

## RESEARCH

# Selection of Early Soybean Inbred Lines Using Multiple Indices

Gabriel de Siqueira Gesteira,\* Adriano Teodoro Bruzi, Roberto Kazuhiko Zito, Vanoli Fronza, and Neylson Eustáquio Arantes

## ABSTRACT

Soybean [*Glycine max* (L.) Merr.] has been studied and enhanced for most of its economically important traits. Previous research has studied the association among them and the effect of the genotype  $\times$  environment interaction; however, less is known about their correlation considering absolute maturity, as well as the use of multiple selection indices to study the genotype  $\times$  environment interaction and select superior cultivars. Regarding this, the aim of the present study was to identify lines that associate precocity, good yield performance, and high oil and protein contents in the grains, as well as to estimate the correlation among these traits and study the effect of genotype  $\times$  environment interaction using a standardized multiple selection index. Trials were conducted in two crop seasons in the state of Minas Gerais, Brazil, with 39 lines in 13 evaluation environments. The experiments were conducted in a randomized complete block design with four replications, and the grain yield, absolute maturity, and oil and protein contents in the grains were evaluated. The results indicated high experimental precision and accuracy, with significant differences among lines for all traits. High-magnitude correlations between evaluated traits were found, highlighting the negative correlation between absolute maturity and protein content in the grains. The genotype  $\times$  environment interaction was also significant, and the use of the multiple selection index was efficient to identify superior and stable inbred lines by the genotype + genotype  $\times$  environment (GGE) biplot method, which explained 82.23% of the genotype  $\times$  environment interaction effect. Lines 27 and 31 stood out from the others because they associated stability and good performance for all evaluated traits.

G. de Siqueira Gesteira, Genetics and Plant Breeding Dep., Luiz de Queiroz College of Agriculture (ESALQ/USP), Piracicaba, São Paulo, Brazil; A.T. Bruzi, Plant Sciences Dep., Federal Univ. of Lavras (UFLA), Lavras, Minas Gerais, Brazil; R.K. Zito, EMBRAPA Soybean, Londrina, Paraná, Brazil; V. Fronza, EMBRAPA Wheat, Passo Fundo, Rio Grande do Sul, Brazil; N.E. Arantes, Triangle Foundation, Uberaba, Minas Gerais, Brazil. Received 3 May 2018. Accepted 7 June 2018. \*Corresponding author (gabrielgesteira@usp.br). Assigned to Associate Editor Stella Kantartzi.

**Abbreviations:** CV, coefficient of variation; GGE, genotype + genotype  $\times$  environment; PC, principal component; SMSI, standardized multiple selection index; VCU, value for cultivation and use.

**S**OYBEAN [*Glycine max* (L.) Merr.] is one of the most important crops in the international market, being the fourth most produced and consumed crop worldwide. It has average oil and protein contents of 20 and 40%, respectively, which makes it an excellent raw material for feed preparation, oil extraction, and biodiesel production, as well as containing cosmetic, nutraceutical, and pharmaceutical properties (EMBRAPA, 2013).

Brazil is the second largest soybean producer worldwide and is one of the largest exporters of its products and byproducts (CONAB, 2016). The country stands out in the international market because it shows great potential for agricultural production and for the possibility of harvesting two crops successively in the rainy season (summer) in some regions, which allows for greater production in the same area, and hence greater economic returns.

After the beginning of the second harvest in Brazil, a good part of the soybean breeding programs turned to the development of earlier cultivars, aiming to anticipate the harvest and to enable the planting of the crop in succession. However, this reduction in the crop cycle may influence the expression of other traits, especially those of greater economic interest, such as grain yield, size, and architecture of plants, resistance to biotic and abiotic factors, and

Published in Crop Sci. 58:2494–2502 (2018).  
doi: 10.2135/cropsci2018.05.0295

© Crop Science Society of America | 5585 Guilford Rd., Madison, WI 53711 USA  
All rights reserved.

the oil and protein contents in the grains. It is known that there is a positive and large magnitude correlation between grain yield and absolute maturity in soybean, which means that the reduction in the crop cycle would generally imply a reduction in the average yield of cultivars.

The estimation of phenotypic, genetic, and environmental parameters allows identifying and studying the traits of interest for breeding programs. It is also possible to estimate the correlation among them, aiming to verify if the observed variation in one trait can influence the variation observed in another trait (Cruz et al., 2012). Several studies have been conducted with the aim of estimating these parameters for several traits in soybean, as well as their correlations (Carvalho et al., 2002b; Lopes et al., 2002; Carpentieri-Pípolo et al., 2005; Peluzio et al., 2005; Almeida et al., 2010; Nogueira et al., 2012; Rigon et al., 2012; Gesteira et al., 2015).

Although most of these studies deal with quantitative traits, no studies were found in the literature involving the correlation between oil and protein contents with absolute maturity in soybean. The study of the correlation among these traits is extremely important, since it allows identifying if the change in the absolute maturity of commercial cultivars can affect the oil and protein contents present in the grains.

In value for cultivation and use (VCU) experiments, several traits are usually evaluated. An alternative for contemplating more than one characteristic in the selection is the adoption of an index. This strategy has been applied in the soybean on some occasions (Soares et al., 2015; Silva, 2016). However, it should be noted that there are no reports on the use of this tool in the selection of soybean cultivars for absolute maturity, yield, and grain quality.

The genotype  $\times$  environment interaction is often reported for different traits in soybean (Carvalho et al., 2002a; Lima et al., 2008; Vasconcelos et al., 2010; Marques et al., 2011; Barros et al., 2012; Bueno et al., 2013; Silva et al., 2015; Soares et al., 2015) and other crop species, like barley (*Hordeum vulgare* L.), triticale ( $\times$ *Triticosecale* Wittm. ex A. Camus), and wheat (*Triticum aestivum* L.) (Kendal, 2015; Kendal and Sener, 2015; Dogan et al., 2016; Kendal and Sayar, 2016; Kendal et al., 2016; Kendal and Tekdal, 2016). When the existence of interaction is verified, a strategy that can be adopted is the adaptability and stability analysis through the genotype + genotype  $\times$  environment (GGE) biplot method. Several studies report the use of this method, especially for the trait grain yield. However, results on the study of adaptability and stability are not available when considering multiple traits in soybean.

Based on the above, the present study was performed with the aim of obtaining information on the association among the traits of grain yield, absolute maturity, and oil and protein contents in grains; studying the genotype  $\times$

environment interaction (crop seasons and sites) in the choice of cultivars using multiple index selection; and identifying soybean lines that associate early maturity, good yield, and grain quality.

