GWAS of a soybean breeding collection from South East and South Kazakhstan for resistance to fungal diseases

A. Zatybekov[1, 2], S. Abugalieva[1], S. Didorenko[3], A. Rsaliyev[4], Y. Turuspekov[1] ✉

[1] Institute of Plant Biology and Biotechnology, Almaty, Kazakhstan
[2] Kazakh National Agrarian University, Almaty, Kazakhstan
[3] Kazakh Research Institute of Agriculture and Plant Growing, Almalybak, Almaty Region, Kazakhstan
[4] Research Institute for Biological Safety Problems, Gvardeiskiy (Otar), Zhambyl Region, Kazakhstan

Soybean (*Glycine max* (L.) Merr) is an essential food, feed, and technical culture. In Kazakhstan the area under soybean is increasing every year, helping to solve the problem of protein deficiency in human nutrition and animal feeding. One of the main problems of soybean production is fungal diseases causing yields losses of up to 30 %. Modern genomic studies can be applied to facilitate efficient breeding research for improvement of soybean fungal disease tolerance. Therefore, the objective of this genome-wide association study (GWAS) was analysis of a soybean collection consisting of 182 accessions in relation to fungal diseases in the conditions of South East and South Kazakhstan. Field evaluation of the soybean collection suggested that *Fusarium* spp. and *Cercospora sojina* affected plants in the South region (RIBSP), and *Septoria glycines* – in the South East region (KRIAPP). The major objective of the study was identification of QTL associated with resistance to fusarium root rot (FUS), frogeye leaf spot (FLS), and brown spot (BS). GWAS using 4 442 SNP (single nucleotide polymorphism) markers of Illumina iSelect array allowed for identification of fifteen marker trait associations (MTA) resistant to the three diseases at two different stages of growth. Two QTL both for FUS (chromosomes 13 and 17) and BS (chromosomes 14 and 17) were genetically mapped, including one presumably novel QTL for BS (chromosome 17). Also, five presumably novel QTL for FLS were genetically mapped on chromosomes 2, 7, and 15. The results can be used for improvement of the local breeding projects based on marker-assisted selection approach.

Key words: soybean; fusarium root rot; frogeye leaf spot; brown spot; GWAS; SNP; QTL mapping.

Полногеномный анализ ассоциаций с устойчивостью к грибным болезням в коллекции сои в условиях Юго-Восточного и Южного Казахстана

А. Затыбеков[1, 2], С. Абугалиева[1], С. Дидоренко[3], А. Рсалиев[4], Е. Туруспеков[1] ✉

[1] Институт биологии и биотехнологии растений Комитета науки Министерства образования и науки Республики Казахстан, Алматы, Казахстан
[2] Казахский национальный аграрный университет, Алматы, Казахстан
[3] Казахский научно-исследовательский институт земледелия и растениеводства, пос. Алмалыбак, Алматинская область, Казахстан
[4] Научно-исследовательский институт проблем биологической безопасности Комитета науки Министерства образования и науки Республики Казахстан, пос. Гвардейский, Жамбылская область, Казахстан

Соя (*Glycine max* (L.) Merr) – важная пищевая, кормовая и техническая культура. В Казахстане площадь под соей увеличивается с каждым годом, что обусловлено ее важностью в решении проблемы дефицита белка в питании людей и кормлении животных. Одной из основных проблем производства сои являются грибные болезни, вызывающие потери урожая до 30 %. Для повышения эффективности селекции, направленной на улучшение устойчивости сои к болезням, могут быть использованы современные геномные технологии. Таким образом, целью настоящего исследования был полногеномный анализ ассоциаций (GWAS) в коллекции сои, состоящей из 182 образцов, на устойчивость к грибным болезням в условиях Юго-Восточного и Южного Казахстана. В результате полевой оценки коллекции сои обнаружены растения, пораженные *Fusarium* spp. и *Cercospora sojina* в Южном регионе (НИИПББ) и *Septoria glycines* – в Юго-Восточном регионе (КазНИИЗиР). Исследование было нацелено на идентификацию локусов количественных признаков (ЛКП), связанных с устойчивостью к основным заболеваниям, таким как фузариоз корневой гнили (FUS), церкоспороз (FLS) и септориоз (BS). GWAS с использованием 4442 SNP-маркеров (single nucleotide polymorphism) матрицы Illumina iSelect позволил идентифицировать 15 ассоциаций маркер–признак (МТА) на устойчивость к трем болезням на двух разных стадиях роста. Генетически картированы два ЛКП как для FUS (хромосомы 13 и 17), так и для BS (хромосомы 14 и 17), включая один предположительно новый ЛКП для BS, который был идентифицирован на хромосоме 17. Кроме того, пять предположительно новых ЛКП для FLS были идентифицированы на хромосомах сои 2, 7 и 15. Результаты исследования могут быть использованы для улучшения селекционных программ, в том числе маркер-опосредованной селекции.

Ключевые слова: соя; фузариоз корневой гнили; церкоспороз; септориоз; GWAS; SNP; ЛКП картирование.

✉ e-mail: yerlant@yahoo.com

Soybean (*G. max* (L.) Merrill.) is one of the most important legumes in the world due to its high nutritional value and protein content (Masuda, Goldsmith, 2009). In Kazakhstan, this crop is mainly cultivated in the South East region. According to the Agency for Statistics of the Republic of Kazakhstan, in 2017 soybean was grown in an area of 137.4 thousand hectares (http://www.fcc.kz/attachments/article/4325). For further development of the soybean industry, the Government of Kazakhstan has declared a new initiative to expand the soybean area to 400 thousand hectares by 2020 to ensure its yield at 1 million tons (Zatybekov et al., 2017).

The productivity of soybean largely depends on availability of well-adapted cultivars with approptiate flowering and maturity times to match various ecological environments of the country (Zhang et al., 2007). Our previous study based on evaluation of 120 soybean accessions in three different regions of Kazakhstan (Abugalieva et al., 2016) has confirmed the results of observations in other parts of the world (Contreras-Soto et al., 2018; Copley et al., 2018), which underline the importance of suitable flowering time for plant adaptation in a particular environmental niche.

