

UDC 635.655:631.527.8

doi: 10.15389/agrobiology.2017.5.905rus

doi: 10.15389/agrobiology.2017.5.905eng

## GENETIC SOURCES REQUIRED FOR SOYBEAN BREEDING IN THE CONTEXT OF NEW BIOTECHNOLOGIES

(review)

M.A. VISHNYAKOVA<sup>1</sup>, I.V. SEFEROVA<sup>1</sup>, M.G. SAMSONOVA<sup>2</sup>

<sup>1</sup>Federal Research Center the N.I. Vavilov All-Russian Institute of Plant Genetic Resources, Federal Agency of Scientific Organizations, 42-44, ul. Bol'shaya Morskaya, St. Petersburg, 190000 Russia, e-mail m.vishnyakova.vir@gmail.com (corresponding author);

<sup>2</sup>Peter the Great St. Petersburg Polytechnic University, 29, ul. Politechnicheskaya, St. Petersburg, 195251 Russia, e-mail m.g.samsonova@gmail.com

ORCID:

Vishnyakova M.A. orcid.org/0000-0003-2808-7745

Seferova I.V. orcid.org/0000-0003-3308-9198

The authors declare no conflict of interests

Supported by a subsidy from Federal Target Program (agreement No. 14.575.21.0136 of 26.09.2017)

Received June 11, 2017

### Abstract

Soybean is a strategic crop of multipurpose use. Production and consumption of soybeans are increasing year by year, with new uses appeared. Soybean can become one of the key plants in bioeconomics. Food, fodder, technical, medical and pharmaceutical use of soybeans is diversified and requires specialized varieties with the target traits. This poses new challenges for breeders and, accordingly, for holders of germplasm collections that supply source material for breeding. VIR soybean collection for many years serves as a genetic source for breeding. Based on long-term phenotyping, the accessions are systematized by a number of traits. Rapid development of new molecular technologies, e.g. marker-assisted selection (MAS) and genomic breeding are targeted to optimize both creation of new varieties and searching for the necessary genotypes. A number of agronomically important quantitative trait loci (QTL) have been found for soybean (Y. Xu, J.H. Crouch, 2008; D.C. Leite et al., 2016; Y. Ma et al., 2016; H. Liu et al., 2017), and putative candidate genes have been determined (E.Y. Hwang et al., 2014; J. Zhang et al., 2015; J. Zhang et al., 2016). This allows quick and targeted search for genotypes in germplasm collections, and necessitates relevant knowledge of the gene pool, i.e. the trait variability, the industrial uses, including the use of alternative values, etc. The purpose of this paper is an overview of the genetic diversity of VIR soybean collection in the context of modern breeding needs, in particular the creation of specialized varieties for target use, taking into account the crop studying and diversifying in the world, as well as developing new breeding technologies. It is shown that the VIR soybean collection contains genetic sources for high grain quality, i.e. high in protein and low in antinutritional substances, improved in oil and soy milk characteristics, etc. Breeding early matured varieties for all soybean producing regions based on relevant gene sources is urgent. For all the traits discussed, the paper gives the modern data on genetic control, genomic organization and mapping genes and QTL. It is concluded that the range of soybeans uses should be based on a diversity of specialized varieties with specified parameters for target use and different adaptive abilities.

Keywords: *Glycine max* (L.) Merr., soybean, VIR collection, genetic resources, initial material, QTL, genes, breeding, specific uses, grain quality, early maturation

Soybean *Glycine max* (L.) Merr. is a strategic crop for multipurpose use. The production of soybeans is steadily increasing in the world and in the Russian Federation. In 2016, the world soybean areas were 120.31 million hectares [1]. In the Russian Federation over the last 10 years, they have increased by almost 1,500,000 hectares (up to 2,228,485 hectares in 2016). Commercial crops have expanded, including regions for which soybeans were not previously traditional. For 10 years, the crop yield has also increased from 9.2 centner/ha in 2007 to 15.6 centner/ha in 2016 (Official Internet Portal of the Federal Service of State Statistics, <http://www.gks.ru/>). To a large extent, this is due to the success of breeding.

Currently, 210 soybean varieties are registered in the State Register of Selection Achievements admitted to use in the Russian Federation. Their significant share (32 %) comprises varieties of foreign origin [2]. Meanwhile, there are at least 43 institutions engaged in the selection of soybean in the Russian Federation. The geography of these institutions, like the soybean acreage, has a certain dynamics. Since the 1990s, most of the institutions working with soybeans are in the European part of Russia, while earlier production and selection of soybeans in our country were considered the prerogative of the Far East. An increasing variety of soil and climatic conditions makes specific demands on the source material in connection with the need to create adaptive varieties, as for each region, genotypes with a certain photoperiodic sensitivity, requirements for heat, moisture, tolerance to edaphic stressors of the terrain, etc. are needed.

As science develops, along with new production and processing capacities, new directions of soybean use arise. The soybean use is also diversified within its traditional range as a food, fodder and technical crop. The need for pharmaceutical and medical purposes is increasing. Soybean, as a raw material for the production of biofuels and organic fibers, can become one of the key vegetative objects in developing a bioeconomy. According to scientists and the US Department of Energy, biodiesel from soybeans is the only alternative fuel that meets all environmental requirements [3]. The production of fibers from soy protein has long been a reality. Fabrics from it are called textiles of the 21st century [4, 5]. All these innovations require the creation of new varieties that determine the quali-milk, okara and tofu-based products, which are increasingly recognized among the population, especially in southern Russia. In many countries of the world, interest in the use of soybean as a vegetable crop, which was popular in the ancient times in the East, is reviving, the direction so far little known in the Russian Federation. Soybean grain is also suitable for products with preventive and therapeutic properties in a number of diseases. Therefore, since 1990, soybean is classified as functional food [6, 7]. Along with the use of traditional soybean feed (meal, cake, fat soy flour, feed phosphatides, straw, chaff, green mass, etc.) for many types of farm animals and poultry, such feeds are increasingly used in unprecedentedly fast-growing industry of fish farming, as evidenced by the reports of the international conference 'Aquaculture of America 2017' held in early 2017 in Texas [8].