## MATERIAL AND METHODS

### Conducting the Experiments

The experiments were conducted in 10 cities in the state of Minas Gerais (Brazil), during two crop seasons, with the aim of covering the soybean production area of the state (Table 1). In the first crop season, 17 elite inbred lines were evaluated, and in the second crop season, 20 elite inbred lines from the same program were evaluated, three of which were common to the first crop season. Five commercial cultivars (Treatments 1–5) were used as controls in all experiments, totaling 39 different inbred lines (Table 2). All elite inbred lines evaluated were developed at the EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária) Soybean Breeding Program (in partnership with EPAMIG [Empresa de Pesquisa Agropecuária de Minas Gerais], Triângulo Foundation, EMATER–GO [Agência Goiana de Assistência Técnica, Extensão Rural e Pesquisa Agropecuária] and CTPA [Centro Tecnológico para Pesquisas Agropecuárias]).

The no-tillage system was adopted and the seeds were inoculated with *Bradyrhizobium japonicum* using a liquid inoculant, in the proportion of 1,200,000 bacteria per seed. Seeding was performed mechanically in both crop seasons. The thinning was done 25 d after germination, aiming to guarantee the final stand of 15 plants per linear meter. The other cultural practices were performed according to the procedure shown by Soares et al. (2015).

The 39 treatments were evaluated in randomized complete blocks design with four replicates, each plot consisting of four rows 5 m in length, with 0.5-m spacing between rows. The seeds of the two central rows were harvested, except for the initial and terminal 0.5 m of each row, totaling 4 m<sup>2</sup> of useful area.

The following traits were evaluated:

1. Grain yield (kg ha<sup>-1</sup>), obtained by individual harvesting of each plot, weighing and correction for 13% moisture, and extrapolation of obtained value for the number of kilograms harvested in one hectare.
2. Absolute maturity, comprising the number of days from sowing until maturity, represented by 95% of plants with mature pods in each plot.
3. The percentage protein and oil contents in the grains, determined in whole grains by the technique of the near-infrared reflectance spectroscopy (NIRS), according to Heil (2010). The grains of every line were submitted to duplicate readings through Thermo equipment (Antaris II model) with an integrating sphere with a resolution of 4 cm<sup>-1</sup>.

The obtained data were used to perform individual analyzes of variance per site in each crop season following the model

$$y_{ij} = \mu + g_i + b_j + \varepsilon_{ij}$$

where  $y_{ij}$  is the phenotypic observation of line  $i$  in block  $j$ ,  $\mu$  is the overall average,  $g_i$  is the effect of line  $i$ ,  $b_j$  is the effect of

**Table 1. Crop seasons, trial sites, latitude and longitude (geographical), annual rainfall, and correspondent environments.**

Crop season	City	Latitude	Longitude	Annual rainfall	Environment
		°S	°W	mm	
2014–2015	Conceição das Alagoas	19.9147	48.3883	1485.6	1
2014–2015	Iraí de Minas	22.0989	46.1897	1309.0	2
2014–2015	Presidente Olegário	18.4178	46.4181	1504.0	3
2014–2015	Sacramento	19.8653	47.4400	1630.2	4
2014–2015	Uberaba	19.7667	47.9333	1528.5	5
2014–2015	Muzambinho	21.3758	46.5256	1488.2	6
2015–2016	Conceição das Alagoas	19.9147	48.3883	1485.6	7
2015–2016	Presidente Olegário	18.4178	46.4181	1504.0	8
2015–2016	Sacramento	19.8653	47.4400	1630.2	9
2015–2016	Uberaba	19.7667	47.9333	1528.5	10
2015–2016	Água Comprida	20.0564	48.1089	1409.0	11
2015–2016	Buritzeiro	17.3511	44.9622	1086.9	12
2015–2016	Pedrinópolis	19.2278	47.4622	1605.6	13

block *j*, and  $\varepsilon_{ij}$  is the error associated with the observation of line *i* in block *j*.

Later, since the data were unbalanced and the analysis was based on a mixed model, a joint deviance analysis was performed considering all the environments (site-year combination), following the model

$$y_{ijk} = \mu + g_i + a_k + b_{j(k)} + ga_{ik} + \varepsilon_{ijk}$$

where  $y_{ijk}$  is the phenotypic observation of line *i* in block *j* in the environment *k*,  $\mu$  is the overall average,  $g_i$  is the effect of line *i* (fixed nature),  $a_k$  is the effect of environment *k* (random nature),  $b_{j(k)}$  is the effect of block *j* in environment *k* (random nature),  $ga_{ik}$  is the effect of the interaction line *i* and environment *k* (random nature), and  $\varepsilon_{ijk}$  is the error associated with the observation of line *i* in block *j* in the environment *k*.

The joint deviance analysis was performed using the restricted maximum likelihood method, with a significance test of fixed effect by the *F* value, and the random effects by the likelihood-ratio test. The averages obtained in all analyses

**Table 2. Treatments and their corresponding cultivars (1–5) and inbred lines (6–39).**

Treatment	Line	Treatment	Line
1	NA 5909 RR	21	BRRY34-0901
2	BMX Potência RR	22	RRMG11-55821
3	BMX Desafio RR	23	BRRR12-67015
4	Anta 82 RR	24	BRRR12-67205
5	NA 7337 RR	25	BRRY34-1172
6	BR09-1554	26	BRRY34-1175
7	RRMG09-99717	27	BRRY45-10190
8	RRMG11-53504	28	BRRY45-10378
9	RRMG11-54115	29	BRRY45-10473
10	RRMG11-56011	30	BRRY45-10649
11	RRMG11-57119	31	BRRY45-10729
12	BRRR12-67204	32	BRY34-1245
13	BRRR12-70304	33	RRMG12-69204
14	BRY23-0068	34	RRMG12-78605
15	BRY23-0080	35	RRMG13-85603
16	BRRY34-0591	36	RRMG13-85908
17	BRY23-0035	37	RRMG13-87803
18	BRY23-0089	38	RRMG13-88501
19	BRY23-0062	39	RRMG13-97816
20	BRRY34-0784		

were compared by Scott and Knott (1974) at 5% probability. Experimental accuracy was measured by estimating the coefficient of variation (CV) and accuracy (Resende and Duarte, 2007), according to the estimators below:

$$CV = (\sigma/\bar{y})100$$

where CV is the coefficient of variation expressed as a percentage,  $\sigma$  is the standard deviation of the trait, and  $\bar{y}$  is the average of the trait.

$$r = \left[ \sqrt{1 - (1/F_c)} \right] 100$$

where *r* is the accuracy expressed as a percentage and  $F_c$  is the calculated *F* value.

