Selection of drought-tolerant and drought-sensitive lines (*Glycine max*) using principal component analysis

Mohammad Hassanvand[1], Mahdi Changizi[1]*, Shahab Khaghani[1], Masoud Gomarian[2], Ezatollah Sedaghatfar[2]

1. Department of Genetic and Plant Breeding, Arak Branch Islamic Azad University, Arak, Iran
2. Department of Plant Protection Arak Branch, Islamic Azad University, Arak, Iran

Abstract

In a study to select sensitive and tolerant soybean lines to drought stress and superior genotypes, 80 soybean lines were obtained from crosses of Karbin × Fora cultivars along with L.17 cultivar as control in the Lurestan-Saraab Research Farm during crop rotation 2016-2017. The cultivars were evaluated in the form of a lattice design 9×9 with two replications. The results of the analysis of variance showed significant differences between the control and experimental lines and also between the blocks for different traits. According to the results of stepwise regression and principal component analysis (PCA), traits were grouped into four components, accounting for about 62% of the total variation. The evaluation of grain yield potential and accurate PCA evaluation of the traits led to the selection of superior lines in the experiment. Results showed that plant height, number of pods per plant, number of seeds per pod, and 100-seed weight can be used as the criteria for selection of soybean crops to improve seed yield in drought stress condition.
Keywords: drought stress, selection criteria, soybean cultivars, stepwise regression

Hassanvand, M., M. Changizi, Sh. Khaghani, M. Gomarian, and **E. Sedaghatfar. 2022.** 'Selection of drought-tolerant and drought-sensitive lines (Glycine max) using principal component analysis'. *Iranian Journal of Plant Physiology 12* (2), 4145-4151.

Introduction

Oily grains constitute the second largest world food reserves after grains. Soybean is one of the most important oily seeds in the world with the highest production in the world. It contains protein in addition to rich fatty acids. In fact, soybean accounts for more than half of the global production of oilseed and more than a quarter of the protein used globally for human food and animal feed (Wang et al., 2018).

In Iran, this plant has been assigned the second rank in terms of production (Shariati et al., 1379). Therefore, the production of high yield resistant cultivars is very important and can be of great help to agriculture and the national economy. An important goal in soybean rearing is producing cultivars with higher seed yields, to interact with each other to increase the yield of high-yielding genotypes, and then to select high-yielding

* *Corresponding Author*
E-mail Address: mchangizi47@gmail.com
Received: June, 2021
Accepted: November, 2021

Table 1
ANOVA of the different traits in soybean Karbin × Fora cultivars

		Mean of squares							
S.O.V	df	Percent Germination	Color Flower	Plant height (centimeter)	Number Sub branch	Day before flowering	day maturity	Birth rate (Day)	Grain filling period (Day)
Repeat	1	3.85^{*}	0.22^{ns}	29389.65^{**}	83.06^{ns}	6.72^{ns}	7.13^{*}	7.56^{*}	6.72^{**}
Untreated treatment	80	10.8^{ns}	0.4^{ns}	91.44^{*}	0.39^{ns}	0.85^{ns}	0.94^{**}	0.83^{**}	0.81^{**}
Block in Repeat	16	10.45	0.19	25.47^{*}	0.36^{ns}	0.8^{ns}	0.84^{**}	0.81^{**}	0.83^{*}
Inside block error	64	17.13	0.26	33.42	0.51	0.92	0.92	0.95	0.94
CV		15.9%	9.7%	14.2%	15.6%	20.3%	22.1%	10.9%	21.9%

		Mean squares								
S.O.V	df	Number of pods per plant	Number of seeds per pod	Number of seeds per plant	Seed yield per plant (G)	Yield in plot (G)	Grain yield per hectare (kg ha)	Dry weight whole plant (G)	Seed color	Weight of 100 seeds (G)
Repeat	1	7688^{**}	0.05^{**}	14393.39^{**}	2080.91^{**}	5790204^{**}	144755112^{**}	1701.4^{**}	0.006^{ns}	784.96^{**}
Untreated treatment	80	15.13^{*}	0.005^{**}	47.25^{*}	6.42^{**}	22496.25^{**}	562407.5^{**}	2561.85^{**}	0.09^{ns}	9.75^{**}
Block in Repeat	16	17.08^{**}	0.00025^{**}	76.35^{**}	10.41^{**}	8216.28^{**}	205407.28^{**}	2907.02^{**}	0.12^{ns}	9.73^{*}
Inside block error	64	18.66	0.0003	128.14	12.96	10208.7	255217.7	4319.03	0.15	16.97
CV		17.3%	15%	24.3%	11.8%	18.7%	18.7%	24.2%	3.5%	15.2%

* and ** show significant at p≤0.05 and p≤0.01%, respectively.