Herein, we have given an incomplete list of soybean use, but those already listed make us reconsider the traditional way of deriving new varieties from existing ones which are considered good in this or that region and often called universal. Selection should focus on specialized varieties with high quality and an increased level of target ingredients. All this requires a parental material with identified traits of interest specific for the intended purpose [9, 10].

However, perhaps the main reason to know the existing diversity of the gene pool is the radical change in breeding methods with the inevitable transition from traditional to genomic technologies on the threshold of which we are standing. Soybeans have been subjected to transgenic manipulation for a long time, but in this review we will focus on those molecular technologies that generate new conceptual level for searching genotypes of interest in germplasm collections. Soybean has already been involved in MAS, based on description and mapping many quantitative trait loci (QTL) in *Glycine max* genome which determine, in particular, grain quality, resistance to diseases, tolerance to some abiotic stressors, vitamin E content, seed weight, etc. [11-14]. The genomic selection of soybeans has been laid [15]. To date, whole-genome sequencing revealed candidate genes determining seed protein and oil content [16], seed size [17], and made it possible to find several new loci and clarify linkage groups of

known ones associated with early flowering, early maturity and plant height in early ripening soybeans [18].

The gene pool stored in the VIR collection was systematized basically on phenotypic characters obtained in studding the accessions in different soil and climatic conditions and in laboratory tests. However, the task for a curator is to know the world trends and priorities, both recent and future, in use and selection of a crop. It is necessary for seeking parental forms and donors to be involved in deriving varieties for the intended use. The possibility of such a search is based on the wide variability of traits revealed in the VIR soybean collection, and knowledge about the differentiation of the gene pool on these traits.

The purpose of this paper, in the context of current needs in varieties for target use, is to overview genetic diversity of the VIR soybean collection, which for many years served as a source of material for breeding, with a focus on the world's achievements in modern breeding technologies, soybean studying and diversified use.

The quality of raw materials is the main requirement for agricultural crops, regardless of their use. For soybean, it is primarily the grain quality, i.e. the composition and content of protein and oil, antinutrients. In VIR collection, soybeans have traditionally been evaluated for protein, oil and trypsin inhibitors in seeds. Very often breeders want to find in the gene pool sources for breeding a variety with high levels of both protein and oil, using their total value. It was found that over the last 30-40 years this value increased from 49.7 to 66.3 %, mainly due to a decrease in the proportion of the seed coat, and practically reached the biological limits of the crop [9]. However, the accumulation of protein and oil in the soybean seed usually reveals a strong negative relationship that cannot be broken [19]. The negative correlation between the traits can vary from  $r = -0.25$  to  $r = -0.93$  [20] at high heritability of 0.89-0.93 [21]. Moreover, the total plant productivity often negatively correlates with the protein content, although this relationship is weaker than that between protein and oil levels [9, 22]. It is shown, however, that there are no permanent metabolic barriers between these traits. The amount of seed proteins can be increased by backcrossing. Thus, a 4.7 % increase required eight cycles of recurrent selection [23]. It was also reported about a 5.6-6.9 % increase in the portion of proteins without reducing seed oil level and grain yield under univariate intravarietal selection [24].

It is known that modification variability of protein and oil levels in seeds is high, and oil content is much less influenced by environment conditions than protein content [25]. According to the assessment of biochemical indicators of 936 soybean samples (Kuban branch of the Vavilov All-Russian Institute of Plant Genetic Resources — VIR, steppe zone of the Kuban plain, Krasnodar Krai), the seed protein content ranged from 23.5 to 48.0 %, and the oil content ranged from 13.8 to 27.2 %. The relationship between the protein and oil content was negative ( $r = -0.57$ ). In this, 24 donors of high protein content (45.1-50.0 %) and 111 donors of high oil content (24.1-27.1 %) were identified. A total of 12 samples were simultaneously high in protein and oil (from 20.1 to 22.0 %), and 22 samples of different ripeness groups showed high oil content and high seed production, and all samples with high protein content had an average or low productivity [26]. When assessing 189 samples in the Adler branch of the VIR (Black Sea coast, Krasnodar Krai), the seed protein content was higher (34.4-51.0 %), and the oil content was lower than in the conditions of Kuban (14.7-24.8 %). The relationship between the protein and oil levels was also negative ( $r = -0.63$ ). An apparent correlation between seed productivity and protein and oil content was not found [27, 28].

Since the seeds protein and oil level is resultant of interaction of multiple

genes and loci, influenced by the environment, traditional selection for this trait in a number of world breeding centers is accompanied by an analysis of linkage groups to identify QTLs that determine oil and proteins in different genotypes and under different environmental conditions [29-31]. Several QTLs affecting seed protein content have been detected. Loci located on the Gm20 chromosome have been identified in many mapping populations. The genes present in these loci are described in the offspring from crossing NILs (near-isogenic lines) with the contrast trait values [32]. In particular, the genes encoding the potential protein regulator of Mov34-1 family, heat shock protein Hsp22.5 and AT synthetase were identified. QTLs associated with protein synthesis and related traits in combination with genomic analysis may contribute to the rapid selection of soybean QTLs that are significant for protein accumulation and the identification of candidate genes regulating seed protein content [33].

