



Application of Multivariate Analysis for Genotype Evaluation in Soybean

V. Perić^(✉), M. Srebrić, A. Nikolić, and S. Mladenović-Drinić

Maize Research Institute “Zemun Polje”, Belgrade, Serbia
vesnaperic@yahoo.com

Abstract. Multivariate analysis is a frequently used approach in breeding studies. Applied to multiple trait data, Principal Component Analysis (PCA) gives an opportunity to graphically display the relationships between important agronomic traits, evaluate the varieties on the basis of multiple traits, identify valuable breeding sources and recommend possible selection strategies. This study was conducted to evaluate agronomic performance of 31 early varieties (maturity group 0) from the soybean collection of the Maize Research Institute “Zemun Polje” (Belgrade, Serbia). PCA analysis was applied to identify the best performing genotypes considering multiple traits (seed yield, major yield components and parameters of technological quality of grain), and to determine the level of trait interdependence. Genotypes were tested in a randomized complete block design with three replications during two years (2011 and 2012) at two locations in Serbia (Zemun Polje and Pančevo). An analysis of variance (mixed model) exhibited significant effects of genotype (G), environment (E), and genotype \times environment (G \times E) for most of the traits. According to PCA biplot, seed yield per plant was in positive correlation with number of seeds and pods per plant, indicating that those yield components could be effective selection criteria in breeding for seed yield. Comparison of early-maturing genotypes based on multiple traits revealed two varieties with above-average seed yield per plant and four varieties with high oil and above average protein content, which could be used as potential parents in breeding for important agronomic traits in future breeding programs.

Keywords: Agronomic traits · Correlations · PCA

1 Introduction

Favourable composition of soybean grain makes this crop major protein and oilseed source in many regions of the world. In Serbia, total soybean harvested area was 154249 ha (FAOSTAT 2014), mainly occupied by varieties that belong to 0, I and II maturity group. Although the seed yield is one of the most important breeding goals in soybean breeding programs, released varieties have to meet the requirements of processing industry regarding chemical composition and technological quality of grain. Development of high yielding and good grain quality cultivars is a demanding task due to the negative correlation between the protein and oil content, on one hand (Chung et al. 2003; Bellalui et al. 2009), and negative correlation between proteins and seed yield on the other (Li and Burton 2002; Taški-Ajduković et al. 2010). Thus,

identification of superior genotypes considering multiple traits is the primary goal of many breeding programs (Yan and Rajcan 2002). PCA is a useful method for evaluation of cultivars based on multiple traits, identification of those that are superior in desired traits (Mohammadi and Amri 2011), and visualization of genetic correlation among traits, as well (Yan and Rajcan 2002). The objective of this study was to identify superior genotypes as potential parents in future breeding cycles, examine the correlations between traits and determine effective selection criteria by applying the method of PCA.

2 Materials and Methods

The plant material for this study comprised 31 soybean accessions of the maturity group 0 (early maturing genotypes) from the soybean collection of the Maize Research Institute “Zemun Polje”, originated from different parts of the world. The trials were carried out during two growing seasons (2011 and 2012), at two locations in Serbia (Zemun Polje and Pancevo) and arranged according to randomized block design with 3 replications and two rows per genotype. The experimental plot size was 5 m². Standard agricultural practice was applied. The soil type is slightly calcareous chernozem. During the both research years, particularly in 2012., severe drought occurred, with total rainfalls significantly lower and average temperatures much higher than multiyear average. The samples consisted of 30 plants per genotype collected at the R8 stage. A total of eight agronomically important traits were analyzed: *PH* – plant height, *NN* – node number, *PN* – pod number, *SN* – seed number, *TSW* – 1000 seed weight, *SYP* – seed yield per plant, *PROT* – protein content, *OIL* – oil content. Seed samples were analyzed for protein and oil content (expressed as a percentage on a dry matter basis) using grain analyzer Infraneo, Chopin Technologies ®. The data recorded for each agronomic trait were subjected to analysis of variance (mixed model). PCA was performed in order to determine trait interrelationships, to compare genotypes by multiple traits and identify candidates that could be used as parents. Relationships of traits revealed by biplot were compared to Pearson’s correlation coefficients.