The CV is dependent on the phenotypic average, and thus the lower the average value, the higher the CV. On the other hand, the accuracy is independent of the average magnitude, which is the main advantage of using this parameter when evaluating the experimental accuracy (Resende and Duarte, 2007).

## Estimation of Phenotypic Parameters and Index Selection

With the adjusted phenotypic averages of the lines obtained in the joint analysis, the association among yield, absolute maturity, and oil and protein contents in the grains was estimated, adopting the following estimator:

$$r_p = \text{cov}_{P_{XY}} / \left( \sqrt{\sigma_{P_X}^2 \sigma_{P_Y}^2} \right)$$

where  $r_p$  is the estimator of the phenotypic correlation between the *X* and *Y* traits,  $\text{cov}_{P_{XY}}$  is the phenotypic covariance between the *X* and *Y* traits,  $\sigma_{P_X}^2$  is the phenotypic variance for trait *X*, and  $\sigma_{P_Y}^2$  is the phenotypic variance for trait *Y*.

The significance of the obtained correlations was estimated by the *t* test with *n* - 2 df, where *n* corresponds to the number of evaluated genotypes.

The standardized multiple selection index (SMSI) was calculated based primarily on the standardization of the phenotypic averages obtained in the individual analyses for all lines, using the following estimator:

$$Z_{ik} = (\bar{y}_{ik} - \bar{y}_k) / \sigma_k$$

where  $Z_{ik}$  is the standardized average,  $\bar{y}_{ik}$  is the average of line  $i$  in the environment  $k$ ,  $\bar{y}_k$  is the overall average in the environment  $k$ , and  $\sigma_k$  is the standard deviation in the environment  $k$ .

The index was calculated considering the grain yield, absolute maturity, oil content, and protein content, which were added originating the SMSI, according to the following expression (Mendes et al., 2009):

$$Z_{MIS} = Z_{GY} - Z_{AM} + Z_{Oil} + Z_{Protein}$$

where  $Z_{MIS}$  is the multiple index selection,  $Z_{GY}$  is the standardized average for the trait grain yield,  $Z_{AM}$  is the standardized average for the trait absolute maturity,  $Z_{Oil}$  is standardized average for the trait oil content, and  $Z_{Protein}$  is the standardized average for the trait protein content. The obtained values were added to a constant to eliminate negative values.

### Study of Genotype $\times$ Environment Interaction

The SMSI was used to study the effect of genotype  $\times$  environment interaction, with the objective of including the four main traits associated with line yield and grain quality (yield, absolute maturity, and oil and protein contents in the grains), thus forming a multivariate analysis.

Due to the highly imbalanced degree of the data, the phenotypic values of each absent genotype  $\times$  environment interaction were predicted, aiming to fill in the double-entry tables of the SMSI considering all the environments for each evaluated characteristic, as well as to enable the study of genotype  $\times$  environment interaction. The method used to predict missing values was based on the singular value decomposition in combination with the expectation-maximization algorithm for the iterations to find the value with the maximum likelihood of occurrence, as proposed by Yan (2013).

Cross-validation was also performed to calculate the predictive accuracy through resampling and predicting values from a submatrix, contemplating the balanced data of the second harvest, after which the correlation of the predicted values with the observed phenotypic values was calculated. The value of the predictive accuracy was calculated based on the average of correlations obtained in 1000 resamplings for all traits.

The double-entry tables of the four characteristics of interest, containing the 39 genotypes evaluated in the 12 environments, were filled with the predicted values and used to calculate the SMSI. With the balanced data, the decomposition of genotypic effects and genotype  $\times$  environment interaction was performed following the model proposed by Yan et al. (2000), according to the equation:

$$Y_{ij} - \mu - \beta_j = \lambda_1 \alpha_{i1} \gamma_{1j} + \lambda_2 \alpha_{i2} \gamma_{2j} + \varepsilon_{ij}$$

where  $Y_{ij}$  is the phenotypic average of line  $i$  in the environment  $j$ ,  $\mu$  is the overall average of the experiment,  $\beta_j$  is the main effect of the environment  $j$ ,  $\lambda_1$  is the largest eigenvalue of the first principal component (PC 1),  $\alpha_{i1}$  is the eigenvector of line  $i$  for PC 1,  $\gamma_{1j}$  is the eigenvector of the environment  $j$  for PC 1,  $\lambda_2$  is largest eigenvalue of the second principal component (PC 2),  $\alpha_{i2}$  is the eigenvector of line  $i$  for PC 2,  $\gamma_{2j}$  is the eigenvector of the environment  $j$  for PC 2, and  $\varepsilon_{ij}$  is the error that was not explained by both effects.

The biplot graph was then obtained for the effects of the first two PCs (PC 1 vs. PC 2), being the first component associated

with the yield ratio, which is due only to the genotype characteristics, and the second component associated with the part of the income due to the genotype  $\times$  environment interaction (Yan and Holland, 2010; Yan, 2011). Biplots with axis correction were also obtained by means (mean vs. stability), and the biplot PC 1 vs. PC 2 subdivided into macro-environments with line ranking (which-won-where). Analyses and graphs were generated by the “GGEbiplotGUI” package using R computer software (R Core Team, 2017).

## RESULTS AND DISCUSSION

The results of the joint deviance analysis, considering all environments, are presented in Table 3. When evaluating lines in the final stages of a breeding program (i.e., VCU), it is essential that they be well conducted and that high experimental accuracy be obtained. Considering the parameters used to measure the experimental accuracy and precision, the highest values for accuracy and lowest values for coefficient of variation are desirable. In the present study, as mentioned above, the accuracy (Resende and Duarte, 2007) and the CV (Pimentel-Gomes, 2009) were adopted as tools to infer on the experimental accuracy and precision. The minimum and maximum values obtained for the accuracy and CV were 0.93 and 10.97, respectively, which indicate medium to high experimental accuracy and precision.

Accuracy was affected by the referred trait, as well as by the used analysis strategy. In the individual analyses, lower accuracy was evident. When the joint analysis was adopted, there was an increase in accuracy. This fact can be justified because a greater number of replicates is achieved when the joint analysis is performed, thus guaranteeing greater experimental accuracy and lower error associated with the estimates (Pimentel-Gomes, 2009).