As noted, differentiation that determines the variety character exists within a certain use, for example, for food. A variety of products from soybean is difficult to list. This is oil, soy milk and its derivatives, flour, protein concentrates, isolates, textures, lecithin, etc. However, the relevant specialized varieties are still often produced mainly on the basis of any highly productive forms. In particular, for the varieties suitable for soy milk production, the breeders use large-seed samples with a light seed coat, light hilum and high protein content. These are the characteristics of the soybean variety Donskaya, which is commonly referred to as “milky” (All-Russian Research Institute of Grain Production), and Lakta (All-Russian Research Institute of Oil Crops). Such varieties should also have low trypsin-inhibiting activity, good extractability of solids and improved taste. The assessment of the above traits is rather time-consuming. It seems more constructive to screen the collection to reveal the soybean gene pool polymorphism of the level and proportion of glycinin (11S) and  $\beta$ -conglycinin (7S) protein fractions which constitute 70 % of the total soybean protein. The need for such assessment is determined by the fact that soy milk production necessitates soybean varieties rich in conglycinin as an emulsifier. Varieties with an alternative trait, in turn, should be suitable for protein texturates produced from soybean seeds [34]. Thus, tofu should be manufactured only from seeds with the prevalence of 11S fractions [35]. Soybean varieties are significantly polymorphic on content of these fractions [36-39], which allowed conclusion about the possibility of selection for the predominance of a fraction or even its subunits without reducing the total amount of protein. Such data can also be useful for the tactics of fodder production, since varieties containing more conglycinin are more suitable to fattening pigs, while cattle need more glycinin fractions [40]. The variability of this trait in gene pool may be assessed using molecular screening, since the families of genes encoding for these protein fractions are well known, their genomic organization has been studied and QTLs have been found and mapped [41-43].

When creating oil-bearing varieties, the characterization of the parental forms for the quality of the oil is also not always performed. In the VIR collection, the sources of valuable traits are found and corresponding genotypes are identified to optimize the oilseed selection. These valuable forms are those with high oil content (above 25 %), high levels of linoleic acid (50-52 %) and polyunsaturated acids, with *fan* gene recessive allele, determining the content of seed linolenic acid, and *fap1* gene recessive allele, determining the seed palmitic acid level [43]. The accumulation of oil in soybean seeds is under multiple genes control. Thus, about 110 corresponding QTLs have been mapped. Alleles of 14 genes, which control significant differences in fatty acids content, have been identified in the regions containing QTLs, and the greater number of “high oil”

alleles, the higher oil content is [44]. Interestingly, in samples adapted to high latitudes, the seed oil level is higher than that for those growing in low latitudes, which also correlated with the number of “high oil” alleles. This indicates the additivity of the action of genes controlling the synthesis and quality of oil in soybean. The study of the 10 most widely distributed varieties in China showed that they do not carry all the identified alleles for high oil content, which gives prospects for pyramiding additional alleles.

Modern technologies of improving storability, taste and nutritional qualities of soy oil suggest regulating proportion of certain fatty acids. Therefore, screening the gene pool polymorphism for fatty acids and tocopherols (vitamin E) content is constructive in searching for the parental forms for soybean breeding. In soybean seeds, there are  $\gamma$ -,  $\delta$ - and  $\alpha$ -tocopherols, amounting 60-70, 20-25 and less than 10 % of the total seed tocopherols, respectively [45-46]. It is  $\alpha$ -tocopherol that is the most active as vitamin E in mammals. To date, a QTL has been identified for high  $\alpha$ -tocopherol content, a mapping of this region has been carried out and  $\gamma$ -*TMT3* gene was identified which appears to be responsible for increased  $\alpha$ -tocopherol accumulation [47]. In the VIR collection, the search for genotypes with high  $\alpha$ -tocopherol based on identification of  $\gamma$ -*TMT3* and QTL is a matter of a near future.

The quality of soybean oil is largely depends on concentration of lipoxygenase (coenzyme Q). Lipoxygenase catalyzes synthesis of peroxide products (hexanal, etc.) from unsaturated fatty acids, which determine the “bean” flavor of raw soybean. Soybean seeds are considered as the richest natural source of lipoxygenase [48]. Three lipoxygenase isozymes are described of which lipoxygenase 2 mainly contributes to flavoring. Low content or inactivation of at least one isoenzyme eliminates bitter taste, improving the organoleptic characteristics of soy products. Forms with a low amount of lipoxygenase are also used for selection of edamame, the vegetable soybean which is very popular in Japan and the USA [49]. The presence of lipoxygenase fractions is controlled by the dominant *Lx* gene, and their low content or absence is controlled by the recessive *lx* allele. In the VIR collection, there are samples with *lx* recessive alleles; adding them to the genome of commercial soybean genotypes will reduce lipoxygenase activity to get products that are less prone to undesirable oxidation. To date, these genes have been sequenced, and molecular markers can be used to speed up the search for the desired genotypes [50].

Sugars are also important characters of soybean grain quality. Water soluble sugars of soybean are mainly sucrose disaccharide and stachyose and raffinose trisaccharides. In most varieties, the content of sucrose is 4.0-4.5 %, reaching 7-11 % of dry matter in some. The increased content of sucrose is desirable both for vegetable and feed use [51, 52].

Till now, targeted selection has been carried out for concomitants of soybean oil (phospholipids, tocopherols), biologically active substances (oligosaccharides, isoflavones) and for improving organoleptic parameters [9], so screening of the VIR soybean collection by these indicators has not been done yet.

