sciendo

PROCEEDINGS OF THE LATVIAN ACADEMY OF SCIENCES. Section B,
Vol. 74 (2020), No. 4 (727), pp. 244–251.

DOI: 10.2478/prolas-2020-0039

EFFECT OF POPULATION SIZE ON GENOME-WIDE ASSOCIATION STUDY OF AGRONOMIC TRAITS IN SOYBEAN

Alibek Zatybekov[1], Yerlan Turuspekov[1,2], Botakoz Doszhanova[1], Svetlana Didorenko[3], and Saule Abugalieva[1,#]

[1] Institute of Plant Biology and Biotechnology, Almaty, 050040, KAZAKHSTAN

[2] Kazakh National University by al-Farabi, Almaty, KAZAKHSTAN

[3] Kazakh Research Institute of Agriculture and Plant Industry, Almalybak vil., Almaty region 040909, KAZAKHSTAN

* Corresponding author, absaule@yahoo.com

Communicated by Isaak Rashal

Soybean is increasingly becoming one of the main oilseeds crops in Kazakhstan, with a major growing area in the south-eastern part of the country. The comparative assessment of small-size (108) and medium-size (276 accessions) collections of soybean originated from five regions of the world was made in field trials in Almaty region (south-east of Kazakhstan). The full collection was genotyped using the Illumina iSelect 6K SNP (single nucleotide polymorphism) array that allowed observation of 4651 polymorphic SNPs. The genome-wide association study (GWAS) was conducted using the following six important agronomic traits: flowering time, seed maturation time, plant height, number of fertile nodes, number of seeds per node, and thousand seed weight. In total, GWAS using TASSEL package identified 56 stable QTNs (quantitative trait nucleotdies) for all the studied traits, including 42 QTNs detected using the medium-size collection. It was shown that the medium-size population, in comparison with the small-size population, was preferable for the detection of significant associations. A comparative analysis of the data with previously published results suggested that 36 of the stable 56 QTNs were presumably novel associations. The revealed QTNs can be used in the molecular breeding projects for soybean yield improvement.

Key words: *Glycine max, yield components, Illuminia genotyping array, association mapping.*

INTRODUCTION

Soybean (*Glycine max* L.) is one of the major legume crops in the world with a wide range of usage and economic profits. Products made from soybeans are used in everyday life from human nutrition and animal feed to industrial uses, such as textile fibres and biodiesel production (Han *et al*., 2001; Pimental *et al*., 2005). Due to the increasing demand of this crop for food, chemicals, and the feed industry, global soybean production has increased by 2.7 times over the past 20 years. In 2019/2020, it reached 339 400 million metric tons, and the majority of soybean crops were produced in Brazil, the USA, and Argentina (Anonymous, 2020b). In Kazakhstan, the local climate and the length of the photoperiod limit the areas that are relatively suitable for soybean cultivation. To date, over 90% of soybean production in Kazakhstan has been concentrated in the south-eastern part of the country, mostly in the Almaty region (Makulbekova *et al.,* 2017). In 2019, 229 thousand metric tons of soybeans were harvested in Kazakhstan, which was less than 1% of the global production (Anonymous, 2020a). Soybean is a significant part of the state agricultural programmes, which are focussed on increasing the area under this crop due to its local importance and the export potential.

Soybean breeders around the world, including in Kazakhstan, are challenged with the problem of finding an efficient breeding strategy for the development of cultivars with higher yield, improved seed quality, and resistance to biotic and abiotic stresses (Moose and Mumm, 2008; Ray *et al*., 2013; Makulbekova *et al.,* 2017). The main directions

of the national soybean breeding programme are aimed at increasing yield and quality (Didorenko *et al*., 2015; Abugalieva and Didorenko, 2016; Doszhanova *et al*., 2019) and number of disease resistant accessions (Maui *et al*., 2016; Zatybekov *et al*., 2018). Additionally, it is important to develop cultivars that are well adapted to different climatic conditions of the south-eastern region, and the north and east of the country (Didorenko *et al*., 2015; Abugalieva *et al*., 2016), by using diverse genetic germplasms. Previous pre-breeding works (Abugalieva and Didorenko, 2016; Abugalieva *et al*., 2016), as well as a study on *E* flowering genes and their association with yield in the soybean world collection, including local accessions grown in three different regions of Kazakhstan (Abugalieva *et al*., 2016), have provided the preliminary conditions for the expansion of soybean cultivation.

