Genomic analysis in soybean breeding

Barzanova V.V.*, Novikova A.A.
Innovation center "Biruch-new technologies", Alekseevka, Russia
* e-mail: v.barzanova@brc.efko.ru

The patterns of heredity and genetic variability, which are peculiar to all organisms, are of great practical importance for agriculture. The role of the application of genotype-phenotype correlation analysis in the genomic selection of crops is currently growing. In the course of the work, foreign and Russian soybean breeding works were analyzed and the most significant SSR markers were identified. The most relevant for research are the genes with strongest impact on the traits of interest. To solve this problem, the following strategy was developed: 200 soybean varieties were selected from the VIR collection. Range of phenotypic traits of these grown varieties were analyzed. Obtained data have been recorded in journals with a view to further using it for GWAS analysis. The next stage of work, 50 soybean varieties were sequenced, having the best performance in the following ways: productivity, photoperiodism, drought resistance, protein content in beans, a short growing season. Sequencing has been carried out using Illumina technology. Primary data processing has been carried out using the FastQC utility, which allowed quality control to eliminate inconsistent results associated with the poor quality of the input data. Samtools and Hisat2 programs have been used for further data processing, which allow preparing the reference genome, align the reads, translate them from sam to bam format, sort them and index them. To prepare for the GWAS analysis, the GATK software package has been used:
• search for duplicate reads;
• local reorganization of readings;
• local recoding of readings around indels;
• synchronization of all information between each read and its paired pair;
• calibration of the quality assessment of the information base;
• verification of areas containing true SNP;
• hierarchical merging of bam files into a single gVCF.
As a result, we obtained gVCF files containing information about all 50K markers found in the samples. The next stage of our work will be GWAS analysis of 200 soybean varieties after NGS, which will allow genomic selection to create new appropriate soybean varieties.