Genet.*, 2012, 50(9-10): 770-783 (doi: 10.1007/s10528-012-9519-z).
5. Ren D.X., Miao S.Y., Chen Y.L., Zou C.X., Liang X.W., Liu J.X. Genotyping of the k-casein and β -lactoglobulin genes in Chinese Holstein, Jersey and water buffalo by PCR-RFLP. *J. Genet.*, 2011, 90(1) (<http://www.ias.ac.in/jgenet/OnlineResou-rse/90/e1.pdf>).
6. Porto-Neto L.R., Sonstegard T.S., Liu G.E., Bickhart D.M., Da Silva M.V., Machado M.A., Utsunomiya Y.T., Garcia J.F., Gondro C., Van Tassel C.P. Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. *BMC Genomics*, 2013, 14: 876-888 (doi: 10.1186/1471-2164-14-876).
7. Nijman I.J., Otsen M., Verkaar E.L., De Ruijter C., Hanekamp E., Ochieng J.W., Shamshad S., Rege J.E., Hanotte O., Barwegen M.W., Sulawati T., Lenstra J.A. Hybridization of banteng (*Bos javanicus*) and zebu (*Bos indicus*) revealed by mitochondrial DNA, satellite DNA, AFLP and microsatellites. *Heredity (Edinb.)*, 2003, 90(1): 10-16.
8. Levin H.L., Moran J.V. Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet.*, 2012, 12(9): 615-627 (doi: 10.1038/nrg3030).
9. Pereira V. Insertion bias and purifying selection of retrotransposons in the Arabidopsis thaliana genome. *Genome Biol.*, 2004, 5(10) (<http://www.genomebiology.com/2004/5/10/R79>).
10. Gao L., McCarthy E.M., Ganko E.W., McDonald J.F. Evolutionary history of *Oryza sativa* LTR retrotransposons: a preliminary survey of the rice genome sequences. *BMC Genomics*, 2004, 5(1) (<http://www.biomedcentral.com/1471-2164/5/18>).
11. Bousios A., Kourmpetis Y.A., Pavlidis P., Minga E., Tsiftaris A., Darzentas N. The turbulent life of Sirevirus retrotransposons and the evolution of the maize genome: more than ten thousand elements tell the story. *Plant J.*, 2012, 69(3): 475-488 (doi: 10.1111/j.1365-3113X.2011.04806.x).
12. Bousios A., Darzentas N. Sirevirus LTR retrotransposons: phylogenetic misconceptions in the plant world. *Mobile DNA*, 2013, 4 (<http://www.mobilednajournal.com/content/4/1/9>).
13. Kalendar R., Grob T., Regina M., Suoniemi A., Shulman A. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. *Theor. Appl. Genet.*, 1999, 98: 704-711 (doi: 10.1007/s001220051124).
14. Botstein D., White R.L., Scolnik H., Davis R.W. Construction of a genetic map

- in man using restriction fragment length polymorphism. *Am. J. Hum. Genet.*, 1980, 32: 314-331.
15. Belyayev A., Kalendar R., Brodsky L., Nevo E., Shulman A.H., Raskina O. Transposable elements in a marginal plant population: temporal fluctuations provide new insights into genome evolution of wild diploid wheat. *Mobile DNA*, 2012, 1 (<http://www.mobilnjournal.com/content/1/1/6>).
 16. Remeslo V.N., Kolomatskii A.V. *Nauka i chelovechestvo: dostupno i tochno o glavnom v mirovoi nauke (Mezhdunarodnyi ezhegodnik)* [Science and the mankind: the main achievements (international annual issue)]. Moscow, 1980: 105-117.
 17. Kozub N.A., Sozinov I.A., Sozinov A.A. *Tsitologiya i genetika*, 2008, 3: 87-93.
 18. *Katalog nauchno-tekhnicheskoi produktcii. Vypusk 7-8: Nauchnoe obespechenie APK Sibiri* [Scientific basis of agroindustrial complex in Siberia. Issue 7-8]. Novosibirsk, 2009: 59.

Scientific events



IV INTERNATIONAL WORKSHOP «GENETICS FOR AGRICULTURE: GENOME BIOTECHNOLOGIES FOR AGROINDUSTRIAL COMPLEX»

(October 15 октября, 2014, Institute of Genetic and Cytology of NAS of Belarus, Minsk)

The aim of the workshop is to inform scientists and experts about recent achievements in the genetics of agricultural plants, acceleration of practical application of modern genetics in plant breed, development of the relationship between scientists and experts.

Contacts and information: <http://www.gens.by>, office@igc.bas-net.by