Increasing the nutritional and fodder value of agricultural crops, endowing food varieties with the qualities of highly functional products is the prerogative of selection technologies, referred to as “bio-fortification”. The creation of varieties containing more vitamins, bioactive substances, antioxidants, oils with an optimized fatty acid composition, a reduced content of antinutrients, etc. is aimed at combating the so-called hidden hunger. This is facilitated by high variability in the content of macro and micronutrients in plants. Unfortunately, in Russia this field has not yet found a proper development and the gene pool has not been screened for these characteristics. However, the impressive results of legume biofortification

obtained abroad (including traditional methods of selection) are known. As for soybean, the most famous example of bio-fortification is high-oleic oil, and also oil with a reduced content of saturated fatty acids. DuPont Pioneer obtained a variety for high-oleic food oil (80-85 %) through transgenesis. Varieties for technical purposes also have been created with containing up to 50 % linolenic acid in oil in contrast to 3 % characteristic of food varieties. Usually, the proportion of oleic acid in the soybean oil does not exceed 23 %. The high-oleic variety does not differ from the control ones in total proteins, oil, hydrocarbons, coarse fiber, certain amino acids, vitamins, ashes, minerals, trypsin inhibitors, oligosugars [53]. By traditional selection, low palmitic acid lines were created [54, 55]. In Japan, after chemical mutagenesis and x-ray treatment of seeds, lines with a high content of saturated fatty acids were revealed to increase the oil stability and to produce margarines and other solid and semi-solid fats. In total, the proportion of palmitic and stearic acid in soybean is usually 5 %, reaching 38 % [55] for new lines. A form with increased vitamin E content, more resistant to herbicides, was obtained by physical mutagenesis in the United States [56].

Molecular study revealed 35 QTLs associated with accumulation of Ca, Mg, Fe, Zn and P in soybean seeds. It creates the prerequisites for bio-fortification by MAS [57]. In addition, the researchers explore possibilities to improve utilization of these elements from soil by plants. For example, the QTL qPE8 was identified on the GM08 chromosome, which also contains gene *GmACP1* encoding the acidic phosphatase, a putative candidate gene for high phosphorus utilization. Overexpression of this gene in soybean root hairs leads to a 2.3-fold increase in acid phosphatase activity and an 11-20 % increase in the efficiency of phosphorus utilization compared to the control [58].

Our brief overview of successes in soybean biofortification is done in order to emphasize the need to study the soybean gene pool, i.e. variability of traits and the branches of industry in which they can be used, including alternative trait manifestations. In this, the variability of proteinase inhibitors in soybean seeds can be seen as an example. The study of the VIR collection accessions carried out earlier showed high activity of these antinutrients in the samples of cultivated and Ussuri soybean, and their low level (several times less) in wild Australian species [59, 60]. These species, constituting the tertiary gene pool, were often considered as agronomic potential for crosses with cultivated soybeans [61]. However, presently, the samples with an alternative (high) activity of trypsin inhibitors may be of great importance, since the participation of protease inhibitors in plant protection from pests and diseases has been shown, and most importantly, anticarcinogenic and radioprotective functions of protease inhibitors have been proven [62, 63]. Both of these properties are most pronounced in the Bowman-Birk protease inhibitor. In biochemical studies of the polymorphism of soybean varieties from VIR collection with regard to Bowman-Birk protease inhibitor and Kunitz-type soybean trypsin inhibitor, using a developed methodological approach to mass screening, the sources of raw material for the pharmaceutical industry have been identified. In addition, it has been shown that it is more reliable to identify soybean varieties by proteinase inhibitor electrophoretic spectra than by reserve proteins or DNA fragments detected by RAPD (random amplification of polymorphic DNA) [64].

A key role in expanding the geography of soybean cultivation plays high-yield varieties with a shorter vegetation period. The reaching of ripeness is relevant for all regions of soybean cultivation in Russia. In the southern regions, the early-ripening soybean varieties will optimize time of harvesting. Early ripeness can also ensure the growth of soybean in the more humid northern areas. This problem, posed by N.I. Vavilov in the early 1930s [65],

becomes especially relevant now due to warming and frequent droughts.

Expansion of soybean crops to the north is a worldwide trend. The main soybean country in the world, the United States, has long come to the need to create fast-ripening varieties, which led to the emergence of a new classification of ripening groups. If in the 1980s there were 10 such groups and the classification began with MG1 (maturity group 1), the appearance of early maturing forms led to an increase in the number of ripeness groups to 13 and the appearance of groups MG0, MG00, MG000 [66]. In recent years, due to expansion of soybean crops to the high-latitude regions of China, MG0000 category with the fastest ripening to date has been added to this classification [67].

Crop adaptation to the cultivation zones is ensured by the time of flowering and maturation. The molecular genetic basis of precocity is not yet fully understood, but 10 loci determining the periods of flowering and maturation and designated as *E* loci have been identified by classical methods. Of these, four (*E1*, *E3*, *E4* and *E7*) are classified as photoperiodic genes. It was suggested that allelic variants and combinations of these genes determine the variability of maturation period [68]. The dominant alleles *E6*, *E9* and *J* lead to early blooming, while the dominant alleles of other genes, on the contrary, slow the flowering and the onset of ripeness [69]. The *E1*, *E2*, *E3*, *E4* and *E9* loci were categorized as QTLs. The casual genes of these loci are, respectively, the genes of the DNA-binding protein *Q3*, *GIGANTEA*, *Phytochrome A3*, *Phytochrome A2* and *GMFT2a* [69-73]. Casual genes of other loci have not been identified. In the field, the locus *E1* [74] exerts the greatest influence on the flowering time. Thus, molecular screening on the basis of early maturity is still a matter of the future. However, we have data of phenotypic assessment reflecting soybean ripeness periods in different Russian regions, including the northernmost point of the world's soybean cultivation (VIR, Leningrad Province). Breeders are given varieties corresponding to one or another region with regard of ripening groups, including ultra-ripe genotypes of the "northern ecotype", which, in our opinion, corresponds to the American ripeness groups MG00-MG000 [75-77].