Another important factor that severely limits the soybean productivity worldwide is susceptibility to harmful diseases (Yang X.B. et al., 2001; Vidic et al., 2013). For instance, 25 known diseases posed a constant threat to the productivity of soybean in the USA (Mueller et al., 2010). In China, out of eight most common diseases, six are caused by fungi. In Russia, reports suggest there are up to 32 soybean diseases (Zaostrovnykh, 2005; Kurilova, 2010; Polozhieva, Dubovitskaya, 2015). In Kazakhstan, more than ten fungal diseases of soybean have been identified (Mombekova et al., 2013; Didorenko et al., 2014), and with expansion of the area under the crop, it is an obvious necessity to study the genetic background associated with the tolerance to harmful pathogens.

The damage caused by various diseases is determined by environmental conditions, the biology and spread of the parasite, and the characteristics of breeding material (Faske et al., 2014). Different parts of the plant, including seeds, sprouts, roots, shoots, leaves, and beans can be severely affected by these diseases. In this respect, all soybean diseases can be separated into three large groups: 1) diseases of seeds, sprouts, and seedlings; 2) patches that affect various parts of the plant; 3) diseases that cause the plants to wilt (Faske et al., 2014). In general, the total yield loss from susceptibility to fungal diseases can reach up to 40 % (Zaostrovnykh, 2005). J.K. Pataky and S.M. Lim (1981) reported that soybean yield loss due to *S. glycines* was associated with reduction of seed weight. M.D. Dias et al. (2016) identified a highly significant correlations ($p < 0.01$) between yield and soybean *Colletotrichum truncatum* incidence on pods ($r = –0.85$). About 90 kg/ha of soybean grain were lost for each 1 % increment in the disease incidence.

Currently, a large number of genes controlling the resistance to various diseases and pests have been identified (Prabhu et al., 1999; Wang J. et al., 2010; Vidic et al., 2013). Several soybean mapping populations were developed for genetic localization of QTL and genes associated with the soybean diseases, such as rhizoctonia root rot (RRR, caused by *Rhizoctonia solani*) (Zhao et al., 2005), fusarium root rot (FUS, caused by *Fusarium* spp.) (Stacey, 2008), phytophthora root rot (PRR, caused by *Phytophthora sojae*) (Zhang et al., 2013), frogeye leaf spot (FLS, caused by *C. sojina*) (Mian et al., 1999) and sclerotinia stem rot (SCL, caused by *Sclerotinia sclerotiorum*) (Zhao et al., 2015). The majority of these studies were based on use of SSR (simple sequence repeat) microsatellite markers. However, with the development of SoySNP50K iSelect SNP (single nucleotide polymorphism) array (Song et al., 2013), most of the modern studies rely on the use of SNP markers, which are crucial for genome-wide association studies (GWAS) (Klein, 2007). GWAS is based on use of whole genome genotyping and a detailed phytopathological and morphological description of collections with a high level of genetic diversity (Klein, 2007). A survey of recent reports has shown successful use of GWAS for studying soybean resistance to fungal diseases (Bao et al., 2015; Iquira et al., 2015; Schneider et al., 2016; Qin et al., 2017).

The purpose of this study was to assess the tolerance of the soybean germplasm collection represented by 182 accessions from major soybean growing regions from all around the world to most harmful diseases spreading in the South and South East of Kazakhstan. The GWAS was applied for identification of marker-trait associations for resistance to FUS, FLS and BS.

Materials and methods

The analyzed soybean collection consisted of 182 accessions, including 18 released cultivars and prospective breeding lines from Kazakhstan (Zatybekov et al., 2017). The accessions represented 12 countries from 5 geographic regions, including Western and Eastern Europe, North America, East Asia, and Kazakhstan. The collection was tested in the experimental plots of Research Institute of Biological Safety Problems (RIBSP, southern Kazakhstan) and Kazakh Research Institute of Agricultural Plant Production (KRIAPP, south-eastern Kazakhstan) in 2016–2017. Despite the environment similarities of the two localities, the conditions of soybean growth were different, as KRIAPP tested plants in irrigated, and RIBSP – in non-irrigated sites. Plants were grown in 1 meter long rows with a 30-cm distance between adjacent rows and a 5-cm gap between plants within rows. In total, the data for mean values of eight agronomic traits of the 182 soybean accessions harvested in two environments were subjected to further statistical analysis. The eight traits included the following data: days to seedling emergence (VE), days to flowering time (R2), days to development of pods (R4), days to full maturity of seeds (R8), plant height (PH), number of seeds per plant (NSP), thousand seeds weight (TSW), and yield per plant (YP).

Disease resistance analysis was carried out in relation to the three fungal diseases spread in the regions. In the South East the plants were analyzed for resistance to BS (caused by *S. glycines*), while in the South they were tested for resistance to FLS (*C. sojina*) and FUS (caused by a group of unidentified *Fusarium* pathogens indicated in this study as *Fusarium* spp.). The plant resistance to fungal diseases of the leaf surface was characterized based on a nine-point scale, where point 1 stood for highly resistant (no symptoms), 3 – for resistant (5–19 % foliage affected), 5 – for partially resistant (20–49 % of the foliage affected), 7 – for susceptible (50–79 % of the foliage affected), and 9 – for highly susceptible (up to 80 % of the foliage affected) (Hnetkovsky et al., 1996). The plant resistance to root rot was characterized based on a five-point scale,

where point 1 stood for healthy root without infection symptoms, 2 – for slight cortical necrosis or vascular discoloration, 3 – for moderate cortical necrosis or vascular discoloration, 4 – for extensive cortical or vascular tissue distroyed, and 5 – for withering and dying of a plant (Leath, Carroll, 1982). The disease severity after infection as a percentage of the affected area and the healthy part of plants was noted as well. Therefore, the resistance to all three diseases were marked with number 1 (for instance, FUS1), and the diseases severity were marked as number 2 (for instance, FUS2). Statistical analyses of obtained data were calculated by using Statistical Package for the Social Sciences (SPSS 16.0) (https://www.ibm.com/analytics/data-science/predictive-analytics/spss-statistical-software) computer programs.