There were significant differences among lines for all evaluated traits. This fact can be explained because the tested lines have genetic differences, due to their different genetic background. There were also significant effects of environments on the expression of traits. The environmental effect in this study is due to the combination of predictable and unpredictable factors (Allard and Bradshaw, 1964). Both factors are present because the different lines were tested in different sites in the state of Minas Gerais, Brazil, as well as in two crop seasons. The unpredictable factors are particularly important for VCU experiments, since the lines must be tested for at least 2 yr to quantify this effect and minimize errors in the recommendation of cultivars (Kaster and Farias, 2012).

The genotype  $\times$  environment interaction was also significant, which suggests that the lines did not show coincident behavior for all the traits in the different evaluation environments. Genotype  $\times$  environment interaction for soybean in Minas Gerais has been reported in the literature (Gesteira et al., 2015; Silva et al., 2015; Soares et al., 2015). There are some alternatives that can be adopted in the occurrence of the genotype  $\times$  environment interaction to

**Table 3. Summary of joint deviance analysis, considering all environments, for all evaluated traits.**

Random effect	df	Grain yield		Absolute maturity		Oil		Protein	
		Variance	P value	Variance	P value	Variance	P value	Variance	P value
Line × Environment	456	22.57	<2 × 10 <sup>-16**</sup>	16.36	<2 × 10 <sup>-16**</sup>	0.20	3 × 10 <sup>-4**</sup>	0.59	4 × 10 <sup>-16**</sup>
Replication(environment)	39	6.12	9 × 10 <sup>-12 **</sup>	0.61	2 × 10 <sup>-5**</sup>	–	–	–	–
Environment	12	97.24	<2 × 10 <sup>-16**</sup>	57.40	<2 × 10 <sup>-16**</sup>	0.15	5 × 10 <sup>-10**</sup>	0.27	1 × 10 <sup>-8**</sup>
Residual	1482 (507)†	56.48	–	10.35	–	0.70	–	0.67	–
Fixed effect	df	F	Pr(> F)	F	Pr(> F)	F	Pr(> F)	F	Pr(> F)
Line	39	7.92	0.00	34.18	0.00	8.30	0.00	10.34	0.00
CV‡ (%)	–	10.97	–	8.27	–	3.80	–	3.42	–
Accuracy	–	0.93	–	0.99	–	0.94	–	0.95	–
Avg.	–	3582.94	–	114.90	–	21.98	–	37.34	–

\*\* Significant at the 0.01 probability level.

† The residual for oil and protein contents is given in parentheses.

‡ CV, coefficient of variation.

identify the best lines. In this study, we chose to identify the most stable lines (Kang and Gauch, 1996).

The adjusted averages of the 39 lines evaluated in the 13 environments for all the characteristics are presented in Table 4. The grain yield averages ranged from 2584.90 to 4515.87 kg ha<sup>-1</sup> for Lines 9 and 18, respectively, with an overall average of 3582.94 kg ha<sup>-1</sup>. Absolute maturity ranged from 94.64 to 134.39 d also for Lines 9 and 18. For the oil content trait, the averages ranged from 19.26 to 23.67% for Lines 8 and 6, whereas the protein content ranged from 34.64 to 41.12% for Lines 18 and 8, respectively.

When studying several traits in breeding programs, it is appropriate to quantify the existence of an association among them. In this study, the association among the traits grain yield, absolute maturity, and oil and protein contents present in the grains were obtained (Table 5). The correlation between grain yield and absolute maturity was positive and of high magnitude. Thus, it is expected that to increase the productive potential of cultivars, later lines should be selected (Sediyama, 2015). However, it should also be mentioned that it is possible to obtain early and highly productive cultivars. This fact is possible due to the new strategy adopted by soybean breeding programs in Brazil, focusing on the selection of semidetermined and undetermined growth habit cultivars. In these groups, it is possible to reduce the crop cycle, as well as to associate good productive potential (Sediyama, 2015).

Several reports show the success on the use of early soybean cultivars. Gesteira et al. (2015) commented that one of the main advantages on the adoption of these cultivars in Minas Gerais is the possibility of cultivating the second crop (mid-crop season). The reduced-cycle cultivars allow the planting of crop in succession to be performed in advance, thus ensuring greater utilization of the rainy season, the agricultural areas, and higher yields for the producers.

The association between grain yield and protein content was negative and highly significant, similar to the association between absolute maturity and protein content.

**Table 4. Adjusted averages for evaluated traits and for standardized multiple selection index (SMSI) in all inbred lines, obtained from the joint analysis.**

Line	Grain yield	Absolute maturity	Oil	Protein	SMSI
	kg ha <sup>-1</sup>	d	%		
1	3793.58g†	109.42s	23.33c	35.77p	8.35
2	3670.90i	115.00m	22.39h	36.99l	7.30
3	4202.25c	124.92e	22.54g	36.70m	7.20
4	3782.70g	123.50g	21.90l	37.14k	6.27
5	4000.09d	129.64b	21.73n	36.25n	4.77
6	3775.20g	109.81r	23.67a	37.12k	9.04
7	3582.05k	109.19t	22.69f	37.60i	8.48
8	2807.65t	98.14x	19.26u	41.12a	6.69
9	2584.90u	94.64z	21.96l	38.13f	7.07
10	3444.50n	103.75w	21.86m	37.51j	8.02
11	3171.87q	105.89v	22.10k	37.74h	7.76
12	3544.23l	111.25p	22.94d	37.68h	8.33
13	3684.32i	112.19o	22.12k	37.77h	7.40
14	4439.92b	129.85b	21.17q	37.02l	5.48
15	3963.82d	126.85d	21.78m	35.13q	4.91
16	3480.83m	120.39h	22.34i	36.11o	5.84
17	3869.05f	128.85c	22.43h	36.06o	6.40
18	4515.87a	134.39a	22.02k	34.64r	5.56
19	3774.25g	126.94d	22.15k	36.20n	6.18
20	3907.04e	124.19f	21.70n	37.64i	6.98
21	3626.20j	120.10i	22.47g	35.19q	5.95
22	2860.50s	95.60y	21.20q	38.10f	7.73
23	3753.86g	111.41p	21.42o	38.06f	7.75
24	3581.81k	109.98r	22.17k	37.77h	8.06
25	3184.95q	115.98k	21.30p	38.87c	6.27
26	3543.88l	116.14k	20.88s	38.75d	6.52
27	3534.24l	109.73r	23.56b	37.62i	9.30
28	3313.81o	112.68n	22.34i	38.16f	7.45
29	3581.94k	117.19j	22.25j	37.64i	6.89
30	3035.19r	116.01k	22.87d	36.30n	5.26
31	3647.19j	107.88u	22.22j	38.41e	9.14
32	3206.58q	128.70c	22.10k	35.78p	2.58
33	3707.02h	110.73q	21.84m	37.94g	7.96
34	3619.55j	112.84n	20.98r	38.35e	7.03
35	3641.83j	109.98r	20.63t	39.45b	7.91
36	3506.70m	115.73l	21.32p	38.38e	6.74
37	3715.23h	109.13t	22.79e	35.79p	7.44
38	3410.08n	109.91r	21.65n	37.85h	7.19
39	3269.19p	112.38o	21.07r	37.44j	5.52
Avg.	3582.94	114.90	21.98	37.34	6.94

† Averages followed by the same letter belong to the same group, according to Scott and Knott (1974), at 5% probability.