Table 2
Mean comparison of different traits between control line and Karbin × Fora cultivars

	Average Genotype	Average control line	Difference line and control line	LSD Test
Percent Germination	97.6	100	2.4	0.94
Color Flower	1.3	1	0.3	0.7
Plant height	100.4	100.2	0.2	0.83
Number Sub-branch	3	3	0	0.74
Day before flowering	48.7	48	0.7	0.8
The day maturity	128.7	128	0.7	11.23
Birth rate(Day)	73	72	1	12.8
Grain filling period (Day)	29	28	1	10.91
Number of pods per plant	35	28.4	6.6**	0.25
Number of seeds per pod	2.3	2	0.3**	0.073
Number of seeds per plant	59	66	7	0.18
Seed yield per plant (G)	15.1	14.35	0.75	0.65
Yield in the plot (G)	460	398.5	61.5	0.87
Grain yield per hectare (kg/ha)	4600	3985	615	0.86
The dry weight of the whole plant (G)	880.9	927.85	46.95	14.9
Color Seed	1.16	1	0.16	0.27
Weight of 100 seeds (G)	23.77	21.65	2.12**	0.011

LSD (P=1%)

genotypes. Selection based on yield is a difficult undertaking. This is because the attribute function is small and polygenic and is heavily influenced by environmental factors. Therefore, to improve yield, the attributes that are highly correlated with and less affected by the environment can be considered, based on which to make choices. Selection for morphological and physiological traits is easy and accurate and their inheritance is desirable. Selection based on these traits is a safe and fast way to screen vegetation communities and improve grain yield. Therefore, better control on environmental influences during breeding programs to increase grain yield can be achieved through indirect selection for attributes that are well-correlated with yield and are less affected by the environment (Calish et al., 2011). Since correlation measures the degree of genetic linkage between two or more traits, it is of particular importance in producing desirable cultivars.

Materials and Methods

This experiment was carried out in Saraab Changhaei Agricultural Research Farm in Khorramabad, Iran during 2016-2017. The farm is located at 33°, 29′ North and 48°, 18′ East, and 1175 meters above sea level with clay loamy soil texture. Land preparation operations included uniform plowing, disc, leveling, and planting rows at appropriate times. To conduct this research, 80 lines of the F8 generation soybeans were studied. The resultant cross-linking strain of inbred soybeans is due to a cross between two cultivars Karbin × Fora. Karbin is a high-yielding drought-tolerant, and Fora is a drought resistant parent cultivar. In addition to biometric study, grain yield and effective traits on yield potential in this population are underway to identify promising lines that have the privileged characteristics of the two parents.

The evaluated lines were cultivated in a simple lattice design with two replications on 2-meter lines with a spacing of 50 cm in 9 blocks of 9 with a control. Irrigation was carried out at normal intervals. During the growing season, physiological traits under study included germination percentage, flowering time, pod formation, plant height, flowering end, number of pods per plant, number of seeds per pod, pod formation, pod day ending, seed formation, seed filling, beginning of the examination. Sampling for grain yield components was considered based on 5 plants per plot. To determine the grain yield after harvest, two lines of each plot (one square meter) were harvested by hand and the seeds were weighed. After harvest, 1000 seed weight and grain yield per hectare were determined. To determine the uniformity of the land, the analysis of variance was performed for all traits related to the control of each replication. The statistical analysis of the data was based on descriptive statistics including minimum, maximum, mean, standard deviation, and diversity coefficient related to the traits. Genetic correlations were also estimated using genetic variance-covariance. To test the significance of genetic correlation coefficients, a correlation coefficient table with n-2 degrees of freedom (n = number of observations) was used. Simple correlation coefficients between the traits were calculated. Stepwise regression was used for a descriptive model between studied traits and grain yield. To eliminate the correlation between independent variables, reduction of data volume and better interpretation of relationships from factorization to factorization were used. SPSS21, SAS9, and MINITAB16 were used to perform statistical calculations.