For the enhancement of local breeding projects, marker-assisted selection (MAS) has been proposed as one of the most powerful and effective modern approaches (Collard and Mackill, 2008; Moose and Mumm, 2008; Kumar *et al*., 2011). One of the most powerful approaches for understanding the genetic background of the genetic control of the complex traits is genome-wide association study (GWAS) (Korte and Farlow, 2013). This approach has been successfully applied worldwide to search for associations between DNA markers and soybean important agronomic traits associated with photosynthetic features (Lu *et al*., 2018), abiotic (Chen *et al*., 2019; Do *et al*., 2019) and biotic (Vuong *et al*., 2015; Schneider *et al*., 2016) stresses, yield components (Zhang *et al*., 2016; Contreras-Soto *et al*., 2017), and seed quality traits (Hwang *et al*., 2014; Vaughn *et al*., 2014). In Kazakhstan, GWAS has been applied to the study of the agronomic traits and resistance to fungal diseases in the soybean collection included in the genetic resources in Kazakhstan (Zatybekov *et al*., 2017, 2018). Still, the limitation of those studies was a small number of analysed accessions in the breeding pool, restricted to nearly a hundred lines. The analysis of the soybean genome-wide association studies that involve small (~100 accessions) (Contreras-Soto *et al*., 2017; Li *et al*., 2019), medium (~300 accessions) (Wang *et al*., 2019; Ikram *et al*., 2020), and large populations (500 and more accessions) (Fang *et al*., 2017; Lee *et al*., 2019) suggested that the effectiveness of GWAS was higher with enlargement of the size of the studied population. On the other hand, analysing a large population in multiple repetitions is costly and often is challenging for a small research group. Hence, it is important to test the efficiency of GWAS using medium-size collections in comparison to small-size populations. Therefore, the main purpose of this GWAS was to reveal quantitative trait nucleotides (QTN) for critical agronomic traits using a genetically diverse panel of soybean consisting of 276 accessions genotyped for 6000 SNP (single nucleotide polymorphism) markers and phenotyped in environmental conditions of Kazakhstan. The other aim was to test the robustness of GWAS using medium-size (276 accessions) versus small-size population (108 accessions) studied in the same environment but in different years.

MATERIALS AND METHODS

Plant material. The plant material consisted of two sets of the soybean collection, designated as Set 1 and Set 2 (Suppl., Table S1), which were studied on experimental plots of the Kazakh Research Institute of Agriculture and Plant Industry (KRIAPI). The information on the meteorological and climate conditions of the KRIAPI is given in Supplementum, Table S2. Set 1 consisted of 276 accessions from West and East Europe, East Asia, North America, and Kazakhstan, and studied in 2018 and 2019. Set 2 was comprised of the same subsets of Set 1, consisting of selected 108 accessions but tested in four years (2015, 2016, 2018, 2019). While Set 2 included 13 promising local lines, Set 1 integrated 34 accessions from Kazakhstan, involving 20 commercial cultivars (Suppl., Table S1) (Zatybekov *et al*., 2017).

Phenotypic evaluations of the collection. The soybean collection was grown in three randomised replicates in one square metre plots. In total, nine agronomic traits connected with the vegetation period, plant morphology, and yield components were studied. The list of traits included the flowering time (VER2), seed maturation time (R2R8), plant height (PH), number of lateral branches (NLB), number of fertile nodes (NFN), number of seeds per plant (NSP), thousand seeds weight (TSW), yield per plant (YP), and yield per plot (YpP). Evaluation protocols for each agronomic trait were standardised and measured according to the method by Korsakov *et al*. (1968). The determination of soybean maturation groups for each accession was done based on Fehr and Cavines *et al*. (1977). Set 1 was divided into seven seed maturation groups, while Set 2 was represented by three early maturity groups (Suppl., Table S3). Statistical analyses of the field data was conducted using the SPSS 22.0 and STATISTIKA 13.2 software packages.

Genotyping and genetic variation study. The panels of the studied collection were genotyped using the soybean 6K Illumina iSelect SNP array at Traitgenetics GmbH (Gatersleben, Germany). The Illumina Infinium procedure was performed according to the manufacturer's protocol. SNP genotyping was performed using the Illumina Genome Studio software version V2011.1 (Illumina Inc., 2018). A total 4651 SNP markers were used for further analysis after removing all monomorphic markers and markers with minor allele frequency (MAF) < 0.05. Since preliminary testing of accessions using SSR markers showed very high homogeneity within accessions (data not published), DNA from one plant per accession was used for SNP genotyping.

Association mapping analysis. Analysis of the population structure was carried out using STRUCTURE (v2.3.4.) software with a Bayesian Markov Chain Monte Carlo (MCMC) approach, based on the admixture and correlated frequency models (Evanno *et al*., 2005). Numbers of hypothetical groups ranging from k = 1 to k = 10 were assessed using 50 000 burn-in iterations, followed by 100 000 recorded iterations. The output from STRUCTURE was to determine the delta K value (ΔK) in STRUCTURE HARVESTER

(Earl and von Holdt, 2012). On the basis of the final *k* values, the Q-matrix for the three identified clusters was developed. GWAS for the above-listed traits was conducted using TASSEL 5.0 (v20191212) (Bradbury *et al*., 2007) based on the MLM (K+Q) (Zhang *et al*., 2010). Minimal threshold bars for low- and high-significance QTNs were $p < 0.0001$ and $p < 0.000001$, respectively. The genetic map was drawn using MapChart v.2.3 software (Voorrips, 2002). Pairwise LD between the markers based on their correlations (R^2) was calculated using TASSEL. R statistical software was used to plot the correlation between pairwise R^2 and the genetic distance (LD decay plot) (R Core Team, 2018).

RESULTS

Phenotypic variation of the main agronomic traits in the collection. The largest set of the study (Set 1) consisted of 276 accessions phenotyped in south-east of Kazakhstan for two years (2018 and 2019). Set 2, which is a subset of Set 1, consisted of selected 108 predominantly early flowering accessions harvested in 2015, 2016, 2018, and 2019. As expected, the ranges of variation in three plant adaptivity related traits, such as FT (VER2), SMT (R2R8), and PH, were significantly higher in Set 1 than in Set 2 ($p < 0.0001$). The Pearson's correlation test suggested that the average yield was positively correlated with FT and SMT (Fig. 1).

