

# Microsatellites and phenotypic characteristics used to select soybean lines in a Brazilian university breeding program

L.B. Sousa<sup>1</sup>, O.T. Hamawaki<sup>1</sup>, A.P.O. Nogueira<sup>3</sup>, D.B.O. Cardoso<sup>2</sup>, A.C. Dias<sup>4</sup>, V. Oliveira<sup>6</sup>, R.O. Batista<sup>7</sup>, A.M. Bonetti<sup>3</sup> and R.L. Hamawaki<sup>5</sup>

<sup>1</sup> Instituto de Ciências Agrárias, Universidade Federal de Uberlândia, Uberlândia, MG, Brasil

<sup>2</sup> Programa de Pós-Graduação em Agronomia, Instituto de Ciências Agrárias, Universidade Federal de Uberlândia, Uberlândia, MG, Brasil

<sup>3</sup> Instituto de Genética e Bioquímica, Universidade Federal de Uberlândia, Uberlândia, MG, Brasil

<sup>4</sup> Programa de Pós-Graduação em Genética e Bioquímica, Universidade Federal de Uberlândia, Uberlândia, MG, Brasil

<sup>5</sup> Southern Illinois University Carbondale, Carbondale, IL, United States

<sup>6</sup> Fazenda Prata, Corumbá de Goiás, GO, Brasil

<sup>7</sup> Instituto de Ciências Agrárias, Universidade Federal dos Vales do Jequitinhonha e Mucuri, Unaí, MG, Brasil

Corresponding author: D.B.O. Cardoso  
E-mail: danieludia13@hotmail.com

Genet. Mol. Res. 19 (1): gmr18105  
Received August 06, 2019  
Accepted January 27, 2020  
Published February 29, 2020  
DOI <http://dx.doi.org/10.4238/gmr18105>

**ABSTRACT.** Plant breeders need to understand the genetic diversity and relationships between genotypes to choose parents of segregating populations. We evaluated the genetic diversity of 35 soybean genotypes and determined their potential as parent lines in breeding programs by analyzing their agronomic traits and microsatellite markers. Phenotypic analysis was carried out in the field at Fazenda Capim Branco, a research station of the Federal University of Uberlândia. Nine microsatellite markers and seven agronomic traits were evaluated in 35 conventional soybean genotypes. Genotypes were grouped by UPGMA and Tocher cluster analyses, and molecular analysis was used to calculate polymorphism for each microsatellite locus. Nine microsatellite markers amplified 26 alleles, ranging from two to four. Polymorphism varied

between 0.29 and 0.66, and averaged 0.44. The number of days to maturity was the characteristic that contributed the most to yield (29.44%). Seven potential parents (G11, G12, G16, G21, G22, G26 and G33) had average grain yields higher than 5000 kg ha<sup>-1</sup>. UPGMA and Tocher methods identified groups from the matrix of phenotypic and molecular data. Hybrids of G11, G12, G16, G22, G26 and G33 genotypes segregated promising populations with superior genetic variability.

**Key words:** *Glycine max*; Diversity; SSR; Parent selection; Yield potential

## INTRODUCTION

Soybean (*Glycine max*) is the most important legume crop worldwide, and it is the main source of many products (grain, oil and soy flour) that are key to various production chains (Sediyama et al., 2015). The soybean cultivated area has undergone a great expansion in recent years, across many environmental conditions, from the highest to the lowest latitudes (Sediyama et al., 2012).

Brazil is the second largest soybean producer in the world. The planted area in the 2018/2019 season was 35.79 million hectares, with a total registered grain yield of 120 million tons. Brazil is also the world's largest soybean exporter and is expected to expand this lead in the coming years (CONAB, 2018).

Increases in soybean production are the consequence of the opening of new agricultural frontiers combined with progress in breeding, which obtained annual genetic gains of 0.59%, or 18.3 kg.ha<sup>-1</sup>, from 2006 to 2016 (Mezzalira, 2017), contributing to effectively meet growing food demand (Sentelhas et al., 2015). Brazilian breeding programs aim for adaptation to different environments by concentrating on seed physiological quality, nutritional composition of grain, suitability for human consumption, disease resistance, adaptability and, especially, grain yield (Almeida et al., 2011; Santos et al., 2011; Zhang et al., 2013). Genetic variability is essential to a breeding programs success. Breeders recommend out-crossing superior and divergent cultivars to form a base population. Hybrid combinations have high heterosis and segregated generations that also have a high chance to present transgressive genotypes (Almeida et al., 2011; Cruz et al., 2012; Streck et al., 2017).