So, at the present development of molecular selection, the diversity of soybean genetic resources from the VIR collection should be used more efficiently. The variability of the traits in gene pool and the molecular methods that optimize identification of gene pool differentiation by the target properties contribute to purposeful and targeted selection of parental forms for breeding. The range of soybean application areas should be based on the varieties with specified parameters for different uses and different adaptive capabilities.

## REFERENCES

1. WAP 09-17. September 2017. Circular Series. USDA, Foreign Agr. Service. Available <https://apps.fas.usda.gov/psdonline/circulars/production.pdf>. Accessed August 22, 2017.
2. Gosudarstvennyi reestr selektsionnykh dostizhenii, dopushchennykh k ispol'zovaniyu [State Register of Breeding Achievements]. Moscow, 2017 (in Russ.).
3. Hill J., Nelson E., Tilman D., Polasky S., Tiffany D. Environmental, economic, and energetic costs, and benefits of biodiesel and ethanol biofuels. *PNAS USA*, 2006, 103: 11206-11210 (doi: 10.1073/pnas.0604600103).
4. Yi-you L. The soybean protein fibre — a healthy and comfortable fibre for the 21<sup>st</sup> century. *Fibres and Textiles in Eastern Europe*, 2004, 12(2/46): 8-9.
5. Brooks M.M. Soybean protein fibres — past, present and future. In: *Woodhead publishing series in textiles. V. 47. Biodegradable and sustainable fibres*. Cambridge, 2005: 398-440 (doi: 10.1533/9781845690991.398).
6. Petibskaya V.S. *Nauchno-tehnicheskii byulleten' VNIIMK*, 2002, 126: 76-83 (in Russ.).
7. Nekrasova T.E. *Masla i zhiry*, 2005, 11(57): 2-4 (in Russ.).
8. Forging new frontiers. Aquaculture America. *Int. Conf. and Exposition*. Texas, USA, 2017. Available <http://www.aquafeed.com>. Accessed August 22, 2017.
9. Petibskaya V.S., Kucherenko L.A., Zelentsov S.V. *Nauchno-tehnicheskii byulleten'*