**Table 5. Pearson estimated correlations between all evaluated traits, considering adjusted averages shown in Table 4.**

Trait	Absolute maturity	Oil	Protein
Grain yield	0.7204**	0.2181	-0.4998**
Absolute maturity		0.1151	-0.6306**
Oil			-0.5884**

\*\* Significant at the 0.01 probability level.

It was evident that the lower the absolute maturity, the higher the protein content in the grains, according to the obtained negative correlation and high magnitude. A possible explanation is related to the positive association between oil content and grain yield, as well as the negative association between oil and protein contents in the grains (Rincker et al., 2014; Gesteira et al., 2015), which may be related to a lower degradation rate of proteins present in the grains with the decrease of the crop cycle. The higher the productive potential, the longer the crop cycle and higher the oil content in the grains hence lower the protein content (Rincker et al., 2014), which explains the negative association obtained.

The association between the oil and protein contents in soybean has been reported in the literature (Carvalho et al., 2002b; Lopes et al., 2002; Carpentieri-Pípolo et al., 2005; Peluzio et al., 2005; Almeida et al., 2010; Nogueira et al., 2012; Rigon et al., 2012) and corroborates the results obtained in the present study. It can be seen that the strategy adopted by soybean breeding programs in Brazil has been focused on obtaining more productive cultivars. Thus, an increase in the percentage of oil content in the grains and hence a reduction in the percentage of protein content is expected. However, it should be emphasized that although there is a reduction in the percentage of protein content in the grains, the absolute yield of protein per area may be higher due to the increase in yield.

The index selection is a very useful tool to identify and select lines that associate good attributes for several traits. The results for the SMSI are presented in Table 4. The main objective of the index selection was to identify more productive lines with lower crop cycles and high oil and protein contents in the grains. It can be observed that Line 27 was more prominent by SMSI (i.e., it was associated with early cycle, good yield, and high grain quality [high oil and protein contents]).

The use of index selection in the soybean crop has been reported in the literature. In a study performed by Soares et al. (2015), the authors adopted a multiple-index selection and verified that it was efficient in the selection of new soybean cultivars, corroborating the results obtained in this study.

When phenotypic data are available in different environments, an alternative for the identification of more stable cultivars is the use of methods to study the genotype  $\times$  environment interaction. Several studies considering the

genotype  $\times$  environment interaction in many crop species are presented in the literature (Kendal, 2015; Kendal and Sener, 2015; Dogan et al., 2016; Kendal and Sayar, 2016; Kendal et al., 2016; Kendal and Tekdal, 2016). In the present study, as already mentioned, the GGE biplot (Yan et al., 2000) was adopted, considering the SMSI. The study of genotype  $\times$  environment interaction is important because it allows obtaining information about the adaptability and stability of genotypes, especially when it is detected significance in the interaction effect. Analysis by GGE biplot is an efficient tool for this study in multisite competition trials (Bhan et al., 2005) and was also considered useful in the study of genotype  $\times$  environment interaction, being a good tool to generate high-quality images and define recommendations for breeders (Kendal and Sayar, 2016). According to Oliveira et al. (2003), the analysis should be performed with the lowest number of axes (PCs) and the highest percentage of explanation possible, because as the number of axes increases, the percentage of noise also increases, which reduces the prediction power of the analysis.

In this context, it is worth pointing out that, in this condition, it is possible to identify stable lines considering not only the trait grain yield, but also important traits such as absolute maturity and grain quality. Previous studies of genotype  $\times$  environment interaction and relationship between traits were performed in durum wheat (*Triticum durum* Desf.) (Kendal, 2015; Kendal and Sener, 2015). The authors found strong relationships between grain yield, grain components, and quality traits, suggesting that some of them can be used for indirect selection due to the presented high association. There is only one report in the literature on an interaction study considering multiple-index selection in soybean; however, it only involved the traits of yield and physiological quality of seeds (Silva, 2016). There were no reports in the literature involving absolute maturity, yield, and grain quality.

After the SMSI analysis, the missing values in the double-entry tables were predicted for the four characteristics of interest, aiming to recalculate the SMSI in a balancing scenario and to allow the study of genotype  $\times$  environment interaction. The validation of predictions was also performed by calculating the predictive accuracy based on 1000 resamplings, whose values were 0.71, 0.88, 0.59, and 0.73 for the yield, absolute maturity, and oil and protein contents in the grains, respectively. The obtained values were of high magnitude, proving the efficacy and reliability of predictions and data used to study the genotype  $\times$  environment interaction.

The analysis of genotype  $\times$  environment interaction using the GGE biplot method was efficient, and the first two PCs explained a large part (82.23%) of the observed genotype + genotype  $\times$  environment effect. This result corroborates those obtained by Amira et al. (2013), who verified in trials with early soybean lines that the first two

PCs explained 86.6% of the interaction effect. Similar results were found in triticale study, where the PC 1 and PC 2 explained 77.50% of the genotype  $\times$  environment interaction effect (Kendal and Sayar, 2016), whereas different results were found in spring barley studies, where the PC 1 and PC 2 explained 52.14 and 68.30% of the interaction effect, respectively (Dogan et al., 2016; Kendal and Tekdal, 2016).

This method considers the environment effect as irrelevant in cultivar selection, since the adaptive capacity of genotypes is more important than the environmental conditions (Camargo-Buitrago et al., 2011). In the GGE biplot graph, the first PC of the graph (abscissa) is associated with adaptability, also presenting high correlation with average performance, whereas the second PC (ordinates) is associated with stability (Yan et al., 2000). It is worth mentioning that the use of SMSI in the GGE biplot analysis will not indicate the best performing lines for all the considered traits, but the lines that show averages above the overall averages of these traits in the experiment, associated with phenotypic stability.