DNA samples were extracted and purified from single seeds of individual cultivars using commercial kits (Qiagene, CA, USA). The DNA concentration for each sample was adjusted to 50 ng/µl. All samples were genotyped using the soybean 5403 SNP Illumina iSelect array (Song et al., 2013) at the Traitgenetics GmbH (Gatersleben, Germany). The Illumina Infinium procedure was performed according to the manufacturer's protocol. SNP genotype analysis was carried out using the Illumina Genome Studio software (GS V2011.1). Population genetic analysis and principal coordinate analysis were performed using GenAlEx 6.5 (Peakall, Smouse, 2012).

The SNP dataset was filtered using a 10 % cutoff for missing data and markers with minor allele frequency >0.10 were considered for GWAS. Numbers of hypothetical groups ranging from $k = 1$ to 10 were assessed using 50,000 burn-in iterations followed by 100,000 recorded Markov-Chain iterations. To estimate the sampling variance of population structure inference, five independent runs were carried out for each k by the STRUCTURE software (Pritchard et al., 2000). The output from STRUCTURE was analyzed for delta K value (ΔK) in STRUCTURE HARVESTER (Evanno et al., 2005). On the basis of the final k values, Q-matrix for three identified clusters was developed. GWAS for resistance to the most harmful fungal diseases of soybean in South East and South Kazakhstan were studied using 4,442 SNP filtered against minor alleles. GWAS based on the MLM model, including options with Q and K matrices, was conducted using the TASSEL 5 software (Bradbury et al., 2007). Pairwise LD between markers was measured using linkage disequilibrium parameter (r^2) between alleles using R studio (Wimmer et al., 2012). The LD decay rate was estimated as the chromosomal distance at which the average pairwise correlation coefficient (r^2) dropped to a half of its maximum value (Wimmer et al., 2012). The GWAS for resistance to diseases and spread of the diseases during the plant growth was run separately, and as the first evaluation was marked as FUS1, FLS1, and BS1, the evaluations for the spread of the diseases during the plant growth were marked as FUS2, FLS2, and BS2.

Results

Diseases resistant
Field trial results at the experimental stations of the South East and South regions suggested a clear difference in the development of FUS, FLS, and BS on the leaf surfaces at the adult stage of the plant growth. While results at the RIBSP (South
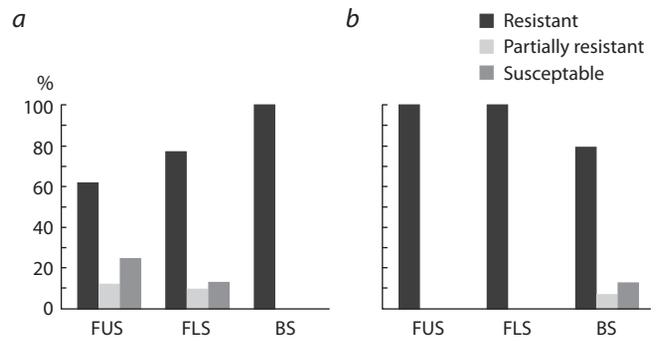

Fig. 1. Resistance of the soybean collection to three studied fungal diseases in the South and South East regions of Kazakhstan:

a – experimental plot of Research Institute of Biological Safety Problems; *b* – experimental plot of Kazakh Research Institute of Agricultural Plant Production.

region) showed the occurrence of FUS and FLS, and luck of BS symptoms, the data at the KRIAPP (South East region) allowed the identification of only BS development (Fig. 1). At the RIBSP site, FUS showed stronger influence on plant growth as only 62.1 % of plants were resistant, whereas 77.4 % was resistant to FLS, and 100 % resistant to BS (see Fig. 1, *a*). At KRIAPP the collection showed 79.3 % resistance to BS with no signs of FUS and FLS throughout the plant growth period (see Fig. 1, *b*).

Three-way ANOVA suggested that the origin of the accessions was significantly associated with reducing of YP in relationship to the development of FUS and FLS in South Kazakhstan (Table 1). Among tested accessions, there were clear examples of association between the high resistance and yield. For instance, cultivar (cv.) 'Santana' from France was highly resistant to FUS and had high YP (8.7 ± 0.26 g), while cv. 'Chernovickaya 7' from Ukraine was highly susceptible to FUS with the YP of only 1.8 ± 0.15. The collection included ten accessions from East Asia, which showed complete resistance to all three diseases.

Phenotypic variation of the collection
Comparative assessment of five groups of samples in the studied soybean collection in two regions for 2016–2017 has not revealed sharp differences in the main agronomic traits. The varieties from North America and West Europe showed good potential in NSP, while the varieties from East Asia – the highest potential in TSW. Pearson correlation demonstrated that NSP negatively correlated with TSW ($p < 0.001$) and positively – with YP ($p < 0.001$). Overall highest average yield in the collection of 182 accessions for the two fields was recorded for the Supra variety from Canada (23.0 ± 3.86), followed by Cheremosh (18.7 ± 2.89) from Ukraine and Slaviya (19.3 ± 5.42) from Russia. Among the accessions from Kazakhstan, the Mysula variety showed the best YP (18.3 ± 3.45). It is interesting that in 2016, 14.8 % of the collection showed a higher yield result in comparison with standard variety Zhansaya, and in 2017 the number of higher yield accessions was even bigger (54.4 %).

Multivariate ANOVA suggested that FUS and FLS affected each studied trait both in RIBSP and KRIAPP, except FLS was not a factor for variation in TSW (see Table 1). On the other

Полногеномный анализ ассоциаций с устойчивостью к грибным
болезням в коллекции сои в условиях Юго-Восточного и Южного
Казахстана

А. Затыбеков, С. Абугалиева
С. Дидоренко, А. Рсалиев, Е. Туруспеков

2018
22•5

Table 1. Multivariate ANOVA for the main agronomic traits by SPSS

Factors	Traits	df	FUS	FLS	BS	Factors	Traits	df	FUS	FLS	BS
Resistance	NSP	8	***	***	ns	Origin	NSP	44	**	*	ns
	TSW	8	*	ns	ns		TSW	44	ns	ns	ns
	YP	8	***	***	ns		YP	44	*	*	ns
	VER2	8	**	*	ns		VER2	44	ns	ns	ns
	VER8	8	***	**	**		VER8	44	ns	*	ns
Region	NSP	17	ns	ns	ns	Resistance × Region × Origin	NSP	89	ns	ns	ns
	TSW	17	ns	ns	ns		TSW	89	ns	ns	ns
	YP	17	ns	**	ns		YP	89	ns	ns	ns
	VER2	17	ns	ns	ns		VER2	89	ns	ns	ns
	VER8	17	*	ns	ns		VER8	89	ns	ns	ns

N o t e : df – degree of freedom. The F values are provided with significance level indicated by the asterisks. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ns – not significant.