3 Results and Discussion

Analysis of variance of the mixed model showed high significant differences for all the sources of variation (Table 1). Predominant influence of environment was determined for all examined traits; the smaller part is attributed to genotype’s effects, while the genotype × environment interaction contributed least to total variation.

First two axes (PC1 and PC2) on PC biplot accounted for 59.0% of the variation of the standardized data (Fig. 1). Along the PC1, genotypes were mostly allocated on the basis of *PN*, *SN*, *SYP* and *PROT*, while PC2 was mainly related to *PH*, *NN*, *TSW* and *OIL*. Majority of genotypes were intermediate for a number of traits, distributing close to the biplot origin. Biplot revealed group of genotypes (Turska 2, Apache, Atlas, Lambert) characterized by high oil and above average protein content (21% and 39%, respectively; data not shown). Highlighted varieties may be of importance for the

Table 1. Mean squares from analysis of variance for eight agronomic traits in soybean

Source of variation	Trait							
	PH	NN	PN	SN	TSW	SYP	PROT	OIL
G	48.8**	34.1**	20.8**	17.0**	65.4**	22.6*	51.9**	46.9**
E	831.7**	349.4**	420.4**	318.3**	752.4**	629.1**	239.4**	116.9**
G × E	8.7**	9.0**	5.5**	3.9**	14.9**	4.8**	9.2**	6.6**

PH – plant height, NN – node number, PN – pod number, SN – seed number, TSW – 1000 seed weight, SYP – seed yield per plant, PROT – protein content, OIL – oil content, G – genotype, E – environment, G × E – genotype × environment interaction

*P < 0.05; ** P < 0.01

simultaneous improvement of both traits, which are in soybean negatively correlated. Variety K-1 had the highest PROT and maximum TSW, but due to a lower SYP compared to the mean of the group, this variety has a potential for breeding for the grain size and protein content but not for the seed yield.

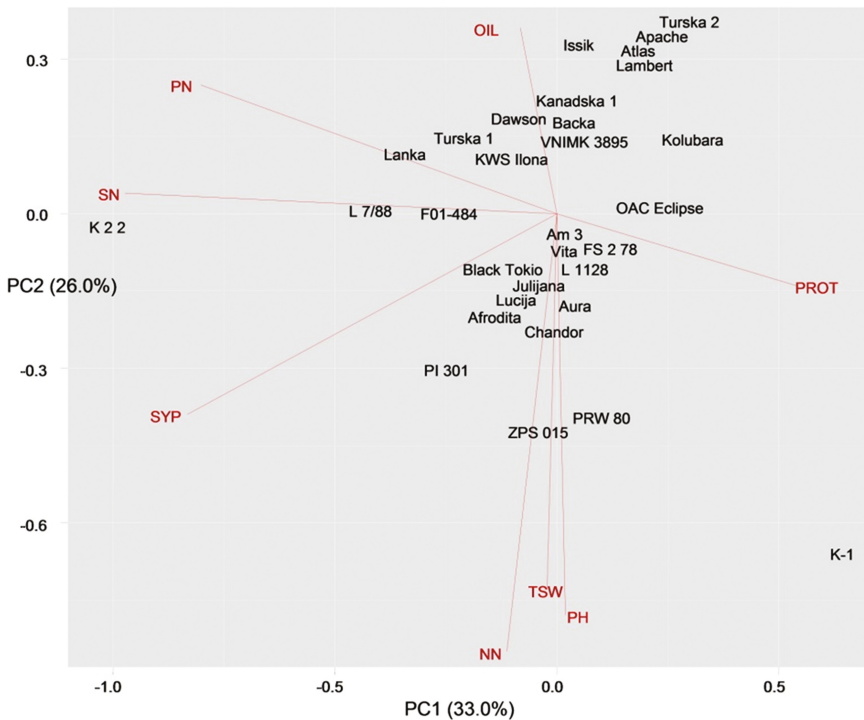


Fig. 1. PCA biplot of 31 soybean genotypes for 8 agronomic traits (PH – plant height, NN – node number, PN – pod number, SN – seed number, TSW – 1000 seed weight, SYP – seed yield per plant, PROT – protein content, OIL – oil content)

Genotype K 2 2 had the largest *PN* and *SN*, while for *TSW* was intermediate, achieving the highest *SYP* in the group. Cultivar Lanka produced larger *PN*, *SN* and *TSW* as compared to the group average, and achieved the second ranked *SYP*. Those genotypes are a useful source for breeding for seed yield through a parallel selection of multiple components, in soybean usually negatively correlated.