In the collection, the number of accessions from Kazakhstan increased from 13 in Set 2 to 34 in Set 1. The enlargement of the collection occurred because of addition of 20 commercial cultivars in the experiment, where ten of those cultivars belonged to maturation group II and eight to group III (Suppl., Table S3). The local standard (check) cultivar Zhansaya belonged to group II and showed 160.3 ± 0.8 g/m^2 in average yield. In total, thirteen accessions outperformed the local standard in YpP, including five cultivars from China and three each from Korea and Russia. The highest averaged YpP was recorded for the Chinese accession Hei lun 48 (302.4 ± 98.4 g/m^2), which also belonged to maturation group II.

Effect of population size on the identification of marker-trait associations. GWAS for six of nine agronomic traits scores was performed using the Java-based TASSEL with the application of the MLM (K+Q) model (Suppl., Table S4). The genotyping data consisted of polymorphic 4651 SNPs that were used for map drawing using MapChart v.2.3, and spanned 20 chromosomes with an average length of 47.4 Mb. The length of the chromosomes ranged from 37.3 Mb in Gm16 to 62.2 Mb in Gm18. The number of markers per chromosome varied from 185 in Gm11 to 295 in Gm13, with the mean number equal to 232.6. The average density of the SNP map was one marker per 246 Kb. Linkage disequilibrium (LD) decays at 5.4 Mb for the whole genome at R^2 of 0.1 (Suppl., Fig. S1).

The STRUCTURE HARVESTER suggested that the optimal number of groups was K = 3 (Suppl., Fig. S2). A comparative clusterisation of two sets for five studied groups

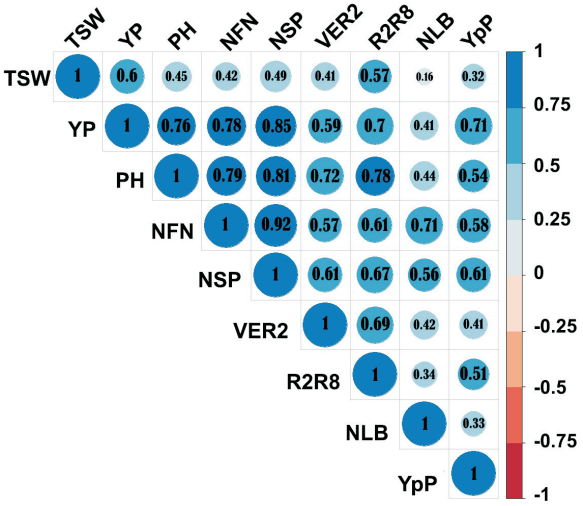


Fig. 1. Pearson correlation analysis among agronomy traits in the studied collection. VER2, flowering time; R2R8, seed maturation time; PH, plant height; NLB, number of lateral branches; NFN, number of fertile nodes; NSP, number of seeds per plant; TSW, thousand seeds weight; YP, yield per plant; YpP, yield per plot. Coefficients with significance at $p < 0.05$ are highlighted in red for negative correlations and blue for positive correlations. The intensity of colors and the size of circles indicates relationship strength according to the scales given on and bottom side of the plot. (the bigger circle and darker color – the higher significance).

Table 1. Clustering of soybean accessions with different origins for the two studied sets using output data from STRUCTURE, %

Origin	Number of accessions	Clusters in Set 1			Number of accessions	Clusters in Set 2		
		Q1	Q2	Q3		Q1	Q2	Q3
East Europe	120	24.17	36.67	39.17	70	15.71	35.71	48.57
West Europe	23	8.70	0	91.30	6	66.67	33.33	0
East Asia	55	7.27	5.45	87.27	3	66.67	33.33	0
North America	44	6.82	11.36	81.82	16	62.50	25.00	12.50
Kazakhstan	34	2.94	20.59	76.47	13	46.15	0	53,85

with different origins is given in Table 1 and Supplementum, Fig. S3. Particularly, in Set 1, the distribution of accessions in cluster Q3 from Kazakhstan (76.5 %) was similar to accessions from North Amerika (81.3 %). In Set 2, most local accessions were spread in three clusters congruently with genotypes from Eastern Europe (Table 1).

The population structure. The complete GWAS results are shown in Supplementum, Table S3, including the *p* value using MLM, minor allele frequency (MAF), phenotypic variance explained by markers R^2, and allele effects of the significant QTN. The Manhattan and quantile-quantile (QQ) plots of the study are presented in Supplementum, Fig. S4. The GWAS of the two studied sets for the six evaluated traits in each two-year trial, and their averages, allowed the identification of 56 stable QTNs, suggesting that each QTN was found in at least two studied environments,

Table 2. The number of identified stable QTNs using TASSEL package for two studied sets of soybean accessions.

Traits	Set 1 (n = 276)	Set 2 (n = 108)	Total
	2018, 2019	2015, 2016, 2018, 2019	
VER2	4	3	7
R2R8	6	5	11
PH	11	4	13
NFN	9	2	10
NSP	11	1	11
TSW	1	4	4
Total	42	19	56

including average values, over two years. In total, 42 and 19 stable QTNs were found in Set 1 and Set 2, respectively (Table 2).