- en' *VNIIMK*, 2006, 2(135): 115-116 (in Russ.).
10. Zaitsev N.I., Bochkarev N.I., Zelentsov S.V. *Maslichnye kul'tury*, 2016, 2(166): 3-11 (in Russ.).
  11. Xu Y., Crouch J.H. Marker-assisted selection in plant breeding: from publications to practice. *Crop Sci.*, 2008, 48: 391-407 (doi: 10.2135/cropsci2007.04.0191).
  12. Leite D.C., Pinheiro J.B., Campos J.B., Di Mauro A.O., Unêda-Trevisoli S.H. QTL mapping of soybean oil content for marker-assisted selection in plant breeding program. *Genet. Mol. Res.*, 2016, 15(1): gmr.15017 (doi: 10.4238/gmr.15017685).
  13. Ma Y., Kan G., Zhang X., Wang Y., Zhang W., Du H., Yu D. Quantitative Trait Loci (QTL) mapping for glycinin and  $\beta$ -conglycinin contents in soybean (*Glycine max* L. Merr.). *J. Agric. Food Chem.*, 2016, 64(17): 3473-3483 (doi: 10.1021/acs.jafc.6b00167).
  14. Liu H., Cao G., Han Y., Jiang Z., Zhao H., Li W. Identification of the QTL underlying the vitamin E content of soybean seeds. *Plant Breeding*, 2017, 136(2): 147-154 (doi: 10.1111/pbr.12454).
  15. Heffner E.L., Sorrells M.E., Jannink J.-L. Genomic selection for crop improvement. *Crop Sci.*, 2009, 49: 1-12 (doi: 10.2135/cropsci2008.08.0512).
  16. Hwang E.Y., Song Q.J., Jia G.F., Specht J.E., Hyten D.L., Costa J., Cregan P.B. A genome-wide association study of seed protein and oil content in soybean. *BMC Genomics*, 2014, 15: 1 (doi: 10.1186/1471-2164-15-1.10.1186/1471-2164-15-11).
  17. Zhang J., Song Q., Cregan P., Jiang G.-L. Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theor. Appl. Genet.*, 2016; 129: 117-130 (doi: 10.1007/s00122-015-2614-x).
  18. Zhang J., Song Q., Cregan P.B., Nelson R.L., Wang X., Wu J., Jiang G.L. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics*, 2015, 16: 217 (doi: 10.1186/s12864-015-1441-4).
  19. Wilson R.F. Seed composition. In: *Soybeans: improvement, production and uses*. American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America. Madison, Wisconsin, USA, 2004: 621-677.
  20. Korsakov N.I., Myakushko Yu.P. *Trudy po prikladnoi botanike, genetike i seleksii*, 1976, 57: 13-19 (in Russ.).
  21. Stobaugh B., Florez-Palacios L., Chen P., Orazaly M. Agronomic evaluation of high-protein and high-oil soybean genotypes for specialty markets. *Journal of Crop Improvement*, 2017, 31(2): 247-260 (doi: 10.1080/15427528.2017.1287807).
  22. Chung J., Babka H.L., Graef G.L., Staswick P.E., Lee D.J., Cregan P.B., Shoemaker R.C., Specht J.E. The seed protein, oil and yield QTL on soybean linkage group I. *Crop Sci.*, 2003, 43(3): 1053-1067 (doi: 10.2135/cropsci2003.1053).
  23. Wilcox J.R. Increasing seed protein in soybean with eight cycles of recurrent selection. *Crop Sci.*, 1998, 38(6): 1536-1540 (doi: 10.2135/cropsci1998.0011183X003800060021x).
  24. Myakushko Yu.P. *Seleksiya i semenovodstvo soi na Severnom Kavkaze. Avtoreferat doktorskoi dissertatsii* [Soya: breeding and seed production]. Leningrad, 1976 (in Russ.).
  25. Bellaloui N., Reddy K.N., Bruns A., Gillen A. M., Mengistu A., Zobiolle L.H.S., Fisher D.K., Abbas H.K., Zablutowicz R., Kremer R.J. Soybean seed composition and quality: interactions of environment, genotype, and management practices. In: *Soybeans: cultivation, uses and nutrition*. J. Maxwell (ed.). Nova Science Publishers, 2011: 1-42.
  26. Seferova I.V., Nekrasov A.Yu., Silaeva O.I., Kiyashko N.I., Teter Z.Yu., Kiva T.I., Nikishkina M.A. *Soya. Iskhodnyi material dlya seleksii soi v Krasnodarskom krae. Katalog mirovoi kollektzii VIR* [Soya. Genetic resources for use in breeding in the Krasnodra Territory]. St. Petersburg, 2008, vypusk 782 (in Russ.).
  27. Seferova I.V., Boiko A.P., Shelenga T.V., Sholukhova T.A. *Trudy po prikladnoi botanike, genetike i seleksii*, 2014, 175(3): 34-41 (in Russ.).
  28. Boiko A.P., Seferova I.V., Shelenga T.V., Sholukhova T.A., Andreeva T.M. *Soya. Iskhodnyi material dlya seleksii v yuzhnykh regionakh Rossiiskoi Federatsii. Katalog mirovoi kollektzii VIR*. St. Petersburg, 2014, vypusk 817 (in Russ.).
  29. Jun T.H., Van K., Kim M.Y., Lee S.H., Walker D.R. Association analysis using SSR markers to find QTL for seed protein content in soybean. *Euphytica*, 2008, 162: 179-191 (doi: 10.1007/s10681-007-9491-6).
  30. Phansak P., Soonsuwon W., Hyten D.L., Song Q., Cregan P.G., Graef G.L., Specht J.E. Multi-population selective genotyping to identify soybean [*Glycine max* (L.) Merr.] seed protein and oil QTLs. G3. *Genes, Genomes, Genetics*, 2016, 6(6): 1635-1648 (doi: 10.1534/g3.116.027656).
  31. Cao Y., Li S., Wang Z., Chang F., Kong J., Gai J., Zhao T. Identification of major quantitative trait loci for seed oil content in soybeans by combining linkage and genome-wide association mapping. *Front. Plant Sci.*, 2017, 8: 1222 (doi: 10.3389/fpls.2017.01222).
  32. Bolton Y.T., Joseph B., Cannon S.B., Graham M.A., Diers B.W., Farmer A.D., May G.D., Muehlbauer G.J., Specht J.E., Tu Z.J., Weeks N., Xu W.W., Shoemaker R.C., Vance C.P. Complementary genetic and genomic approaches help to characterize the linkage group I seed protein QTL in soybean. *BMC Plant Biol.*, 2010, 10: 41 (doi: 10.1186/1471-2229-10-41).