The angles of trait vectors on the biplot suggest that the seed yield per plant showed positive correlation with *SN*, *PN* and *TSW*, negative correlation with *PROT*, and a weak negative correlation with *OIL*.

Table 2. Pearson’s correlation coefficients between the traits

PH	0.72**	-0.23	-0.06	0.20	0.04	-0.06	-0.05
NN		-0.06	0.13	0.29	0.25	0.16	-0.11
PN			0.81**	-0.32	0.51**	-0.21	0.17
SN				-0.15	0.78**	-0.43*	0.07
TSW					0.49**	0.08	-0.29
SYP						-0.37*	-0.09
PROT							-0.57**
	NN	PN	SN	TSW	SYP	PROT	OIL

PH – plant height, *NN* – node number, *PN* – pod number, *SN* – seed number, *TSW* – 1000 seed weight, *SYP* – seed yield per plant, *PROT* – protein content, *OIL* – oil content

*P < 0.05; ** P < 0.01

Pearson’s correlation coefficients (Table 2) were in agreement with the correlations displayed by biplot. Similar to the findings by Sudaric and Vratarić (2002), seed yield had a positive and highly significant correlation with *SN*, and to a lesser extent a positive and significant correlation with *PN* and *TSW*, suggesting that *SN* present the one of the most effective selection criteria. A moderate, negative and highly significant correlation observed for oil and protein content was in agreement with several studies (Popovic et al. 2013; Chung et al. 2003). Although not expected, the negative, low and non-significant relationship between *SYP* and *OIL* found in this research was also reported by Eskandari et al. (2013) and Yan and Rajcan (2002), and could be explained by the heterogeneous genetic background of studied genotypes and specific interaction of each genotype with the environment, which possibly influence the consistency of the correlations (Bellaloui et al. 2009).

4 Conclusion

Comparison of early-maturing soybean varieties based on multiple traits highlighted those that could be used as potential parents in breeding for agronomically important traits. Among the observed yield components, seed and pod number per plant were determined as the most reliable selection criteria.

Acknowledgments. This study is a part of a scientific project TR 31068 supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

References

- Bellaloui N, Smith JR, Ray JD, Gillen AM (2009) Effect of maturity on seed composition in the early soybean production system as measured on near-isogenic soybean lines. *Crop Sci* 49:608–620
- Chung J, Babka HL, Graef GL, Staswick PE, Lee DJ, Cregan PB, Shoemaker RC, Specht JE (2003) The seed protein, oil, and yield QTL on soybean linkage group I. *Crop Sci* 43:1053–1067
- Eskandari M, Cober ER, Rajcan I (2013) Genetic control of soybean seed oil: II. QTL and genes that increase oil concentration without decreasing protein or with increased seed yield. *Theor Appl Genet* 126(6):1677–1687
- FAOSTAT (2014) Food and agriculture organization. <http://faostat.fao.org/>
- Li H, Burton JW (2002) Selecting increased seed density to increase indirectly soybean seed protein concentration. *Crop Sci* 42:393–398
- Mohamadi R, Amri A (2011) Graphyc analysis of trait relations and genotype evaluation in durum wheat. *J Crop Improv* 25:680–696
- Popović V, Miladinović J, Vidić M, Tatić M, Sikora V, Ikanović J, Dozet G (2013) Productive and quality characteristics of soybean in agroecological conditions of Sombor, Serbia. *Field Vegetable Crops Res* 50(2):67–74
- Sudarić A, Vratarić M (2002) Variability and interrelationship of grain quantity and quality characteristics in soybean. *Die Bodenkultur* 53(3):137–142
- Taski-Ajdukovic K, Djordjevic V, Vidic M, Vujakovic M (2010) Subunit composition of seed storage proteins in high protein soybean genotypes. *Pesquisa Agropecuária Brasileira* 45(7):721–729
- Yan W, Rajcan I (2002) Biplot analysis of test sites and trait relations of soybean in Ontario. *Crop Sci* 42:11–20
- Yan W, Fregeau-Reid J (2008) Breeding line selection based on multiple traits. *Crop Sci* 48: 417–423