The largest proportion of identified stable QTNs was expectedly in Set 1 (75%), confirming that the medium-sized population in comparison with the small-sized population was preferable in the detection of significant associations, as phenotypic variation in Set 1 was higher than in Set 2 ($p < 0.0001$). When the number of QTNs in Set 1 was compared with QTNs identified in Set 2, it was found that only 5 QTNs (11.9 %) in Set 1 matched with QTNs in Set 2. Notably, 26 of 42 QTNs in Set 1 had a higher significance value than $p < 0.000001$, which suggests that 61.9% of associations were represented as major QTLs in the studied set.

The lowest number of QTNs was identified in TSW (4) and the largest number in PH (by 13 associations). Interestingly, four of seven QTNs for VER2, and six of 11 QTNs for R2R8, have been published previously (Suppl., Table S4). Therefore, the identified SNP markers in those marker-trait associations may play an essential role in local soybean breeding projects for improvement of yield. Additionally, from Supplementum, Table S4, it can be seen that 36 of the 56 stable QTNs are presumably novel associations that require further validation study. Twenty-eight of those possibly new associations were trait-specific associations, while the remaining QTNs were significant for two and more studied traits (Figs. 2 and 3), and showed pleiotropic effects.

The physical positions of several QTNs were in close proximity of flowering-related genes (Figs. 2 and 3). Notably, the QTN VER2.ipbb.10.1 (Gm10.43894668) associated with VER2 is physically located in the proximity of Gigantea 3 (GI3) from 45294735 bp to 45316121 bp in Soybase.org database, which is responsible for the regulation of photoperiodism (Watanabe *et al*., 2011). In R2R8 analysis, several QTNs were found to be linked with other flowering-related genes. For instance, the QTN R2R8.ipbb.15.2 (Gm15.12317776) was found to be 0.9 Mbp from the *PHYB2*, which is responsible for the inhibitory role in floral initiation (Wu *et al*., 2013). On chromosome 16, the QTN R2R8.ipbb.16.1 (Gm16.32017661) is located 0.2 Mbp apart from Gigantea 2 (GI2), which is another important gene in the regulation of photoperiodism and flowering (Watanabe *et al*., 2011) (Fig. 3). It is important to note that QTNs associated with VER2, R2R8, and PH were found to be responsible for nine multiple associations (three and more), suggesting that those genetic factors have pleiotropic effects affecting key agronomic traits (Figs. 2 and 3). Particularly, it was revealed that one of those QTNs was associated with VER2, three QTNs with PH, and five QTNs with R2R8 (Figs. 2 and 3).

DISCUSSION

Soybean is a comparatively new crop for the Central Asian region, including Kazakhstan. Traditionally, it is cultivated in the south-eastern region of the country, where breeders have previously preferred germplasms from the early maturity group 00 (Abugalieva *et al*., 2016; Abugaliyeva and Didorenko, 2016; Zatybekov *et al*., 2018). Recent intensive field trials have resulted in the successful introduction of new cultivars from later maturation groups II and III. Therefore, the number of local accessions in this study increased from 13 accessions with an early flowering time to 34 accessions, where 18 of them were commercial cultivars with a comparatively late flowering time (maturation groups II and III). The comparative study of Set 1 in 2018 and 2019 using the average yield has allowed to reveal 128 accessions that outperformed the local standard cultivar Zhansaya (160.3 ± 0.8 g/m^2). Interestingly, the majority of those accessions, including the Chinese accession Hei lun 48 with the highest average yield (302.4 ± 98.4 g/m^2), belong to maturation group II. Field experiments in south-eastern Kazakhstan indicated that local breeders should focus on the breeding of new perspective cultivars from maturation groups II and III.

GWAS with a small population is often associated with an inflated *p*-value, and confounding effects can result in the identification of pseudo QTLs (Teng *et al*., 2009; Wang *et al*., 2012). Still, in many cases, researchers face problems with a limited number of accessions in the collection and/or restrictions in experimental plot facilities, leading to GWAS with a relatively small-sized collection. In these circumstances, an important question arises about the size of the studied population. QQ plots using both small and medium-size populations suggested that the obtained GWAS results for a medium-sized population were reliable. The largest proportion of identified stable QTNs was found in Set 1, as 42 out of 56 QTLs (or 75%) were identified in the medium-size population. The results clearly suggested that the medium-sized population, in comparison with the small-sized population, was preferable in the detection of significant associations. The robustness of GWAS is also evident based on literature survey, which indicated that 20 QTNs and genes of 56 identified stable QTNs had previously been reported elsewhere (Suppl., Table S4). A large bulk of matching QTNs was particularly visible in the plant growth associated traits VER2 and R2R8, as four of seven QTNs for VER2, and six of 11 QTNs for R2R8, had been pub-