33. Patil G., Mian R., Vuong T., Pantalone V., Song Q., Chen P., Shannon G.J., Carter T.C., Nguyen H.T. Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. *Theor. Appl. Genet.*, 2017, 130(10): 1975-1991 (doi: 10.1007/s00122-017-2955-8).
34. Wolf W.J. Soybean proteins: their functional, chemical, and physical properties. *J. Agr. Food Chem.*, 1970, 18(6): 969-976 (doi: 10.1021/jf60172a025).
35. Onodera Y., Ono T., Nakasato K., Toda K. Homogeneity and microstructure of tofu depends on 11S/7S globulin ratio in soymilk and coagulant concentration. *Food Sci. Technol. Res.*, 2009, 15(3): 265-274 (doi: 10.3136/fstr.15.265).
36. Adamovskaya V.G., Molodchenkova O.O., Sichkar' V.I., Kartuzova T.V., Bezkravnaya L.YA., Lavrova G.D. *Visnik Urains'kogo tovaristva genetikiv i seleksioneriv*, 2016, 14(2): 163-167 (in Russ.).
37. Khatib K.A., Herald T.J., Aramouni F.M., MacRithie F., Schapaugh W.T. Characterization and functional properties of soy  $\beta$ -conglycinin and glycinin of selected genotypes. *J. Food Sci.*, 2002, 67: 2923-2929 (doi: 10.1111/j.1365-2621.2002.tb08839.x).
38. Kyoko S., Watanabe T. Differences in functional properties of 7S and 11S soybean proteins. *Journal of Texture Studies*, 1978, 9(1-2): 135-157 (doi: 10.1111/j.1745-4603.1978.tb01298.x).
39. Li J., Matsumoto S., Nakamura A., Maeda H., Matsumura Y. Characterization and functional properties of sub-fractions of soluble soybean. *Biosci. Biotechnol. Biochem.*, 2009, 73: 2568-2575 (doi: 10.1271/bbb.70799).
40. Yaklich R.W.  $\beta$ -Conglycinin and glycinin in high-protein soybean seeds. *J. Agric. Food Chem.*, 2001, 49(2): 729-735 (doi: 10.1021/jf001110s).
41. Beilinson V., Chen Z., Shoemaker R.C., Fischer R.L., Goldberg R.B., Nielsen N.C. Genomic organization of glycinin genes in soybean. *Theor. Appl. Genet.*, 2002, 104: 1132-1140 (doi: 10.1007/s00122-002-0884-6).
42. Li C., Zhang Y.-M. Molecular evolution of glycinin and  $\beta$ -conglycinin gene families in soybean (*Glycine max* L. Merr.). *Heredity*, 2011, 106(4): 633-641 (doi: 10.1038/hdy.2010.97).
43. Vishnyakova M.A., Seferova I.V. V knige: *Identifitsirovannyi genofond v kollektzii VIR i ego ispol'zovanie v seleksii*. St. Petersburg, 2005: 841-850 (in Russ.).
44. Fang C., Ma Y., Wu S., Liu Z., Wang Z., Yang R., Hu G., Zhou Z., Yu H., Zhang M., Pan Y., Zhou G., Ren H., Du W., Yan H., Wang Y., Han D., Shen Y., Liu S., Liu T., Zhang J., Qin H., Yuan J., Yuan X., Kong F., Liu B., Li J., Zhang Z., Wang G., Zhu B., Tian Z. Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.*, 2017, 18: 161 (doi: 10.1186/s13059-017-1289-9).
45. Van Eenennaam A.L., Lincoln K., Durrett T.P., Valentin H.E., Shewmaker C.K., Thorne G.M., Jiang J., Baszis S.R., Levering C.K., Aasen E.D., Hao M., Stein J.C., Norris S.R., Last R.L. Engineering vitamin E content: from *Arabidopsis* mutant to soy oil. *Plant Cell*, 2003, 15: 3007-3019 (doi: 10.1105/tpc.015875).
46. Ujiie A., Yamada T., Fujimoto K., Endo Y., Kitamura K. Identification of soybean varieties with high  $\alpha$ -tocopherol content. *Breed. Sci.*, 2005, 55: 123-125 (doi: 10.1270/jsbbs.55.123).
47. Dwiyantri M.S., Yamada T., Sato M., Abe J., Kitamura K. Genetic variation of  $\gamma$ -tocopherol methyltransferase gene contributes to elevated  $\alpha$ -tocopherol content in soybean seeds. *BMC Plant Biol.*, 2011, 11: 152 (doi: 10.1186/1471-2229-11-152).
48. Axelrod B., Cheesbrough T.M., Laakso S. Lipoxygenase from soybeans. *Methods Enzymol.*, 1981, 71: 441-451 (doi: 10.1016/0076-6879(81)71055-3).
49. Rackis J.J., Hoing D.H., Sessa D.S., Moser H.A. Lipoxygenase and peroxidase activities of soybeans as related to flavor profile during maturation. *Cereal Chemistry*, 1972, 49: 586-597.
50. Reinprecht Y., Luk-Labey S.Y., Yu K., Rajcan I., Ablett G.R., Peter Pauls K. Molecular basis of seed lipoxygenase null traits in soybean line OX948. *Theor. Appl. Genet.*, 2011, 122(7): 1247-1264 (doi: 10.1007/s00122-011-1528-5).
51. Kumar V., Rani A., Goyal L., Dixit A.K., Manjaya J.G., Dev J., Swamy M. Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. *J. Agric. Food Chem.*, 2010, 58(8): 5081-5085 (doi: 10.1021/jf903141s).
52. Santana A.C., Carrao-Panizzi M.C., Mandarino J.M.G., Leite R.S., Silva J.B., Ida E.I. Effect of harvest at different times of day on the physical and chemical characteristics of vegetable-type soybean. *Cienc. Tecnol. Aliment.*, 2012, 32(2): 351-356 (doi: 10.1590/S0101-20612012005000044).
53. Cahoon E.B. Genetic enhancement of soybean oil for industrial uses: prospects and challenges. *AgBioForum*, 2003, 6(1-2): 11-13.
54. Burton J., Wilson R., Brim C. Registration of N79-2077-12 and N87-2122-4, two soybean germplasm lines with reduced palmitic acid in seed oil. *Crop Sci.*, 1994, 34: 313 (doi: 10.2135/cropsci1994.0011183X003400010080x).
55. Rahman S.M., Anai T., Kinoshita T., Takagi Y. A novel soybean germplasm with elevated saturated fatty acids. *Crop Sci.*, 2003, 43: 527-531 (doi: 10.2135/cropsci2003.5270).
56. Stacey M.G., Cahoon R.E., Nguyen H.T., Cui Y., Sato S., Nguyen C.T., Phoka N.,