Results

Following implementation of a lattice design and the analysis of variance for different traits, the value of each trait in the lines was reviewed and corrected as presented in Table 1. Based on the lattice model, an analysis of variance with control was used. The effect of genotype on days to maturity, planting length, number of seeds per pod, grain yield per plant, grain yield per plot, grain yield per hectare were significant ($p \leq 0.01$), while it was significant ($p \leq 0.05$) on plant height and the number of seeds per plant (Table 1). This indicates a significant difference between the lines of the populations being evaluated and the genetic variation acceptable for most of the studied traits. Based on the corrected traits, descriptive statistics analysis counting minimum, maximum, mean, normal deviation, and variation coefficients were calculated. The Table 1 shows information about the descriptive statistics of the studied character in the population as a whole as well as in the control. Among the traits, the grain yield in plots and grain yield per hectare had the highest diversity coefficient (24.1%). Therefore,

Table 3
Descriptive statistical indexes for different evaluated traits of inbreed soybean from Karbin × For a

	$\bar{X}$	$\sqrt{S^2}$	Min	Max	CV	Average control line
Percent Germination	97.6	13.45	85	100	15.9	100
Flower Color	1.3	0.36	1	2	9.7	1
Plant height	100.4	10.24	74	122	9.2	100.2
Number Sub-branch	3	10.33	1	4	15.6	3
Day before flowering	48.7	15.21	48	51	20.3	48
The day maturity	128.7	5.29	128	131	22.1	128
Birth rate(Day)	73	3.68	72	75	10.9	72
Grain filling period (Day)	29	11.50	28	31	21.9	28
Number of pods per plant	35	3.41	24	46	17.3	28.4
Number of seeds per pod	2.3	10.70	1.3	3.2	15	2
Number of seeds per plant	59	14.13	28	89	24.3	66
Seed yield per plant (G)	15.1	39.49	5.7	29	11.8	14.35
Yield in the plot (G)	460	16.57	202	586	24.2	398.5
Grain yield per hectare (kg/ha)	4600	3.54	2020	5860	22	3985
The dry weight of the whole plant (G)	880.9	122.71	373.1	1459.7	21.2	927.85
Color Seed	1.16	0.81	1	2	3.5	1
Weight of 100 seeds (G)	23.77	0.32	16.8	33.4	15.2	21.65

there are the highest diversity in comparison with other traits. The lowest diversity among the studied lines was related to plant height with a diversity coefficient of 9.2%. It should be noted that among the yield components, the number of seeds per plant showed the highest variation. Low varieties of flowers and plant height were 9.7% and 9.2%, respectively, among the offspring resulting from the cross-linking to the fact that each of the parents of the population had different characteristics. Karbin variety had a low altitude but high yield, and the flower color was related to the Fora parent. The results of the comparison of mean lines and control with LSD showed that the studied population was higher in terms of number of pods per plant, number of seeds per pod, and 100 seed weight in the control population (Table 2).

Grain yield as an economic trait is very important. This trait showed that the genotypes had an average yield of 0.460 kg/m^2. In these samples, the minimum and maximum grain yields of 0.203 kg/m^2 (line 67) and 5889 kg/m^2 (line 21) were obtained. A mean of 0.390 kg (cultivar L.17) was recoreded in the control with a minimum of 0.202 kg/m^2 and a maximum of 0.586 kg/m^2. The 80-line yield was higher than that of the control (Table 3).