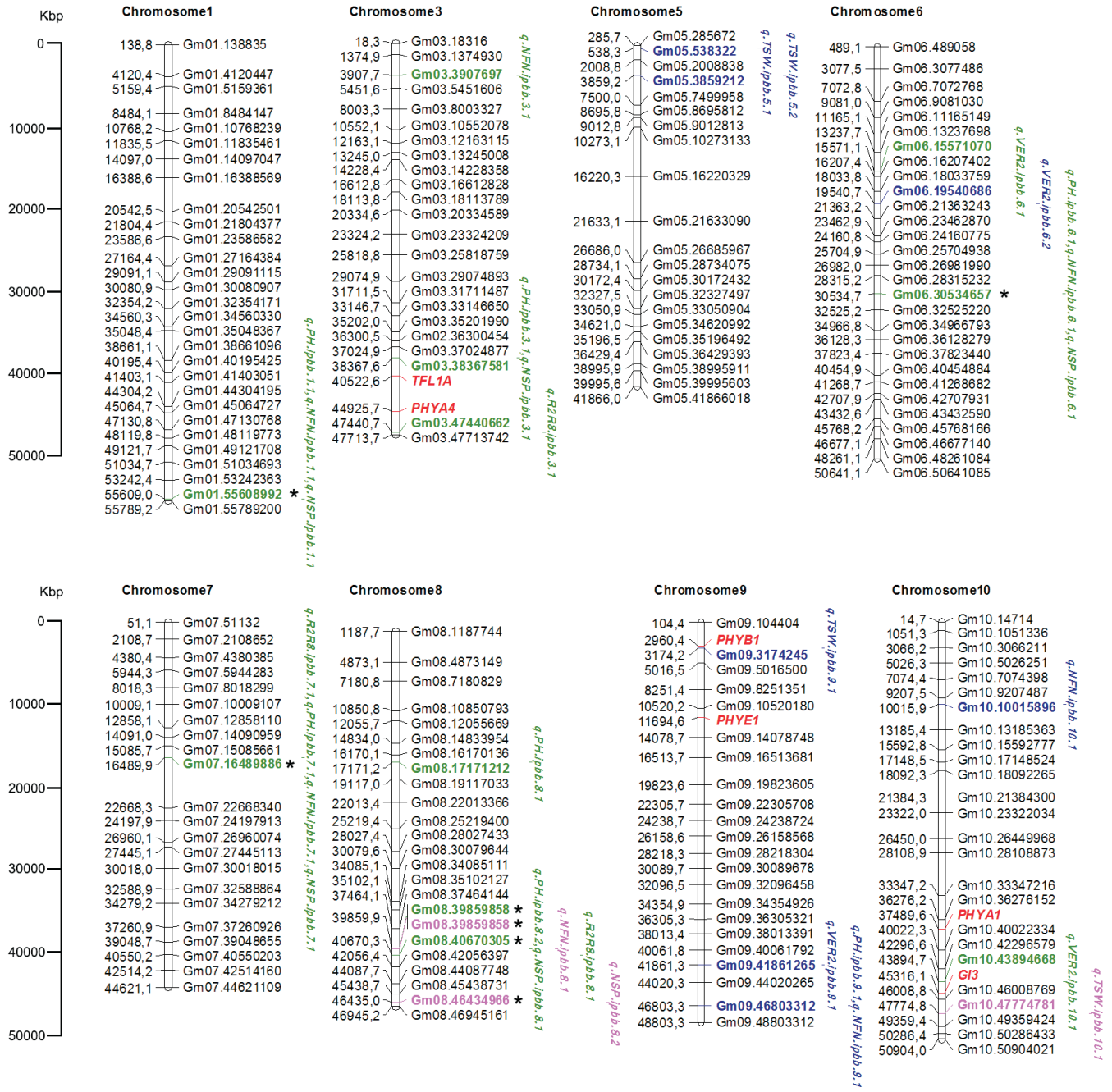


Fig. 2. Location of stable 56 Quantitative Trait Nucleotides (QTNs) identified using TASSEL-based Mixed Linear Methods (MLM). The QTNs identified using Set 1 are shown in green, using Set 2 in blue, and using both sets in pink. The positions of known soybean genes were shown in red. The SNP positions for major QTNs shown by asterisks. Chromosomes from 1 to 10.

lished previously (Suppl., Table S4). Additional evidence to the reliability of identified QTNs for VER2 and R23R8 is the fact that their positions were matched to the physical locations of flowering genes, shown in Figures 2 and 3. For instance, flowering gene *GI3* was found to be located nearby Gm10.43894668, which is the peak SNP marker associated with VER2 (Fig. 3). Another peak SNP markers for four QTNs associated with R2R8 were physically mapped in the proximity of flowering genes *PHYB2*, *GI2* and *E4* (Fig. 3). A total of 36 of the 56 QTNs in the six analysed traits could potentially be novel, but only five of those QTNs were in genic positions, suggesting that additional studies will be required for genomic evaluations of the 31 QTNs located in intergenic positions.

CONCLUSIONS

Field trials using a collection of soybeans with 276 accessions showed that the average yield over two years was positively correlated with late flowering and seed maturation time. The assessment of the TSW suggested that accessions from maturation group II were the most productive in the region. One hundred and twenty-eight accessions of the