It was observed that Lines 6, 7, 27, and 31 showed good general adaptability because they showed the highest SMSI values, as well as superiority in relation to the other lines in the abscissa axis (Fig. 1). Among them, Line 7 showed lower stability because it is more distant from the origin of the ordinate axis. Thus, the Lines 6, 27, and 31 stood out due to their association with good adaptability, stability, and averages higher than the other evaluated lines.

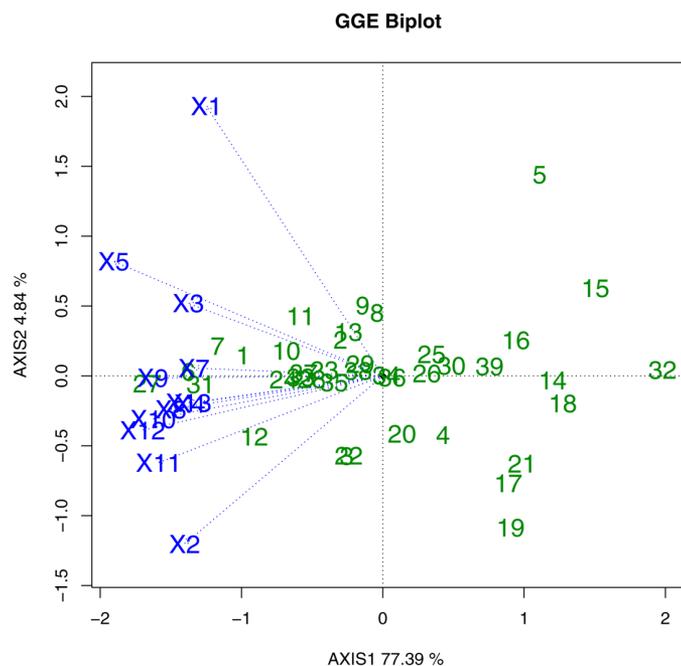


Fig. 1. Biplot with the effects of the first two principal components (PC 1 vs. PC 2) for the standardized multiple selection index (SMSI) of the 39 lines evaluated in 12 environments. The genotypes are represented by the green coloration, whereas the environments are represented by the blue coloration. GGE, genotype + genotype  $\times$  environment.

Following this same reasoning, Environment 13 (Pedrinópolis, Minas Gerais, 2015–2016 crop year) provided the greatest overall adaptability, whereas Environment 5 (Uberaba, Minas Gerais, 2014–2015 crop year) provided the least adaptability.

Figure 2 (mean vs. stability chart) corrects the positioning of axes with the PCs according to the line averages, thus indicating which material showed the best performance considering the SMSI (greater value in the abscissa axis), allowing a simultaneous evaluation of its stability by the ordinate axis. The graph provides a combined analysis of adaptability, stability, and average performance of the lines, which reinforces the importance of the identification of genotypes with superior performance, besides the safety and predictability in the recommendation of cultivars, seeking to guarantee success with the selection. From this perspective, it can be highlighted that although most of the lines (located in the center of graph) showed good stability and adaptability, their averages were inferior to the averages of Lines 27, 6, 31, and 7, mentioned above, considering the four characteristics simultaneously. It is also possible to highlight the superiority of Line 27 in relation to the others, including the point highlighted on the line in reference to the ideal environment. Thus, Line 27 showed the highest average and therefore better performance and grain quality associated with precocity and stability (Fig. 2), reinforcing the results shown in Fig. 1.

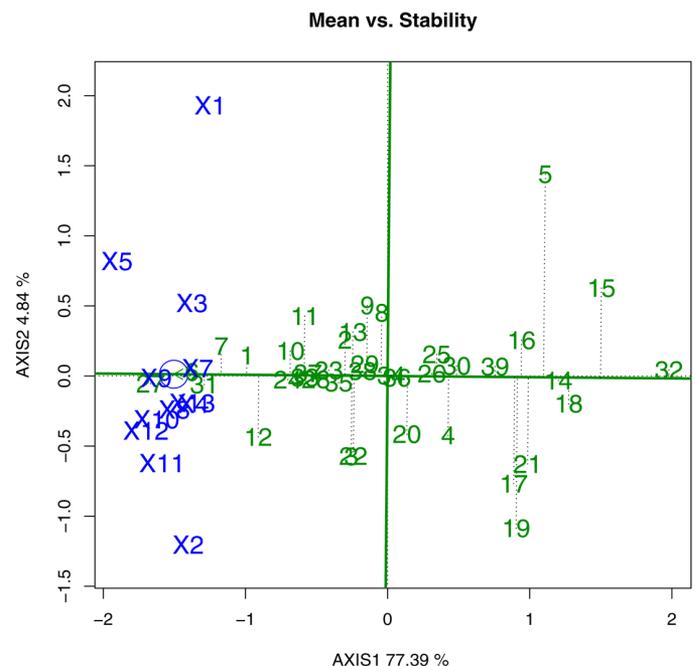


Fig. 2. Biplot of the principal components (PC 1 vs. PC 2) in the main axes, with second positioning of axes in relation to the genotype average for the standardized multiple selection index (SMSI) of the 39 evaluated lines in 12 environments. The genotypes are represented by the green coloration, whereas the environments are represented by the blue coloration.

One of the advantages of a GGE biplot adaptability and stability study is that both axes (PC 1 and PC 2) are in the same unit, which allows the comparison of genotypes in each environment (Yan et al., 2000). A good tool to compare the relative performance of genotypes in environments is the which-won-where chart, which subdivides genotypes and environments into regions called mega-environments, which are delimited by red lines and a polygon. The genotypes and environments contained in each mega-environment have specific adaptability among themselves, and the lines that form the polygon vertices showed the best performance within each region and therefore are the most promising in the environments contained in the referred region.

The obtained graph visualizes the formation of a mega-environment delimited by Line 27 at its extreme vertex (Fig. 3), also encompassing several other lines with a performance superior to the general average of the index. Environments 9, 7, and 4 are also contained in this mega-environment, and it can be suggested that the lines contained in this region showed good specific adaptability for the mentioned environments. However, due to the greater proximity among Lines 27, 6, and 31 with these environments and the fact that both axes present the same unit, it can also be noted that the specific adaptability among these lines and the mentioned environments is greater in relation to the other lines contained in this mega-environment.