To study the effect of traits on grain yield, multiple regression analysis was performed on the studied lines. The grain yield was considered as an independent variable and other components of yield were considered as dependent variables. Plant height, number of pods per plant, number of seeds per pod, number of seeds per plant, and 100-seed weight were introduced into the model. Other traits had no significant effect on the regression model. The difference between lines in terms of grain yield were attributed to differences in these traits. If the grain yield is Y and the above attributes are X1 to X5, then the regression model will be as follows:

$y = - 437 + 1.89 x1 + 3.67 x2 - 11.9 x3 - 23.7 x4 + 17.0 x5$ $R2= 81.71$

It is justified that grain yield changes are function-dependent, so that the function is predictable by predicting the yield of a hypothetical number to predict the yield in future experiments. The R2 value of the model shows that approximately 81.71% of the variance of grain yield is related to five factors, namely plant height, number of pods per plant, number of seeds per pod, number of seeds per plant, and 100- seed weight (Table 5).

PCA analysis

Table 4
Factor analysis of the evaluated traits in cultivars from Karbin × Fora

trait	Principal component Common Index			
	PCA_1	PCA_2	PCA_3	PCA_4
Percent Germination	-0.044	-0.223	-0.243	-0.844
Color Flower	0.055	-0.206	-0.226	-0.745
Plant height	0.006	-0.141	-0.161	-0.794
Number Sub-branch	-0.074	-0.016	-0.036	-0.873
Day before flowering	0.406	0.324	-0.26	-0.394
The day maturity	0.405	-0.24	-0.26	-0.395
Birth rate(Day)	0.404	-0.24	-0.26	-0.396
Grain filling period (Day)	0.404	0.341	-0.261	-0.396
Number of pods per plant	-0.042	0.02	0.125	-0.842
Number of seeds per pod	0.187	0.129	0.109	-0.613
Number of seeds per plant	-0.314	0.301	-0.321	-1.114
Seed yield per plant (G)	-0.548	-0.399	-0.149	-1.048
Yield in the plot (G)	-0.523	-0.225	-0.145	-1.03
Grain yield per hectare (kg/ha)	-0.044	-0.223	-0.143	-0.844
The dry weight of the whole plant (G)	0.055	-0.206	-0.226	-0.745
Color Seed	0.006	-0.141	-0.061	-0.794
Weight of 100 seeds (G)	-0.074	-0.016	-0.036	-0.874
Eigenvalue	3.201	2.278	1.339	1.124
Accumulative variance	22.820	41.330	53.671	62.335

Table 5
Stepwise regression analysis to determine the relative contribution of grain yield components in cultivars from Karbin × Fora

Variant	Standard Regression Index	R^2
a	-463.6	0.707
Plant height	7.477	0.706
Number of pods per plant	1.893	0.719
Number of seeds per pod	3.673	0.875
Number of seeds per plant	-11.94	0.835
Weight of 100 seeds(G)	16.962	0.726

$$Y = -a \pm bx$$

There are many advantages for multivariate statistical analysis in a deep understanding of the data structure. To find out the reason behind the correlation and the relationship between the seventeen traits examined in 80 soybeans, the fraction of the index or factors affecting these attributes were analyzed. The factors evaluated for the traits in this analysis included four main and independent factors justifying 62.4% of the total data, of which the share of the first to fourth factor was 10.71, 12.61, 17.52, and 21.56% respectively (Table 4). The first component with 22.82% of the total variation was mainly due to the grain yield of the plots and the grain yield per plant, which was referred to as the factor of making. The second factor accounted for 18.51% of the total variation and correlating to the number of seeds per plant, the period of grain filling, and the day to the beginning of flowering had positive and high effects, hence being considered as a morphological characteristic. The third factor accounted for 12.34% of the variation and was considered as an important element in the justification of germination percentage, flower color, plant height, and total physiological dry weight. In this research, grain yield assessment showed that lines with high values of the first factor had a high yield and this component can be used in indirect selection. The PCA diagram consisted of four sections, namely A, B, C, and D which had different properties if each of the lines were placed in these sections. But the lines in section C were more resistant to drought stress. Choose the line that dominates the first principal component. In section D, all lines had a low yield in the plot, so this low yield was the source of drought stress. Results of principal component analysis under stress conditions showed that Line 21 was sensitive to drought stress and Line 3 was selected for RNA extraction experiments drought stress. In the same way, the drought-resistant line that had a high yield compared to other lines (line 3) was selected as drought-resistant line (Fig. I). Line 67 and 14 had a higher yield than line 3, but since it had a shorter flowering day, it was not selected for the drought-resistant line because lines can have higher durability against drought that have longer flowering and do not lose their flowers.