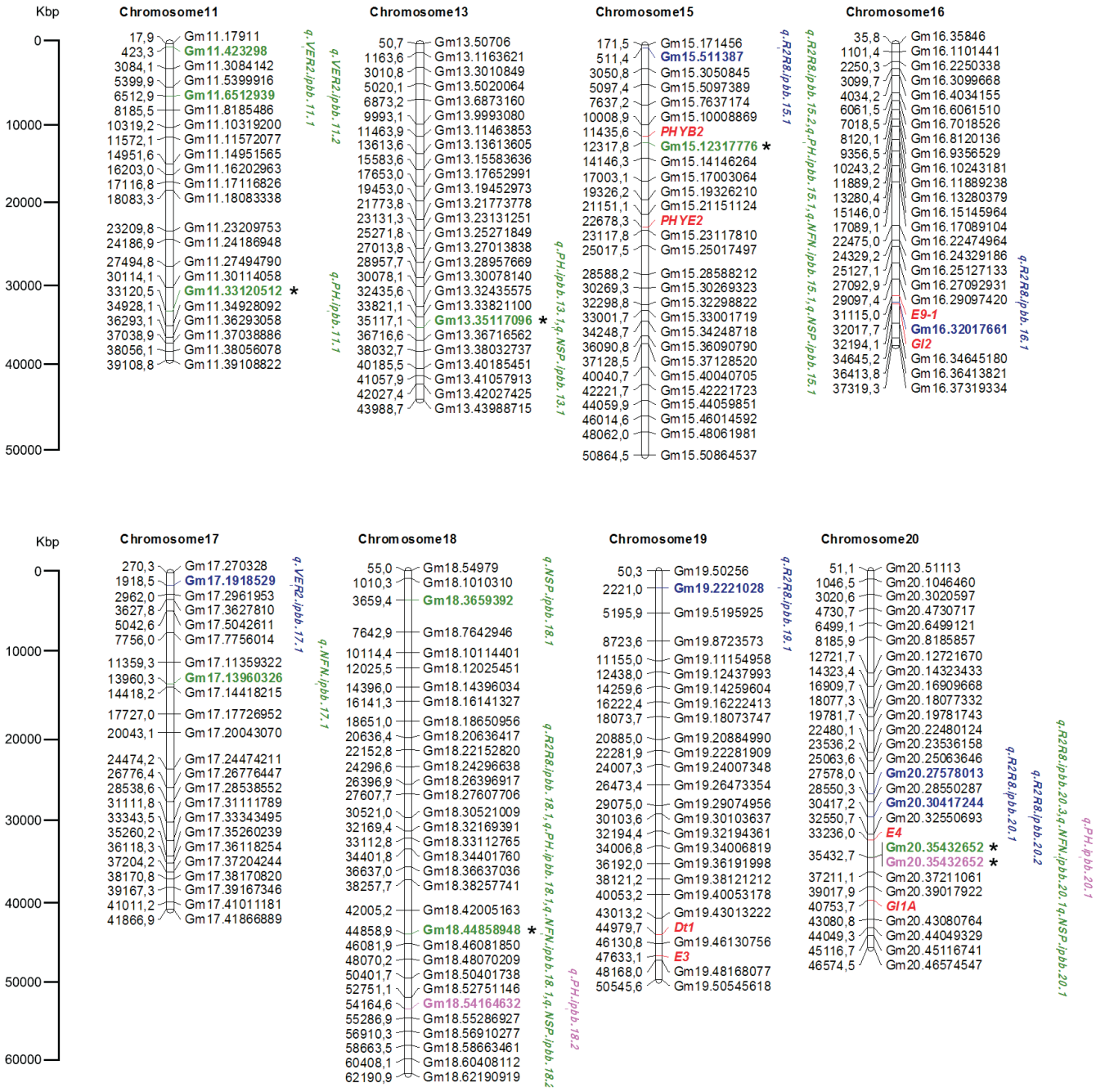


Fig. 3. Locations of stable 56 Quantitative Trait Nucleotides (QTNs) identified using TASSEL-based Mixed Linear Methods (MLM). The QTNs identified using Set 1 are shown in green, using Set 2 in blue, and using both sets in pink. The positions of known soybean genes were shown in red. The SNP positions for major QTNs shown by asterisks. Chromosomes from 11 to 20.

collection outperformed the local check cultivar Zhansaya (maturation group II, 160.3 ± 0.8 g/m^2), and should be further tested in the local breeding projects. The full collection was genotyped using the Illumina iSelect 6K SNP array, allowing the generation of 4651 polymorphic SNPs. GWAS was performed using the TASSEL package with the use of the MLM method. In total, GWAS identified 56 stable QTNs for six studied agronomic traits, including 42 QTNs detected using the medium-sized collection. The results are indicating that the medium-sized population, in comparison with the small-sized population, is preferable in the detection of significant associations. A comparative analysis of the data with previously published results suggested that 36 out of those 56 QTNs were presumably novel marker-trait associations. The identified QTNs can be used in local projects for the improvement of soybean productivity via the marker-assisted selection approach.

SUPPLEMENTARY MATERIALS

Table S1: The list of soybean accessions and their origin. *Table S2*: Location and meteorological data of the experimental site. *Table S3*: The distribution of the soybean collection accessions by maturation group, according to their origin. *Table S4*: A list of identified QTNs for Set 1 and Set2 using MLM (K+Q) method in TASSEL.

Figure S1: Genome-wide linkage disequilibrium in the soybean genome using 4651 SNP. The horizontal red line denotes the LD threshold of 0.1 for pair-wise R2. *Figure S2*: Optimal delta K value in Set 1 of the soybean collection determined by using STRUCTURE HARVESTER. *Figure S3*: The bar plot of clustering of collection using Structure software based on 4651 SNP markers. *Figure S4*: Manhattan and Quantile-quantile (QQ) plots obtained using TASSEL packages based on Set 1.