Table 2. Mean genetic diversity indexes in five soybean groups based on 4442 SNPs

Population	East Europe	West Europe	East Asia	North America	Kazakhstan
N	83	24	10	32	33
Ne	1.93 ± 0.006	1.72 ± 0.005	1.66 ± 0.006	1.67 ± 0.005	1.68 ± 0.005
I	0.76 ± 0.003	0.6 ± 0.003	0.55 ± 0.004	0.58 ± 0.003	0.58 ± 0.004
h	0.45 ± 0.002	0.40 ± 0.002	0.36 ± 0.003	0.38 ± 0.002	0.39 ± 0.002

N o t e : N – number of accessions; Ne – number of effective alleles; I – Shannon index; h – Nei's diversity index.

hand, BS significantly affected the duration of plant growth at the VER8 stage, the only one out of the eight studied traits (see Table 1).

Genetic variation in the soybean collection based on SNP markers

Genotyping of the soybean collection using the Illumina iSelect SNP array revealed 5403 successful SNPs (74.03 % success) with 77.98 % variants being transitions and 22.01 % – transversions. The final data consisted of 4442 polymorphic SNPs spanned on 20 chromosomes with the average length of 47.4 Mb and the average number of SNPs per chromosome of 222.1. The number of markers per chromosome varied from 163 in Gm11 to 286 in Gm13 with the chromosome length ranging from 37.3 Mb in Gm16 to 62.2 Mb in Gm18. The average density of SNP map was one marker per every 213 Kb. The LD decay curve at the threshold $r^2 = 0.1$ was 20 Kbp (Supplemental Fig. 1)[1]. The PCoA allowed to separate 182 accessions based on their breeding origin and were split into five geographically distinct groups (see Supplemental Fig. 2). The smallest group was from East Asia (10 accessions), and the largest – from Eastern Europe (83 accessions, see Table 2). The PCoA analysis based on NeiP data showed that genotypes from two European groups were positioned separately from other three groups by the PCoA1 component (see Supplemental Fig. 2), while PCoA2 effectively separated the remaining three groups. The accessions from North

America and Kazakhstan appeared to be the most close groups, while the accessions from East Asia were genetically more distant from the other four groups (see Supplemental Fig. 2).

Association mapping

A total of 9 SNPs for 15 MTAs at the two stages of plant growth were identified to be associated with the resistance to three fungal diseases (Table 3). For each MTA separate QQ plots were generated to validate the significance of the associations. In addition, the results were statistically validated using a t-test for identification of a false positive MTA. All 15 identified MTAs were significant after t-test application. The results suggested that two MTAs were significant for resistance to FUS, five – for FLS, and two – for BS (see Table 3). Also, the physical position of each critical SNP marker was compared with positions of known QTLs (https://soybase.org/search/qtllist_by_symbol.php). In this study only two out of nine SNPs matched the positions of analogous QTL in a soybean genome (see Table 3).

The largest number of SNP markers in identified MTAs were located in chromosome 2, where mainly QTLs for resistance to *P. sojae* were genetically mapped (Fig. 2). The analysis of genome physical locations of associated SNP markers revealed that 3 SNPs were part of CDS (coding DNA sequence) and remained 6 SNPs were located in intergenic regions (Table 4).

Each SNP in intergenic position was considered for possible functional annotation based on the actual proximity of nearby located genes.

---

[1] Supplementary Materials are available in the online version of the paper: http://www.bionet.nsc.ru/vogis/download/pict-2018-22/appx9.pdf

Table 3. List of MTAs identified using the TASSEL software and matching with locations of corresponding QTL available in the Soybase database (https://soybase.org)

Disease stages	SNP ID	Chr	Position	Allele	Allele freq	p-value	Add	r^2, %	Suggested QTL ID in this study	Known QTLs*
FUS1	Gm13.20484995	13	20484995	G/T	43/134	2.38E−4	1.9	8.1	q*Fus.* spp 13-1	Fusarium lesion length 1-2, Phytoph 9-2
	Gm17.8109237	17	8109237	A/C	134/41	7.54E−5	−2.1	9.4	q*Fus.* spp 17-1	SCN 23-2
FUS2	Gm13.20484995	13	20484995	G/T	43/134	5.05E−4	1.9	7.3	q*Fus.* spp 13-2	Fusarium lesion length 1-2, Phytoph 9-2
	Gm17.8109237	17	8109237	A/C	134/41	1.40E−5	−2.6	11.4	q*Fus.* spp 17-2	SCN 23-2
FLS1	Gm02.9039246	2	9039246	C/T	72/104	1.52E−4	0.4	8.6	q*Cer.s* 2-1-1	Bean pyralid 1-2
	Gm02.10140292	2	10140292	A/G	84/87	1.93E−4	−2.6	8.5	q*Cer.s* 2-1-2	
	Gm07.36875730	7	36875730	C/T	48/130	8.97E−4	1.4	6.6	q*Cer.s* 7-1	SCN 40-4
	Gm15.1007132	15	1007132	A/C	75/103	7.40E−4	13.5	6.6	q*Cer.s* 15-1	Fusarium lesion length 1-3
FLS2	Gm02.6596937	2	6596937	G/T	38/130	8.32E−4	1.1	6.9	q*Cer.s* 2-2-1	Phytoph 14-4, Bean pyralid 1-6
	Gm02.9039246	2	9039246	C/T	72/104	8.78E−5	6.2	9.3	q*Cer.s* 2-2-2	Bean pyralid 1-2
	Gm02.10140292	2	10140292	A/G	84/87	1.69E−4	−4.4	8.6	q*Cer.s* 2-2-3	
	Gm07.36875730	7	36875730	C/T	48/130	7.69E−4	15.2	6.8	q*Cer.s* 7-2	SCN 40-4
	Gm15.1007132	15	1007132	A/C	75/103	9.52E−4	13.5	6.3	q*Cer.s* 15-2	Fusarium lesion length 1-3
BS1	Gm14.4811528	14	4811528	A/C	87/87	7.23E−4	3.2	6.9	q*Sep.g*14-1	SDS 14-10, Sclero 8-2
	Gm17.12684761	17	12684761	C/T	109/62	4.19E−4	−3.5	7.7	q*Sep.g*17-1	