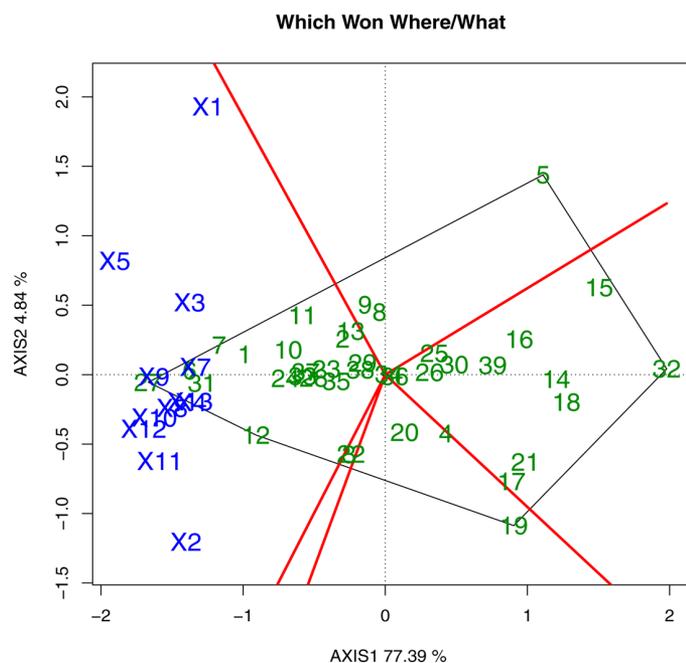


Fig. 3. Biplot of the principal components (PC 1 vs. PC 2) in the main axes, with subdivision of the macro-environments and ranking considering the standardized multiple selection index (SMSI) of the 39 lines evaluated in the 12 environments. The genotypes are represented by the green coloration, whereas the environments are represented by the blue coloration.

Based on this analysis, it can be suggested that Lines 27, 6, and 31 showed superior performance in a larger number of environments and hence contributed less to the interaction than other evaluated lines. These lines associated specific adaptability and good stability with good performance for the traits considered by the SMSI.

## CONCLUSION

Lines 6, 27, and 31 show phenotypic stability associated with good performance in grain yield and quality, considering the SMSI. There is a negative and high-magnitude association between absolute maturity and protein content in the grains, indicating that it is possible to obtain early lines with high protein content in the grains. The GGE biplot method is efficient to quantify the genotype  $\times$  environment interaction when the SMSI is adopted, constituting an important tool for selection and recommendation of cultivars.

## Conflict of Interest

The authors declare that there is no conflict of interest.

## Data Availability

Data are available from the Dryad Digital Repository at doi:10.5061/dryad.51s498d.

## References

- Allard, R.W., and A.D. Bradshaw. 1964. Implications of genotype-environmental interactions in applied plant breeding. *Crop Sci.* 4:503–508. doi:10.2135/cropsci1964.0011183X000400050021x
- Almeida, R.D., J.M. Peluzio, and F.S. Afferi. 2010. Correlações fenotípicas, genotípicas e ambientais em soja cultivada sob condições várzea irrigada, sul do Tocantins. *Biosci. J.* 26:95–99.
- Amira, J.O., D.K. Ojo, O.J. Ariyo, O.A. Oduwaye, and M.A. Ayo-Vaughan. 2013. Relative discriminating powers of GGE and AMMI models in the selection of tropical soybean genotypes. *Afr. Crop Sci. J.* 21:67–73.
- Barros, H.B., T. Sedyama, R.R. Fidelis, and A. Capone. 2012. Adaptabilidade e estabilidade de genótipos de soja por meio de métodos uni e multivariado. *J. Biotechnol. Biodiversity* 3:49–58.
- Bhan, M.K., S. Pal, B.L. Rao, A.K. Dhar, and M.S. Kang. 2005. GGE biplot analysis of oil yield in lemon-grass (*Cymbopogon* spp.). *J. New Seeds* 7:127–139. doi:10.1300/J153v07n02\_07
- Bueno, R.D., L.L. Borges, K.M.A. Arruda, L.L. Behring, E.G. Barros, and M.A. Moreira. 2013. Genetic parameters and genotype  $\times$  environment interaction for productivity, oil and protein content in soybean. *Afr. J. Agric. Res.* 8:4853–4859. doi:10.5897/AJAR2013.6924
- Camargo-Buitrago, I., E.M. Intire, and R. Córdon-Mendoza. 2011. Identificación de mega-ambientes para potenciar el uso de progénies superiores de arroz en Panamá. *Pesquisa Agropecu. Bras.* 46:1061–1069. doi:10.1590/S0100-204X2011000900013
- Carpentieri-Pípolo, V., L.F. Gastaldi, and A.E. Pipolo. 2005. Correlações fenotípicas entre caracteres quantitativos em soja. *Semina Cienc. Agrar.* 26:11–16. doi:10.5433/1679-0359.2005v26n1p11