AUTHORS' CONTRIBUTIONS

Conceptualization, YT and SA; methodology, AZ, YT and SA; formal analysis, AZ, SD and BD; investigation, AZ, BD and YT; resources, SD and SA; data curation, AZ, SD and BD; writing–original draft preparation, AZ; writing–review and editing, YT and SA; visualization, AZ; supervision, YT; project administration, SA All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGMENTS

The authors would like to thank the technical staff of the Kazakh Research Institute of Agriculture and Plant Industry for the agronomic management of the field trials. The authors would like to acknowledge the funding from the Ministry of Education and Sciences of the Republic of Kazakhstan for project AP05131592.

REFERENCES

Abugalieva, S., Didorenko, S., Anuarbek, S., Volkova, L., Gerasimova, Y., Sidorik, I., Turuspekov, Y. (2016) Assessment of soybean flowering and seed maturation time in different latitude regions of Kazakhstan. *PLoS ONE*, **11** (12), 1–11.

Abugaliyeva, A. I., Didorenko, S. V. (2016). Genetic diversity of soybean cultivars belonging to different ripeness groups with regard to performance and quality. *Vavilov J. Gen. Breed.,* **20** (3), 303–310.

Anonymous (2020a). Official website of the Ministry of National Economy of the Republic of Kazakhstan. Statistics committee. Available at: http://stat.gov.kz/. (accessed 28.04.2020).

Anonymous (2020b). Official website of the United States government. Department of Agriculture. World Agricultural Supply and Demand Estimates. Available at: http://usda.gov/ (accessed 28.04.2020).

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**, 2633–2635.

Chen, L., Fang, Y., Li, X., Zeng, K., Chen, H., Zhang H., Yang H., Cao, D., Hao, Q., Yuyan, S. (2019). Identification of soybean drought-tolerant genotypes and loci correlated with agronomic traits contributes new candidate genes for breeding. *Plant Mol. Biol.,* **102**, 109–122.

Collard, B. C., Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philos.Transact. Roy. Soc. B: Biol. Sci.,* **363** (1491), 557–572.

Contreras-Soto, R. Mora, F., de Oliveira, M. A. R., Higashi, W., Scapim, C. A., Schuster, I. (2017). A genome-wide association study for agronomic traits in soybean using SNP markers and SNP based haplotype analysis. *PLoS ONE*, **12** (2).

Contreras-Soto, R. I., Mora, F., Lazzari, F., de Oliveira, M. A. R., Scapim, C. A., Schuster, I. (2017). Genome-wide association mapping for flowering and maturity in tropical soybean: Implications for breeding strategies. *Breeding Sci.,* **67** (5), 435–449.

Didorenko, S. V., Sprygailova, Y. N., Kudaibergonov, M. S., Abugalieva, A. I. (2015). Selection of soybean in East Kazakhstan. *Science World*, **1**, 83–88.

Do, T. D., Vuong, T. D., Dunn, D., Clubb, M., Valliyodan, B., Patil, G., Chen, P., Xu, D., Nhuyyen, H. T., Shannon, J. G. (2019). Identification of new loci for salt tolerance in soybean by high-resolution genome-wide association mapping. *BMC Genom.,* **20** (1), 318.

Doszhanova, B. N., Didorenko, S. V., Zatybekov, A. K., Turuspekov, Y. K., Abugalieva, S. I. (2019). Analysis of soybean world collection in conditions of south-eastern Kazakhstan. *Int. J. Biol. Chem.,* **12** (1), 33–40.

Earl, D. A., von Holdt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour*., **4** (2), 359–361.

Evanno, G., Regnaut, S., Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol*., **14** (8), 2611–2620.

Fang, C., Ma, Y., Wu, S., Liu, Z., Wang, Z., Yang, R., Hu, G. (2017). Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.,* **18** (1), 161.

Fehr, W. R., Cavines, C. E. (1977) *Stages of Soybean Development. Special Report. 87.* Cooperative Extention Service, Iowa State University, Ames, Iowa, USA.

Han, B. Z., Rombouts, F. M., Nout, M. J. R. (2001). A Chinese fermented soybean food. *Int. J. Food Microbiol.,* **65**, 1–10.

Hwang, E.-Y., Song, Q., Jia, G., Specht, J. E., Hyten, D. L., Costa, J., Cregan, P. B. (2014). A genome-wide association study of seed protein and oil content in soybean. *BMC Genom.*, **15** (1), 1.

Ikram, M., Han, X., Zuo, J.-F., Song, J., Han, C.-Y., Zhang, Y.-W., Zhang, Y.-M. (2020). Identification of QTNs and their candidate genes for 100-seed weight in soybean (*Glycine max* L.) using multi-locus genome-wide association studies. *Genes*, **11** (7), 714.

Korsakov, N. I., Makasheva, R. H., Adamova, O. P. (1968). *Methodology for Studying the Collection of Legumes* [Корсаков, Н. И., Макашева, Р. Х., Адамова, О. П. *Методика изучения коллекции зернобобовых культур*]. VIR, Leningrad. 175 pp. (in Russian).

Korte, A., Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods*, **9** (1), 29.

Kumar, J., Choudhary, A. K., Solanki, R. K., Pratap, A. (2011). Towards marker-assisted selection in pulses: A review. *Plant Breed.*, **130** (3), 297–313.

