Genomic analysis in soybean breeding

V.V. Barzanova*, A.A. Novikova

LLC "IC "Biruch-NT"", Alekseevka, Russia

DOI 10.18699/ICG-PlantGen2019-76

* e-mail: v.barzanova@brc.efko.ru

Abstract: Marker-assisted selection is a selection method, which use gains momentum within the framework of the current cultivated crops' selection programs. Microsatellite markers (SSR) belong to the class of rather informative, reproducible and inexpensive DNA-markers and are deemed to be efficient enough for genetic diversity studies. The practice of GWAS-analysis represents a more powerful instrument for determination of genotype-phenotype associations related to the economically valuable traits such as productivity, photoperiodicity, drought-resistance, protein content in the beans, short vegetation period.
Key words: soy; genomic selection; SSR-markers; GWAS.

1. Introduction

Marker-assisted selection represents an efficient selection method, the role of which in the modern cultivated crops' selection programs is becoming more and more important. This selection type allows increasing efficacy of the selection process and shortening the time needed for developing new highly productive soybean varieties.

GWAS (Genome-Wide Association Studies) used for the search of the quantitative trait loci employing gene linkage disequilibrium among neighboring SNPs provides a powerful genetic tool for associating phenotypes and genotypes. It allows detecting the soybean genetic variants that include the highest number of phenotypic traits.

2. Materials and methods

Currently the scientific research programs conducted in the "Biruch" Innovative center actively revolve around genetic selection studies aiming to develop a new highly productive soybean variety acclimatized to the weather conditions of the Central Black Earth region. This project incorporates the following works: selection of highly productive and resistant varieties, their marker analysis, sequencing and further GWAS analysis. The results presented within the framework of this research paper represent a detailed description of a single project stage that will be followed by further studies in the nearest future.

Empirical research employed microsatellite analysis for the soybean varieties' genotyping for the protein content of the seeds and abortiveness of the ovaries. Markers were taken from open sources [2]. The analysis was based on the PCR method (C1000 Touch amplifier by BIO-RAD). Separation of DNA fragments of various sizes was carried out by means of electrophoresis on agarose gel (2 % agarose solution) using an electrophoretic chamber with a power source (manufactured by BIO-RAD). Visualization of PCR products was carried out using the gel imaging documentation system GelDoc XR by BIO-RAD.

The next stage of work involved the preparation of DNA libraries of 140 soybean varieties. We selected 50 varieties that had the best performance against the following parameters: productivity, photoperiodicity, drought-resistance, protein content in the beans, short vegetation period. The sequencing of 50 samples was carried out using Illumina technology at the Skolkovo Science Center and at Novogen Company. After obtaining the sequences, the primary data processing was performed using the FastQC utility. The next stage concentrated on the selection of the suitable software based on the following criteria: the number of mapped reads, the number of properly paired reads, the time of mapping. As a result, BWA, Samtools, Hisat 2, and GATK software utilities were selected as the most efficient.

3. Results and discussion

Microsatellite markers (SSR) are considered the most informative, reproducible and relatively inexpensive class of DNA markers for genetic diversity studies. Therefore, the study of marker-associated economically valuable traits was performed using the method of microsatellite analysis.

Marker analysis of 30 soybean varieties was carried out within the framework of the present research work. Samples were selected on the basis of such characteristic as high adaptability to the soil and climatic conditions of the Belgorod region. These samples were tested for a group of markers responsible for the protein content of the beans and abortiveness of the ovaries:

The protein content: Satt 431; Satt 510; Satt 584; Satt 534; Satt 294; Satt 100; Satt 005; Satt 185; Satt 373; Satt 463; Satt 173; Satt 012; Satt 449; Pt 560; Sp; Pt 565 (Figure 1); abortiveness of the ovaries: Satt 150; Satt 291; Satt 567.

According to the results of the marker analysis focused on the size of amplicons after PCR, we obtained the following results:

- 1 sample shows a high number of protein markers (11 markers) and an average abortiveness rate (2 markers);
- 2 samples with an average number of protein markers (content greater than or equal to 7) and a low abortiveness rate (1 full marker, with a standard size, and 1 marker which size differs in electrophoresis detection);
- 10 samples show an average number of protein markers (content greater than or equal to 7) and the average abortiveness rate (2 markers);
- 8 samples have a low number of protein markers (from 2 to 5) and the average abortiveness rate (2 markers);
- 3 samples show a low indicator of the protein markers' number (from 2 to 5) and a high abortiveness rate (3 markers);