- Clark K.M., Liang Y., Forrester J., Batek J., Do P.T., Sleper D.A., Clemente T.E., Cahoon E.B., Stacey G. Identification of homogentisate dioxygenase as a target for vitamin E biofortification in oilseeds. *Plant Physiol.*, 2016, 172(3): 1506-1518 (doi: 10.1104/pp.16.00941).
57. Ning L., Sun P., Wang Q., Ma D., Hu Z., Zhang D., Zhang G., Cheng H., Yu D. Genetic architecture of biofortification traits in soybean (*Glycine max* L. Merr.) revealed through association analysis and linkage mapping. *Euphytica*, 2015, 204: 353-369 (doi: 10.1007/s10681-014-1340).
  58. Zhang D., Song H., Cheng H., Hao D., Wang H., Kan G., Jin H., Yu D. The acid phosphatase-encoding gene *GmACP1* contributes to soybean tolerance to low-phosphorus stress. *PLoS Genet.*, 2014, 10: 1004061 (doi: 10.1371/journal.pgen.1004061).
  59. Benken I.I., Tomilina T.B. *Nauchno-tekhnicheskii byulleten' VIR*, 1985, 149: 3-10 (in Russ.).
  60. Benken I.I., Nikishkina M.A., Shchelko L.G., Serova T.S. *Trudy po prikladnoi botanike, genetike i seleksii*, 1997, 152: 129-133 (in Russ.).
  61. Singh R.J., Hymowitz T. Exploitation of the wild perennial *Glycine* species for improving the soybean. In: *Harnessing the soy potential for health and wealth*. The Soybean Processors Association of India, 2001: 58-61.
  62. Yavelow J., Finlay T.H., Kennedy A.R., Troll W. Bowman-Birk soybean protease inhibitor as an anticarcinogen. *Cancer Res.*, 1983, 43(5 Suppl): 2454-2459.
  63. Kobayashi H. Prevention of cancer and inflammation by soybean protease inhibitors. *Front. Biosci.*, 2013, 5: 966-73 (doi: 10.2741/E676).
  64. Komissarova Yu.V. *Geterogennost' i polimorfizm ingibitorov proteinaz soi i gorokha*. Avtoreferat kandidatskaya dissertatsiya [Heterogeneity and polymorphism of soybean and pea proteins. PhD Thesis]. St. Petersburg, 1998 (in Russ.).
  65. Vavilov N.I. V knige: *Trudy Noyabr'skoi sessii Akademii nauk SSSR 25-30/XI 1931 g* [Proc. November's session of Academy of Sciences of the USSR, 25-30/XI 1931]. Leningrad, 1932: 250-264 (in Russ.).
  66. Fehr W.R. *Principles of cultivar development. V. 1. Theory and technique*. Macmillan, NY, 1991.
  67. Jia H., Jiang B., Wu C., Lu W., Hou W., Sun S., Yan H., Han T. Maturity group classification and maturity locus genotyping of early-maturing soybean varieties from high-latitude cold regions. *PLoS ONE*, 2014, 9(4): e94139 (doi: 10.1371/journal.pone.0094139).
  68. Jiang B., Nan H., Gao Y., Tang L., Yue Y., Lu S., Ma L., Cao D., Sun S., Wang J., Wu C., Yuan X., Hou W., Kong F., Han T., Liu B. Allelic combinations of soybean maturity loci *E1*, *E2*, *E3* and *E4* result in diversity of maturity and adaptation to different latitudes. *PLoS ONE*, 2014, 9(8): e106042 (doi: 10.1371/journal.pone.0106042).
  69. Zhao C., Takeshima R., Zhu J., Xu M., Sato M., Watanabe S., Kanazawa A., Liu B., Kong F., Yamada T., Abe J. A recessive allele for delayed flowering at the soybean maturity locus *E9* is a leaky allele of *FT2a*, a flowering locus *T* ortholog. *BMC Plant Biol.*, 2016, 16: 20 (doi: 10.1186/s12870-016-0704-9).
  70. Liu B., Kanazawa A., Matsumura H., Takahashi R., Harada K., Abe J. Genetic redundancy in soybean photoresponses associated with duplication of *phytochrome A* gene. *Genetics*, 2008, 180: 996-1007 (doi: 10.1534/genetics.108.092742).
  71. Watanabe S., Hideshima R., Xia Z., Tsubokura Y., Sato S., Nakamoto Y., Yamanaka N., Takahashi R., Ishimoto M., Anai T., Tabata S., Harada K. Map-based cloning of the gene associated with the soybean maturity locus *E3*. *Genetics*, 2009, 182: 1251-1262 (doi: 10.1534/genetics.108.098772).
  72. Watanabe S., Xia Z., Hideshima R., Tsubokura Y., Sato S., Harada K. A map-based cloning strategy employing a residual heterozygous line reveals that the *GIGANTEA* gene is involved in soybean maturity and flowering. *Genetics*, 2011, 188: 395-407 (doi: 10.1534/genetics.110.125062).
  73. Xia Z., Watanabe S., Yamada T., Tsubokura Y., Nakashima H., Zhai H., Anai T., Sato S., Yamazaki T., Lü S., Wu H., Tabata S., Harada K. Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. *PNAS*, 2012, 109(32): E2155-E2164 (doi: 10.1073/pnas.1117982109).
  74. Stewart D.W., Cober E.R., Bernard R.L. Modeling genetic effects on the photothermal response of soybean phenological development. *Agron. J.*, 2003, 95: 65-70 (doi: 10.2134/agronj2003.0065).
  75. Seferova I.V., Misyurina T.V., Nikishkina M.A. Ecologo-geographic estimation of biological potential of early varieties in soya north advance. *Sel'skokhozyaistvennaya Biologiya [Agricultural Biology]*, 2007, 5: 42-47 (in Russ.).
  76. Vishnyakova M., Seferova I. Soybean genetic resources for the production in the Non-Chernozem zone of the Russian Federation. *Legume perspectives (The Journal of the International Legume Society, Novi Sad, Serbia)*, 2013, 1: 7-9.
  77. Seferova I.V. *Maslichnye kul'tury*, 2016, 3(167): 101-105 (in Russ.).