Lee, S., Van, K., Sung, M., Nelson, R., LaMantia, J., McHale, L. K., Rouf Mian, M. A. (2019). Genome-wide association study of seed protein, oil and amino acid contents in soybean from maturity groups I to IV. *Theor. Appl. Gen.,* **132**, 1639–1659.

Li, J., Zhao, J., Li, Y., Gao, Y., Hua, S., Nadeem, M., Sun, G., Zhang, W., Hou, J., Wang, X., Qiu, L. (2019). Identification of a novel seed size associated locus *SW9-1* in soybean. *Crop J.*, **7**, 548–559.

Lu, H., Yang, Y., Li, H., Liu, Q., Zhang, J., Yin, J., Chu S., Zhang, X., Kaiye, Y., Lv, L., Chen, X. (2018). Genome-wide association studies of photosynthetic traits related to phosphorus efficiency in soybean. *Frontiers Plant Sci.,* **9**, 1226.

Makulbekova, A., Iskakov, A., Kulkarni, K. P., Song, J. T, Lee, J. D. (2017). Current status of future prospects of soybean production in Kazakhstan. *Plant Breed. Biotechnol.,* **5**, 55–66.

Maui, Ą. Ą., Sauranbaev, B. N., Orazbaev, K. I. (2016). Soybean pathogens in the south-east of Kazakhstan. *J. Hum. Admin. Sci.,* **3** (5), 20–26.

Moose, S. P., Mumm, R. H. (2008). Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiol.,* **147** (3), 969–977.

Pimental, D., Patzek, T. W. (2005). Ethanol production using corn, switchgrass, and wood, biodiesel production using soybean and sunflower. *Nat. Res.Res.,* **14** (1), 65–76.

Ray, D. K., Mueller, N. D., West, P. C, Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS ONE*, **8** (6), e66428.

Schneider, R., Rolling, W., Song, Q., Cregan, P., Dorrance, A. E., McHale, L. K. (2016). Genome-wide association mapping of partial resistance to *Phytophthora sojae* in soybean plant introductions from the Republic of Korea. *BMC Genomics*, **17** (1), 607.

Teng, W., Han, Y., Du, Y., Sun, D., Zhang, Z., Qiu, L., Sun, G., Li, W. (2009). QTL analyses of seed weight during the development of soybean (*Glycine max* L. Merr.). *Heredity*, **102** (4), 372–380.

Vaughn, J. N., Nelson, R. L., Song, Q., Cregan, P. B., Li, Z. (2014). The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. *G3: Genes|Genomes|Genetics*, **4**, 2283–2294.

Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered*., **93**, 77–78.

Vuong, T. D., Sonah, H., Meinhardt, C. G., Deshmukh, R., Kadam, S., Nelson, R. L., Shannon, J. G., Nguyen, H. T. (2015). Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. *BMC Genom.*, **16** (1), 593.

Wang, H., Smith, K. P., Combs, E., Blake, T., Horsley, R. D., Muehlbauer, G. J. (2012). Effect of population size and unbalanced data sets on QTL detection using genome-wide association mapping in barley breeding germplasm. *Theor. Appl. Genet.*, **124**, 111–124.

Wang, J., Zhao, X., Wang, W., Qu, Y., Teng, W., Qiu, L., Zheng, H., Han, Y., Li, W. (2019). Genome-wide association study of inflorescence length of cultivated soybean based on the high-throughout single-nucleotide markers. *Molecular Genetics and Genomics*, **294** (3), 607–620.

Zatybekov, A., Abugalieva, S., Didorenko, S., Gerasimova, Y., Sidorik, I., Anuarbek, S., Turuspekov, Y. (2017). GWAS of agronomic traits in soybean collection included in breeding pool in Kazakhstan. *BMC Plant Biology*, **17** (S1), 179.

Zatybekov, A., Abugalieva, S., Didorenko, S., Rsaliyev, A., Turuspekov, Y. (2018). GWAS of soybean breeding collection for resistance to fungal diseases in condition of South-East and South Kazakhstan. *Vavilov J. Gen. Breed.,* **22** (5), 536–543.

Zatybekov, A. K., Agibaev, A. Z., Didorenko, S. V., Abugalieva, S. I., Turuspekov, Y. K. (2018). Analysis of resistance to *Septoria glycines* in soybean world collection harvested in south-eastern Kazakhstan. *News NAS of RK. Ser. Agricult. Sci.,* **5**, 44–52.

Zhang, J., Song, Q., Cregan, P. B., Jiang, G. L. (2016). Genome-wide association study. genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theor. Appl. Genet*., **129**, 117–130.

Zhang, Z., Ersoz, E., Lai, C., Todhunter, R., Tiwari, H., Gore, M., Bradbury, P., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet*., **42**, 355–360.

Received 10 July 2020
Accepted in the final form 24 July 2020

POPULĀCIJAS IZMĒRA IETEKME UZ SOJAS AGRONOMISKO PAZĪMJU VISA GENOMA ASOCIĀCIJU PĒTĪJUMA REZULTĀTIEM

Pamatojoties uz 4651 polimorfo SNP (viena nukleotīda polimorfisms) marķieru analīzi sojas 108 un 276 paraugu kolekcijās, parādīts, ka lielākā kolekcijā ir iespējams atrast vairāk statistiski ticamu asociāciju. Atrastas vairākas agrāk nezināmas asociācijas, kuras var izmantot sojas selekcijā.