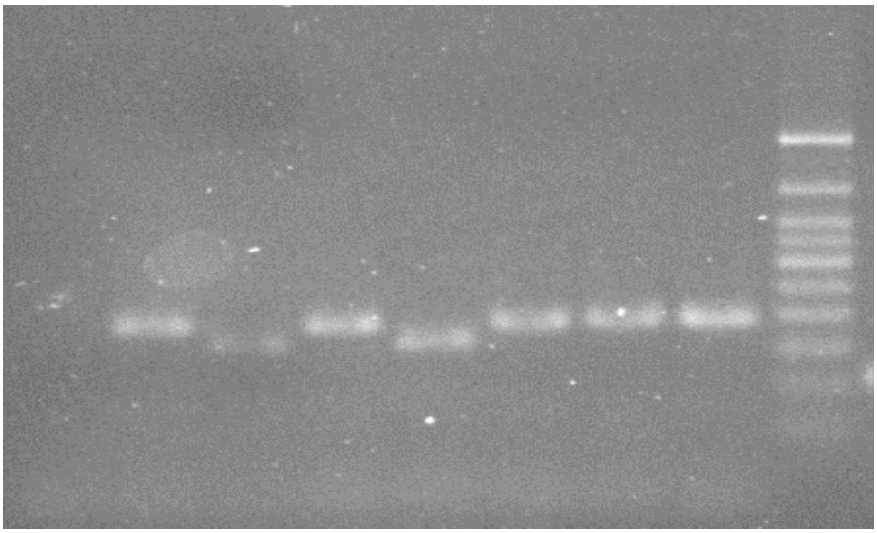

Figure 1. The results of electrophoresis in agarose gel with primers of microsatellite locus Pt 565 (fragment length 183 pairs nucleotides). Original.

- 2 samples show an average number of protein markers (from 5 to 8) and an average abortiveness rate (2 markers);
- 2 samples show a low indicator of the protein markers' number (from 2 to 5) and a low abortiveness rate (1 marker);
- 2 samples show a low indicator of the protein markers' number (from 2 to 5) and a high abortiveness rate (3 markers).

At present, the research in the sphere of genomic selection focuses mainly on finding additional microsatellite markers to increase the credibility of determination of sets of significant polymorphisms associated with certain traits. In prospect, it seems relevant to continue the studies and assess the influence of each marker on the manifestation of the certain traits. Alongside this process, we conduct the phenotyping of the varieties and identifying the most significant polymorphisms associated with economically valuable traits. The analysis results are summarized in a database for subsequent statistical data processing.

The determination of heterozygous plants (true hybrids) is conducted with the help of such markers as Satt 002, Satt 005, Satt 009, Satt 011, Satt 012, Satt 063, Satt 100, Satt 114, Satt 146, Satt 173, Satt 179, Satt 185, Satt 228, Satt 294, Satt 358, Satt 373, Satt 431, Satt 431, Satt 440, Satt 463, Satt 510, and Satt 547.

Another way to improve the credibility of determination of the sets of significant polymorphisms associated with a particular trait is through the use of the GWAS approach.

Currently, the sequencing of 50 soybean varieties has already been carried out and the samples were prepared for further GWAS analysis. Plant sequencing data was first processed using the FastQC utility. An array of the software packages tested for their efficiency in processing the data for comparison of the given samples with the reference sample included Hisat 2, BWA and Bowtie2. The results of the programms' comparison were as follows:

- HISAT2: 211,072,918 reads mapped, 167,295,328 properly coupled (79.26 %); mapping time ~1 hour.
- Bowtie2: 172,162,173 reads mapped; 154,655,842 properly coupled (89.83 %); mapping time ~6 hours.
- BWA: 183,090,700 reads mapped: 163,507,646 properly coupled (90.40%); mapping time ~2 hours.

After a thorough analysis, the BWA utility was chosen as the most productive of the data processing utilities. Samtools and Hisat 2 allowed us to prepare a reference genome, align the reads, convert them from sam* to bam* format, sort and index them. To obtain g.vcf* files, we used the GATK software. As a result, we obtained files containing information about the SNPs.

4. Conclusions

In order to increase the credibility of microsatellite analysis results, it is necessary to increase the number of markers used, assess the connection of each of them to the manifestation of certain traits and accumulate a statistically accurate database of phenotypic traits.

Another way to increase the effectiveness of the analysis of the genotype-phenotype associations (including quantitative traits) is through the use of GWAS analysis. To obtain a representative sample list, it is necessary to carry out the sequencing of other 90 variety samples.

When processing sequencing data, it is essential to use a set of utilities to improve the accuracy and speed of analysis.

References

Efremova V.V., Samelik E.G. Principles and practices of seed selection nowadays. *Scientific Journal*. 2015;106(02).

Perry CreganSoybean and Alfalfa Research Laboratory: https://soybase.org/resources/ssrold.php.

Desta Z.A., Ortiz R. Genomic selection: genome-wide prediction in plantimprovement. *Trends Plant Sci*. 2014;19:592–601.

Newell M.A., Jannink J.L. Genomic selection in plant breeding. *Methods Mol. Biol*. 2014;1145:117–130.

Conflict of interest. The authors declare no conflict of interest.