Fig. II shows the Tolerance Interval Plot for Total yield of optimum condition. The average yield of all lines in these conditions is 4934 kg/ha. Compared to the average yield of Tolerance Interval Plot for Total yield of the dry condition has higher yield (Fig. I vis-à-vis Fig. II).

Discussion

Knowledge of genetic and phenotypic correlation of traits in yield can be a useful tool for increasing plant yield in the selection of superior genotypes. Research has shown a significant correlation between the number of branches per plant, the number of pods per plant, days to flowering, and grain yield per plant (Sharma et al., 1971; Amaranthath et al., 1990; Babaei et al., 2020).

A positive correlation was also reported between the number of pods per plant and soybean yield (Sengupta et al.,1972). Factor analysis is one of the multivariate statistical analysis methods for reducing the volume of data, which can be explained by the correlation between large numbers of variables in the form of a smaller number of independent factors. Effective traits are identified in each agent and the factors are named based on the most effective traits. This method makes it possible to improve the genetic factors due to their related traits (Heydari et al.,2007). This method is used to understand the

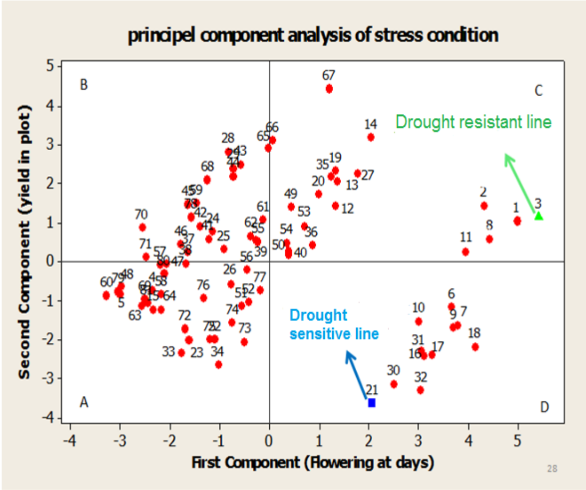

Fig. I. The principal components to select sensitive and drought-tolerant lines

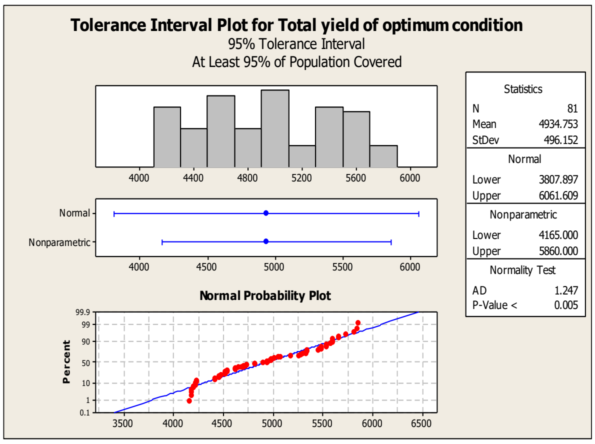

Fig. II. Statistical characteristics of the studied data under normal irrigation conditions of the studied lines

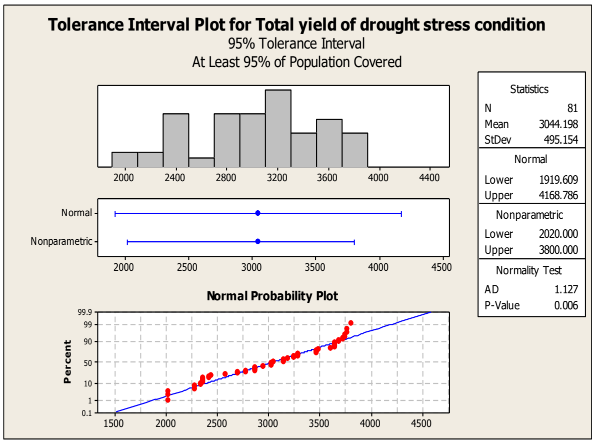

Fig. III Statistical characteristics of the studied data under drought-stress conditions in the studied lines