* Based on the QTL list on SoyBase (https://soybase.org/search/qtllist_by_symbol.php).
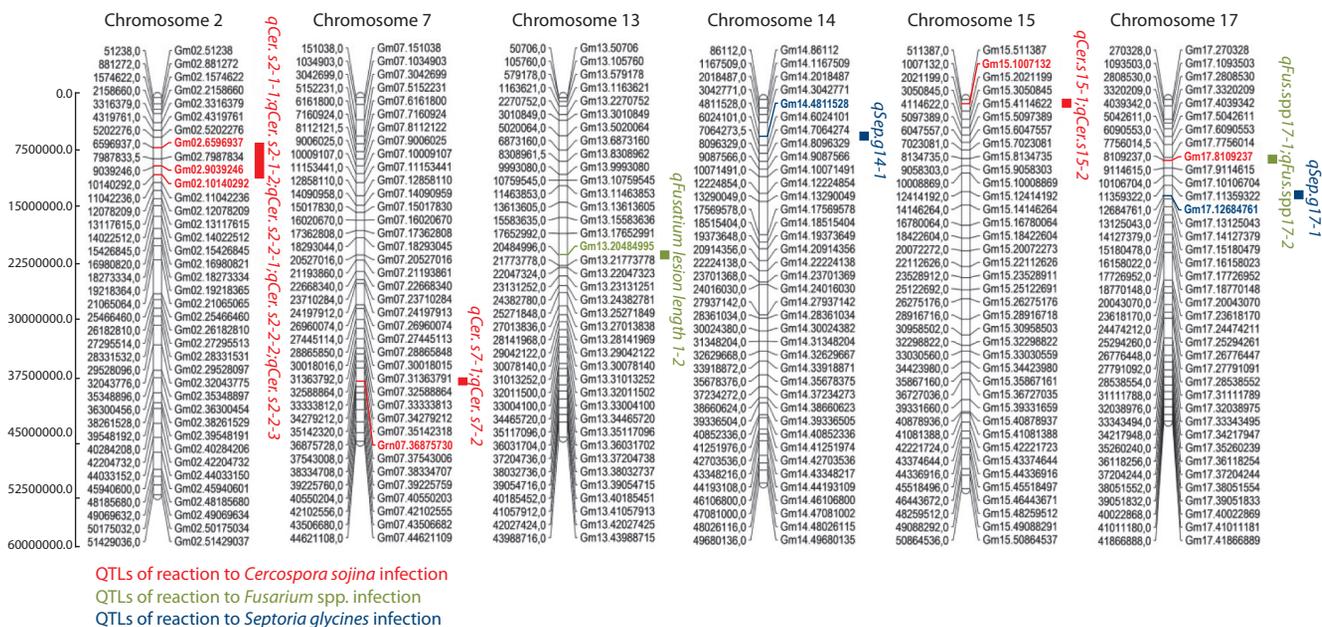

Fig. 2. Genetic map of identified SNP markers in identified MTAs for resistance to the three diseases analyzed in soybean population.

Discussion

The analysis of three diseases from two regions of Kazakhstan revealed strong environmental influence on plant tolerance to studied pathogens, as FUS and FLS were the factor in the South, and BS – in South East parts of the country, respective-ly. ANOVA suggested that FUS and FLS affected nearly all studied traits, except for TSW, which was a factor in case of FUS, but not in case of FLS (see Table 1). The test also sug-gested that the origin of the plant material was essential for NSP and YP in both FUS and FLS studies. A different outcome

Полногеномный анализ ассоциаций с устойчивостью к грибным болезням в коллекции сои в условиях Юго-Восточного и Южного Казахстана

А. Затыбеков, С. Абугалиева
С. Дидоренко, А. Рсалиев, Е. Туруспеков

2018
22 • 5

Table 4. Physical positions of identified SNPs in the soybean genome

Pathogenes of three diseases	SNP ID	Positions in genome	Putative gene	Annotation of putative gene
C. sojina	Gm02.6596937	Intron	Glyma02g08400	Small heat-shock protein (hsp20) family
C. sojina	Gm02.9039246	Intron	Glyma02g10920	Unknown protein
C. sojina	Gm02.10140292	CDS	Glyma02g11935	60S ribosomal protein L34
C. sojina	Gm07.36875730	Intron	Glyma07g31910	Mitochondrial carnitine-acylcarnitine carrier protein
Fusarium spp.	Gm13.20484995	CDS	Glyma13g16550	Domain of unknown function (DUF3511)
S. glycines	Gm14.4811528	Intron	Glyma14g06580	Serine/threonine protein kinase
C. sojina	Gm15.1007132	Intron	Glyma15g01550	AUX/IAA family
Fusarium spp.	Gm17.8109237	CDS	Glyma17g10780	Thyroid receptor interacting protein related
S. glycines	Gm17.12684761	Intron	Glyma17g15990	Vesicle transport v-snare protein vti1-related

was observed in BS analysis, as this disease affected plants only for the duration of the VER8 stage (see Table 1), which shows that the *S. glycines* did not affect plants in RIBSP and was barely important in KRIAPP. It is interesting that all ten studied accessions from East Asia and selected lines from Europe showed strong tolerance to all three diseases and could be directly involved in the breeding processes of soybean for resistance to studied fungi diseases.