- Carvalho, C.G.P., C.A.A. Arias, J.F.F. Toledo, L.A. Almeida, R.A.S. Kiihl, and M.F. Oliveira. 2002a. Interação genótipo  $\times$  ambiente no desempenho produtivo da soja no Paraná. *Pesquisa Agropecu. Bras.* 37:989–1000. doi:10.1590/S0100-204X2002000700013
- Carvalho, C.G.P., C.A.A. Arias, J.F.F. Toledo, M.F. Oliveira, and N.A. Vello. 2002b. Correlações e análise de trilha em linhagens de soja semeadas em diferentes épocas. *Pesquisa Agropecu. Bras.* 37:311–320. doi:10.1590/S0100-204X2002000300012
- CONAB. 2016. Acompanhamento da safra Brasília de grãos. Safra 2015/16. No. 3. Comp. Nac. Abastecimento, Brasília.
- Cruz, C.D., A.J. Regazzi, and P.C.S. Carneiro. 2012. Modelos biométricos aplicados ao melhoramento genético. 4th ed. Editora UFV, Viçosa, Brazil.
- Dogan, Y., E. Kendal, and E. Oral. 2016. Identifying of relationship between traits and grain yield in spring barley by GGE Biplot analysis. *Agric. For.* 62:239–252.
- EMBRAPA. 2013. Tecnologias de produção de soja: Região Central do Brasil 2014. Empresa Bras. Pesqui. Agropecu. Soja, Londrina, Brazil.
- Gesteira, G.S., E.V. Zambiazzi, A.T. Bruzi, I.O. Soares, P.M. Rezende, and K.B. Silva. 2015. Seleção fenotípica de cultivares de soja precoce para a região Sul de Minas Gerais. *Rev. Agrogeoambiental* 7:79–88. doi:10.18406/2316-1817v7n32015730
- Heil, C. 2010. Rapid, multi-component analysis of soybeans by FT-NIR spectroscopy. *Thermo Fisher Sci.*, Madison, WI.
- Kang, M.S., and H.G. Gauch, Jr. 1996. Genotype by environment interaction. CRC Press, Boca Raton, FL.
- Kaster, M., and J.R.B. Farias. 2012. Regionalização dos testes de valor de cultivo e uso e da indicação de cultivares de soja: Terceira aproximação. Empresa Bras. Pesqui. Agropecu. Soja, Londrina, Brazil.
- Kendal, E. 2015. Relationship between chlorophyll and other features in durum wheat (*Triticum turgidum* L. var. *durum*) using SPAD and biplot analyses. *J. Agric. Sci. Technol.* 17:1873–1886.
- Kendal, E., and M.S. Sayar. 2016. The stability of some spring triticale genotypes using biplot analysis. *J. Anim. Plant Sci.* 26:754–765.
- Kendal, E., M.S. Sayar, S. Tekdal, H. Aktas, and M. Karaman. 2016. Assessment of the impact of ecological factors on yield and quality parameters in triticale using GGE biplot and AMMI analysis. *Pak. J. Bot.* 48:1903–1913.
- Kendal, E., and O. Sener. 2015. Examination of genotype  $\times$  environment interactions by GGE biplot analysis in spring durum wheat. *Indian J. Genet.* 75:341–348. doi:10.13140/RG.2.1.2611.0809
- Kendal, E., and S. Tekdal. 2016. Application of AMMI model for evaluation spring barley genotypes in multi-environment trials. *Bangladesh J. Bot.* 45:613–620.
- Lima, W.F., A.E. Pípolo, J.U.V. Moreira, C.G.P. Carvalho, C.E.C. Prete, C.A.A. Arias, et al. 2008. Interação genótipo-ambiente da soja convencional e transgênica resistente a glifosato, no Estado do Paraná, Brasil. *Pesquisa Agropecu. Bras.* 43:729–736. doi:10.1590/S0100-204X2008000600009
- Lopes, A.C.A., N.A. Vello, F. Pandini, M.M. Rocha, and C.Y. Tsutsumi. 2002. Variabilidade e correlações entre caracteres em cruzamentos de soja. *Sci. Agric.* 59:341–348. doi:10.1590/S0103-90162002000200021
- Marques, M.C., O.T. Hamawaki, T. Sedyama, M.R. Bueno, C.D. Cruz, M.S. Reis, and A.P.O. Nogueira. 2011. Adaptabilidade e estabilidade de genótipos de soja em diferentes épocas de semeadura. *Biosci. J.* 27:59–69.
- Mendes, F.F., M.A.P. Ramalho, and A.F.B. Abreu. 2009. Selection index for choosing segregating populations in common bean. *Pesquisa Agropecu. Bras.* 44:1312–1318.
- Nogueira, A.P.O., T. Sedyama, L.B. Sousa, O.T. Hamawaki, C.D. Cruz, D.G. Pereira, and E. Matsuo. 2012. Análise de trilha e correlações entre caracteres em soja cultivada em duas épocas de semeadura. *Biosci. J.* 28:877–888.
- Oliveira, B.A., J.B. Duarte, and J.B. Pinheiro. 2003. Emprego da análise AMMI na avaliação da estabilidade produtiva em soja. *Pesquisa Agropecu. Bras.* 38:357–364.
- Peluzio, J.M., R.D. Almeida, R.R. Fidelis, D.A. Júnior, E.L. Brito, and E.R. Francisco. 2005. Correlações entre caracteres de soja, em Gurupi, Tocantins. *Revista Ceres.* 52:779–786.
- Pimentel-Gomes, F. 2009. Curso de estatística experimental. 15th ed. Fund. Estud. Agrários Luiz de Queiroz, Piracicaba, Brazil.
- R Core Team. 2017. R: A language and environment for statistical computing. R Found. Stat. Comput., Vienna.
- Resende, M.D.V., and J.B. Duarte. 2007. Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Pesqui. Agropecu. Trop.* 3:182–194.
- Rigon, J.P.G., S. Capuani, J.F. Brito Neto, G.M. Rosa, A.D. Wastowski, and C.A.G. Rigon. 2012. Dissimilaridade genética e análise de trilha de cultivares de soja avaliada por meio de descritores quantitativos. *Rev. Ceres* 59:233–240. doi:10.1590/S0034-737X2012000200012
- Rincker, K., R. Nelson, J. Specht, D. Sleper, T. Cary, S.R. Cianzio, et al. 2014. Genetic improvement of US soybean in Maturity Groups II, III, and IV. *Crop Sci.* 54:1419–1432. doi:10.2135/cropsci2013.10.0665
- Scott, A.J., and M.A. Knott. 1974. Cluster analysis method for grouping means in the analysis of variance. *Biometrics* 30:507–512. doi:10.2307/2529204
- Sedyama, T. 2015. Melhoramento genético da soja. Vol. 1. Editora Mecenaz, Londrina, Brazil.
- Silva, K.B. 2016. Plasticidade fenotípica e análise de QTL para qualidade de sementes em soja. Ph.D. diss., Univ. Federal de Lavras, Lavras, Brazil.
- Silva, K.B., A.T. Bruzi, A.M. Zuffo, E.V. Zambiazzi, I.O. Soares, P.M. Rezende, et al. 2015. Adaptability and phenotypic stability of soybean cultivars for grain yield and oil content. *Genet. Mol. Res.* 15:1–15.
- Soares, I.O., P.M. Rezende, A.T. Bruzi, E.V. Zambiazzi, G. Felisberto, A.M. Zuffo, et al. 2015. Adaptability of soybean cultivars in different crop years. *Genet. Mol. Res.* 14:8995–9003. doi:10.4238/2015.August.7.8
- Vasconcelos, E.S., M.S. Reis, C.D. Cruz, T. Sedyama, and C.A. Scapim. 2010. Adaptability and stability of semilate and late maturing soybean genotypes in Minas Gerais state. *Acta Scientiarum* 32:411–415.
- Yan, W. 2011. GGE biplot vs. AMMI graphs for genotype-by-environment data analysis. *J. Indian Soc. Agric. Stat.* 65:181–193.
- Yan, W. 2013. Biplot analysis of incomplete two-way data. *Crop Sci.* 53:48–57. doi:10.2135/cropsci2012.05.0301
- Yan, W., and J.B. Holland. 2010. A heritability-adjusted GGE biplot for test environment evaluation. *Euphytica* 171:355–369. doi:10.1007/s10681-009-0030-5
- Yan, W., L.A. Hunt, Q. Sheng, and Z. Szlavnic. 2000. Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Sci.* 40:597–605. doi:10.2135/cropsci2000.403597x