relationships and structure of yield components and morphological traits of crops. Walton used factor analysis to study plant characteristics and to determine the criteria for choosing suitable wheat yield under drought stress conditions (Tousi et al., 2005; Das et al., 1989). Variety of genotypes can improve the traits. Particularly, the amount of genetic diversity is effective in determining the profitability of selection (Babaei et al.,2014). Sutigihno et al. (1992) and Hwang et al. (2014) after examining seven traits in soybean reported that two traits of 100-seeds weight and number of pods per plant were important attributes in the selection of superior soybeans. Studies also suggested that 100-seed weight, number of seeds per plant, and number of pods per plant had a direct effect on yield and traits, and therefore, are important in choosing the best lines (Mishra et al., 1994; Sutigihno et al., 1992).

References

Amaranthath, K.C. And S.R.I Viswanatha. 1990. Path coefficient analysis for some quantitative characters in soybean. J. Agric. Sci. 24(3): 312-315.

Babaei H R, Razmi N, Hezarjaribi E, Hashemi Jazi M. Study on Adaptability and Grain Yield Stability of Soybean Genotypes [Glycine Max (L.) Merril] Through AMMI & GGE Biplot Analysis. jcb. 2020; 12 (35) :238-250

BABAIE ZARCH, M., & FOTOKIAN, M., & MAHMOODI, S. (2014). EVALUATION OF GENETIC DIVERSITY OF WHEAT (TRITICUM AESTIVUM L.) GENOTYPES FOR MORPHOLOGICAL TRAITS USING MULTIVARITE ANALYSIS METHODS. JOURNAL OF CROP BREEDING, 5(12), 85-98. https://www.sid.ir/en/journal/ViewPaper.aspx?id=416531

Calish, L. and S. Houshmand. 2011.Estimate of heritability and relationship of some durum wheat characters using recombination inbred lines. Electronic Journal Crop Product ,4:223-238(In Persian).

Das, M.L., A. Raman and A.J. Miah. 1989. Correlation and path coefficient and regression studies in soybean. Bangladesh J. Agric. Res. 14(1): 27-29.

Hwang, S., Ray, J.D., Cregan, P.B. et al. 2014. Genetics and mapping of quantitative traits for nodule number, weight, and size in soybean (Glycine max L.[Merr.]). Euphytica 195, 419–434.

Heydari, B., G.A. Saeidi and B.I. Seyed-Tabatabaei. 2007. Factor analysis for quantitative traits and path coefficient analysis for grain yield in wheat. Journal Science, and Technology of Agriculture and Natural Resources, 11: 135-143 (In Persian).

Mishra, A.K., S.A. Ali, R.C. Tiwary, and R.S. Raghuwanshi. 1994. Correlation and path coefficient analysis in segregating populations of Soybean. Int. J. Tropical Agric. 12: 278-281.

Sengupta, K. and Sen, S. 1972. Path-coefcient analysis of some characters influencing seed yield of soybeans (Glycine max. (L.) Merr.). Indian Agriculturist 16: 149-154

Shariati, S. and Ghazi-shahinzadeh, P. 1379. Canola. Agricultural Information Affair. 115PP. (in Persian).

Sharma, S.L. and Juneje, S.L. 1971. Correlation studies for yield and other characters in soybeans (Glycine max. (L.)Merr.). Indian Journal of Agricultural Research 1: 40-45.

Sutigihno, R. and M.S. Sudjono. 1992. Correlation and path coefficient analysis of seven quantitative traits in soybean using multiple regression programs. Penelitian Pertanian (Indonesia) Agric. Res. (Abstract) 9(1): 16-18

Tousi Mogarrad, M., M.R. Ghannadha, M. Khodarahmi and S. Shahabi. 2005. Factor analysis for grain yield and other attributes in bread wheat. Pajuhsh and Sazandegi, 67:9-16(In Persian).

Wang, M., Li, W., Fang, C. et al. Parallel 2018. selection on a dormancy gene during domestication of crops from multiple families. Nat Genet 50, 1435–1441