On the other hand, the genetic study of the collection based on 4,442 polymorphic SNPs indicated a relatively close genetic relationship between the samples from Kazakhstan and North America, as PCoA test placed them together in left upper part of the graph (see Supplemental Fig. 2).

GWAS of the three diseases evaluated at the two stages of plant growth period allowed for identification of nine SNP markers associated with 15 MTAs (see Tables 3, 4). Two SNPs were identified in the GWAS of FUS on chromosomes 13 and 17 (see Fig. 2). The region on chromosome 13 matched with the well-known QTL (Fusarium lesion length 1–2) identified by M. Ellis et al. (2012). The authors had found that the region between the Satt160 and Satt149 markers was significantly associated with resistance to *Fusarium graminearum*. M. Kassem et al. (2006) reported that the Satt160 marker on chromosome 13 appeared to be a significant determinant of seed yield. It is interesting that this region was also associated with resistance to *P. sojae* (Wang H. et al., 2010). The region on the chromosome 17 has the same location with a QTL identified in GWAS for resistance to *Fusarium virguliforme* (Bao et al., 2015; Zhang et al., 2015). SNP marker Gm17.8109237 identified in this study has located approximately 4 Mb from SNP marker ss715611120_C_T identified by J. Zhang et al. (2015) and 6.7 Mb from SNP marker BARC-051665-11191 identified by Y. Bao et al. (2015).

The most significant amount of MTAs was found in GWAS for resistance to FLS (see Tables 3, 4). The SNP locations of identified five QTLs for FLS were mapped on chromosomes 2, 7, and 15, and did not match locations of the QTLs found in the previous study for resistance to this disease (Yang W. et al., 2001; Pham et al., 2015). A literature survey showed that one QTL for resistance to FLS matched the QTL previoulsy mapped on chromosome 13 (Pham et al., 2015), while another

was positioned on chromosomes 16 (Yang W. et al., 2001). Therefore, the MTA found in this study presumably suggested that they are novel QTL for resistance to FLS.

In case of BS, the location of one out of two identified QTLs has matched the same region on chromosomes 14 with a QTL for resistance to sudden death syndrome (SDS) caused by *F. virguliforme* (Anderson et al., 2015). A physical position of associated SNP Gm14.4811528 for this QTL was in proximity of candidate gene Glyma14g06580 (Schmutz et al., 2010). The annotation of the gene is suggesting that it is a serine/threonine protein kinase, which is a common genetic factor often involved in controlling soybean diseases resistance (Cook et al., 2012). However, the identified position of the second MTA on chromosome 17 has possibly been reported for the first time. Therefore, in this study soybean QTL for resistances to FLS and BS were presumably novel ones. As the analyzed population in two regions has shown different reaction to tested diseases, these findings underline the importance of studying a genetically diverse collection of a particular soybean growing in a certain environmental niche. Identified MTAs may facilitate the discovery of new genes for resistance to diseases and a better understanding of genotype × environment interaction patterns. Also, the size and level of genetic variation in the studied genetic panels appear to be critical for the positive outcome of GWAS-based projects. It has been demonstrated that experiments with a sample size less than 384 accessions (Gurung et al., 2014) and large LD blocks (Zanke et al., 2014) might lead to identification of false positive associations. On the other hand, in the study by M.K. Turner et al. (2017), it was shown that smaller panels might allow for detection of false negative associations that would not have been detected in more extensive panels (Oyiga et al., 2017). Therefore, the results of this study using relatively small soybean collection size (*n* = 182) may potentially relate to the above-mentioned findings by M.K. Turner and his coauthors.

With development of new genomic technologies, such as KASP (kompetitative allele-specific polymorphism) (Semagn et al., 2014), the designated SNP markers (see Tables 3, 4) for each of the identified MTAs for resistances to FUS, FLS and BS can be transformed into convenient types of DNA markers to enhance marker-associated selection projects in soybean.

Thus, the results of this study are a further contribution to the genetics and breeding of soybean associated with resistance to main fungal diseases.

Acknowledgements

Conflict of interest

The authors declare they have no conflict of interest.

References

Abugalieva S., Didorenko S., Anuarbek S., Volkova L., Gerasimova Y., Sidorik I., Turuspekov Y. Assessment of soybean flowering and seed maturation time in different latitude regions of Kazakhstan. PLoS One. 2016;11(12):e0166894. DOI 10.1371/journal.pone.0166894.

Anderson J., Akond M., Kassem M., Meksem K., Kantartzi S. Quantitative trait loci underlying resistance to sudden death syndrome (SDS) in MD96-5722 by 'Spencer' recombinant inbred line population of soybean. 3 Biotech. 2015;5(2):203-210. DOI 10.1007/s13205-014-0211-3.

Bao Y., Kurle J.E., Anderson G., Young N.D. Association mapping and genomic prediction for resistance to sudden death syndrome in early maturing soybean germplasm. Mol. Breeding. 2015;35:128. DOI 10.1007/s11032-015-0324-3.

Bradbury P.J., Zhang Z., Kroon D.E., Casstevens T.M., Ramdoss Y., Buckler E.S. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23:2633-2635. DOI 10.1093/bioinformatics/btm308.

Contreras-Soto R.I., Mora F., Lazzari F., Rott de Oliveira M.A., Scapium C.A., Chuster I. Genome-wide association mapping for flowering and maturity in tropical soybean: implications for breeding strategies. Breed. Sci. 2018;67:435-449. DOI 10.1270/jsbbs.17024.

Cook D.E., Lee T.G., Guo X., Melito S., Wang K., Bayless A.M., Wang J., Hughes T.J., Willis D.K., Clemente T.E., Diers B.W., Jiang J., Hudson M.E., Bent A.F. Copy number variation of multiple genes at *Rhg*1 mediates nematode resistance in soybean. Science. 2012;338:1206-1209. DOI 10.1126/science.1228746.

Copley T.R., Duceppe M.O., O'Donoughue L.S. Identification of novel loci associated with maturity and yield traits in early maturity soybean plant introduction lines. BMC Genomics. 2018;19:167. DOI 10.1186/s12864-018-4558-4.

Dias M.D., Pinheiro V.F., Café-Filho A.C. Impact of anthracnose on the yield of soybean subjected to chemical control in the north region of Brazil. Summa Phytopathol. 2016;42(1):18-23. DOI 10.1590/0100-5405/2114.

Didorenko S.V., Sagitov A.O., Kudaibergenov M.S. Main diseases on crops of soybean and methods of dealing with them. AgroAlem. 2014;8(61):42-46. (in Russian)

Ellis M., Wang H., Paul P., St. Martin S.K., McHale L., Dorrance A. Identification of soybean genotypes resistant to *Fusarium graminearum* and genetic mapping of resistance quantitative trait loci in the cultivar Conrad. Crop Sci. 2012;52(5):2224-2233.

Evanno G., Regnaut S., Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 2005;14:2611-2620. DOI 10.1111/j.1365-294X.2005.02553.x.

Faske T., Kirkpatrick T., Zhou J., Tzanetakis I. Soybean diseases. Arkansas Soybean Production Handbook. 2014;11:1-18.

Gurung S., Mamidi S., Bonman J.M., Xiong M., Brown-Guedira G., Adhikari T.B. Genome-wide association study reveals novel quantitative trait loci associated with resistance to multiple leaf spot diseases of spring wheat. PLoS One. 2014;9(9):e108179. DOI 10.1371/journal.pone.0108179.

Hnetkovsky N., Chang S.J.C., Doubler T.W., Gibson P.T., Lightfoot D.A. Genetic mapping of loci underlying field resistance to soybean sudden death syndrome (SDS). Crop Sci. 1996;36:393-400. DOI 10.2135/cropsci1996.0011183X003600020030x.

Iquira E., Humira S., Francois B. Association mapping of QTLs for sclerotinia stem rot resistance in a collection of soybean plant introductions using a genotyping by sequencing (GBS) approach. BMC Plant Biol. 2015;15:5. DOI 10.1186/s12870-014-0408-y.

Kassem M.A.J., Shultz K., Meksem Y., Cho A.J., Wood M.J., Iqbal D., Lightfoot A. An updated 'Essex' by 'Forrest' linkage map and first composite interval map of QTL underlying six soybean traits. Theor. Appl. Genet. 2006;113:1015-1026. DOI 10.1007/s00122-006-0361-8.

Klein R.J. Power analysis for genome-wide association studies. BMC Genetics. 2007;8:58. DOI 10.1186/1471-2156-8-58.

Kurilova D.A. The harmfulness of soybean fusarium depending on the degree of damage to the plants. Maslichnye Kultury. Nauchno-tekhnicheskiy Byulleten Vserossiyskogo NII Maslichnyh Kultur = Oilseeds. Scientific and Technical Bulletin of the All-Russian Research Institute of Oilseeds. 2010;2(144-145):84-89. (in Russian)

Leath S., Carroll R.B. Screening for resistance to *Fusarium oxysporum* in Soybean. Plant Dis. 1982;66(12):1140-1143.

Masuda T., Goldsmith P.D. World soybean production: area harvested, yield, and long-tern projections. Int. Food Agribus. Man. Rev. 2009; 12(4):143-161.

Mian M.A.R., Wang T., Phillips D.V., Alvernaz J., Boerma H.R. Molecular mapping of the *Rcs*3 gene for resistance to frogeye leaf spot in soybean. Crop Sci. 1999;39:1687-1691. DOI 10.1046/j.1439-0523.2001.00563.x.

Mian R., Bond J., Joobeur T., Mengistu A., Wiebold W., Snannon G., Wrather A. Identification of soybean genotypes resistant to *Cercospora sojina* by field screening and molecular markers. Plant Dis. 2009;93:408-411. DOI 10.1094/PDIS-93-4-0408.

Mombekova G.A., Shemshurova O.N., Seitbattalova A.I., Aitkhozhina N.A., Bekmakhanova N.E. Phytopathogens of sugar beet and soybean cultivated in soil and climatic conditions of Almaty region. NAN RK. 2013;4:8-11. (in Russian)

Mueller D., Robertson A., Sisson A., Tylka G. Soybean Diseases. Iowa State Univ. of Sci. and Technol., 2010.

Oyiga B.C., Sharma R.C., Baum M., Ogbonnaya F.C., Léon J., Ballvora A. Allelic variations and differential expressions detected at quantitative trait loci for salt stress tolerance in wheat. Plant Cell Environ. 2017;41(5):919-935. DOI 10.1111/pce.12898.

Pataky J.K., Lim S.M. Effects of septoria brown spot on the yield components of soybeans. Plant Dis. 1981;65:588-590.

Peakall R., Smouse P.E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics. 2012;28:2537-2539. DOI 10.1093/bioinformatics/bts460.

Pham A.T., Harris D.K., Buck J., Hoskins A., Serrano J., Abdel-Haleem H. Fine mapping and characterization of candidate genes that control resistance to *Cercospora sojina* K. Hara in two soybean germplasm accessions. PLoS One. 2015;10(5):e0126753. DOI 10.1371/journal.pone.0126753.

Polozhieva Y.V., Dubovitskaya L.K. Soybean varieties evaluation on prevalence by complex of the root rots agents. Dalnevostochnyi Agrarnyi Vestnik = Far Eastern Agrarian Herald. 2015;3(35):35-38. (in Russian)

Prabhu R.R., Njiti V.N., Bell-Johnson B., Johnson J.E., Schmidt M.E., Klein J.H., Lightfoot D.A. Selecting soybean cultivars for dual resistance to soybean cyst nematode and sudden death syndrome using two DNA markers. Crop Sci. 1999;39(4):982-987.

Pritchard J.K., Stephens M., Rosenberg N.A., Donnelly P. Association mapping in structured populations. Am. J. Hum. Genet. 2000;67:170-181. DOI 10.1086/302959.

Qin J., Song Q., Shi A., Li S., Zhang M., Zhang B. Genome-wide association mapping of resistance to *Phytophthora sojae* in a soybean [*Glycine max* (L.) Merr.] germplasm panel from maturity groups IV and V. PLoS One. 2017;12:e0184613. DOI 10.1371/journal.pone.0184613.

Полногеномный анализ ассоциаций с устойчивостью к грибным
болезням в коллекции сои в условиях Юго-Восточного и Южного
Казахстана

А. Затыбеков, С. Абугалиева
С. Дидоренко, А. Рсалиев, Е. Туруспеков

2018
22•5

Schmutz J., Cannon S.B., Schlueter J., Ma J., Mitros T., Nelson W., Hyten D.L., Song Q., Thelen J.J., Cheng J., Xu D., Hellsten U., May G.D., Yu Y., Sakurai T., Umezawa T., Bhattacharyya M.K., Sandhu D., Valliyodan B., Lindquist E., Peto M., Grant D., Shu S., Goodstein D., Barry K., Futrell-Griggs M., Abernathy B., Du J., Tian Z., Zhu L., Gill N., Joshi T., Libault M., Sethuraman A., Zhang X.C., Shinozaki K., Nguyen H.T., Wing R.A., Cregan P., Specht J., Grimwood J., Rokhsar D., Stacey G., Shoemaker R.C., Jackson S.A. Genome sequence of the paleopolyploid soybean. Nature. 2010;463:178-183. DOI 10.1038/nature08670.

Schneider R., Rolling W., Song Q., Cregan R., Dorrance A.E., McHale L.K. Genome-wide association mapping of partial resistance to *Phytophthora sojae* in soybean plant introductions from the Republic of Korea. BMC Genomics. 2016;17(1):607. DOI 10.1186/s12864-016-2918-5.

Semagn K., Babu R., Hearne S., Olsen M. Single nucleotide polymorphism using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Mol. Breeding. 2014;33(1):1-14.

Song Q., Hyten D.L., Jia G., Quigley C.V., Fickus E.W., Nelson R.L., Cregan P.B. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS One. 2013;8(1):e54985. DOI 10.1371/journal.pone.0054985.

Stacey G. (Ed.) Genetics and Genomics of Soybean. Ser.: Plant Genetics and Genomics: Crops and Models (Vol. 2). Springer, 2008.

Turner M.K., Kolmer J.A., Pumphrey M.O., Bulli P., Chao S., Anderson J.A. Association mapping of leaf rust resistance loci in a spring wheat core collection. Theor. Appl. Genet. 2017;130:345-361. DOI 10.1007/s00122-016-2815-y.

Vidic M., Dordevic V., Petrovic K., Miladinovic J. Review of soybean resistance to pathogens. Ratar. Povrt. 2013;50(2):52-61. DOI 10.5937/ratpov50-4038.

Wang H., Waller L., Tripathy S., St. Martin S.K., Zhou L., Krampis K., Tucker D.M., Mao Y., Hoeschele I., Maroof S.M.A., Tyler B.M., Dorrance A.E. Analysis of genes underlying soybean quantitative trait loci conferring partial resistance to *Phytophthora sojae*. Plant Gen. 2010;3:23-40.

Wang J., Liu C.,Wang J., Qi Z., Li H., Hu G., Chen Q. An integrated QTL map of fungal diseases resistance in soybean (*Glycine max* L. Merr.): a method of meta-analysis for mining R genes. Agric. Sci. China. 2010;9(2):223-232. DOI 10.1016/S1671-2927(09)60087-0.

Wimmer V., Albrecht T., Auinger H., Schoen C. Synbreed: a framework for the analysis of genomic prediction data using R. Bioinformatics. 2012;28(15):2086-2087. DOI 10.1093/bioinformatics/bts335.

Yang W., Weaver D.B., Nielsen B.L., Qiu J. Molecular mapping of a new gene for resistance to frogeye leaf spot of soyabean in Peking. Plant Breed. 2001;120(1):73-78. DOI 10.1046/j.1439-0523.2001.00563.x.

Yang X.B., Feng F. Ranges and diversity of soybean fungal diseases in North America. Phytopathology. 2001;91(8):769-775.

Zanke C., Ling J., Plieske J., Kollers S., Ebmeyer E., Korzun V., Argillier O., Stiewe G., Hinze M., Beier S., Ganal M.W., Roder M.S. Genetic architecture of main effect QTL for heading date in European winter wheat. Front. Plant Sci. 2014;5:217. DOI 10.3389/fpls.2014.00217.

Zaostrovnykh V.I. Soybean diseases. Zaschita i Karantin Rastenii = Plant Protection and Quarantine. 2005;2:49-53. (in Russian)

Zatybekov A., Abugalieva S., Didorenko S., Gerasimova Y., Sidorik I., Anuarbek Sh., Turuspekov Y. GWAS of agronomic traits in soybean collection included in breeding pool in Kazakhstan. BMC Plant Biol. 2017;17(Suppl.4):63-70. DOI 10.1186/s12870-017-1125-0.

Zhang J., Singh A., Mueller D.S., Singh A.K. Genome-wide association and epistasis studies unravel the genetic architecture of sudden death syndrome resistance in soybean. Plant J. 2015;84:1124-1136. DOI 10.1111/tpj.13069.

Zhang J., Xia Ch., Wang X., Duan C., Sun S., Wu X., Zhu Zh. Genetic characterization and fine mapping of the novel Phytophtora resistance gene in a Chines soybean cultivar. Theor. Appl. Genet. 2013; 126:1555-1561.

Zhang L., Kyei-Boahen S., Zhang J., Zhang M., Freeland T., Watson C., Liu X. Modifications of optimum adaptation zones for soybean maturity groups in the USA. Crop Management. 2007;6(1). DOI 10.1094/CM-2007-0927-01-RS.

Zhao G., Ablett G.R., Anderson T.R., Rajcan I., Schaafsma A.W. Inheritance and genetic mapping of resistance to rhizoctonia root and hypocotyl rot in soybean. Crop Sci. 2005;45:1441-1447. DOI 10.2135/cropsci2004.0560.

Zhao X., Han Y., Li Y., Liu D., Sun M., Zhao Y., Lu Ch., Li D., Yang Z., Huang L., Teng W., Qiu L., Zheng H., Li W. Loci and candidate gene identification for resistance to *Sclerotinia sclerotiorum* in soybean (*Glycine max* L. Merr.) via association and linkage maps. Plant J. 2015;82:245-255. DOI 10.1111/tpj.12810.