Genetic divergence among parents can be measured based on agronomic, morphological and molecular traits (Santos et al., 2012). These traits allow the identification of parents to form segregating populations. Genetic complementarity and gamete fusion can bring out new genotypes via sexual reproduction processes (Cruz et al., 2011). However, Brazilian soybean germplasm originated from a narrow genetic base with few ancestral lineages (Yokomizo et al., 2013). It is not possible to determine the genetic dissimilarity between cultivars only by agricultural and morphological differences due to the large number of cultivars in the market. Because of this limitation, DNA markers have been widely used (Silva, 2015).

Molecular markers provide unique tools that can reveal polymorphisms in DNA sequence to discriminate genetic variation between individuals and within populations. These tools differ in variation, advantages and cost limitations, analysis time, accuracy and efficiency (Caixeta et al., 2013). Genetic patterns help to select germplasm sources for specific purposes (Iqbal et al., 2015).

Microsatellites or simple sequence repeats (SSR) have been widely used to evaluate soybean genetic diversity because they have a high degree of polymorphism, co-dominant inheritance, PCR-based detection, allelic diversity and the possibility to identify soybean genome markers position (Narvel et al., 2000; Priolli et al., 2010; Iqbal et al., 2015). The utilization of markers could organize germplasm into gene pools to simplify breeder choice and reduce the number of combinations.

Multivariate cluster analysis of phenotypic and molecular data indirectly selects genotypes and allows separation into homogeneous groups. This separation potentializes the heterogeneity between groups (Cruz et al., 2016). Despite their importance for breeding, studies based on molecular markers cannot replace essential field phenotyping evaluations. The genetic diversity analyses of phenotypic and genotypic traits complement each other. These analyses can select superior genotypes with greater accuracy and adopt introgression breeding methods that transfer valuable traits to commercial cultivars (Mulato et al., 2010).

We evaluated the genetic diversity of 35 soybean genotypes that have the potential of being parents in breeding programs, based on agronomic traits and microsatellite molecular markers.

## MATERIAL AND METHODS

The field experiment (phenotypic analysis) was carried out at the Fazenda Capim Branco (18°52' S; 48°20' W and 805 m of altitude), a research station of the Universidade Federal de Uberlândia (UFU), in Uberlândia, Minas Gerais, during the 2011/2012 growing season. The area where the experiment was carried out has soil classified as Latossolo Vermelho-Escuro (Brazilian classification system) and has been under soybean cultivation for over 10 years. A composite sample of soil was collected before starting the experiment for chemical and physical analyses, and for liming and fertilization purposes.

Soil preparation was done using a conventional tillage system consisting of one plow and two disc-harrow operations/passes. Before sowing, furrows were plowed across the area. The soil was fertilized according to the analysis and recommendations for the crop (EMBRAPA, 2013). We evaluated 35 soybean lines from 14 biparental crosses of the soybean breeding program at UFU. The experimental design was a randomized complete block with three replicates. The plot area consisted of two lines of soybeans, four meters long and spaced by 0.5 m. The seeds were treated with Fludioxonil and Thiamethoxam – commercially available as Maxim® XL and Cruiser® 350 FS, respectively – both at a rate of 200 mL of product for every 100 kg of seeds; seeds were inoculated with *Bradyrhizobium japonicum*. The seeding rate was 20 seeds per meter planted at three cm depth. The plants were thinned at the V1 stage (Fehr and Caviness, 1977) with the goal to establish a plant density of 30 plants per square meter, adopting a population density of 300.000 plants per hectare. Weed control was made with pre and post emergence herbicides, supplemented by hand-weeding when necessary. Diseases and insects were controlled according to appropriate technical recommendations for soybeans (EMBRAPA, 2013).

The agronomic traits that were evaluated are a priority in the UFU's breeding program. These traits were assessed by visual observations and measurements according to soybean stages of development, as proposed by Fehr and Caviness (1977):

a) Plant height at flowering (PHF): the heights, in centimeters, of five randomly selected plants were evaluated in each plot; they consist of the distance between the soil surface and the main stem apex of each plant.

b) Plant height at maturity (PHM): the heights of five plants, randomly selected in each plot, were measured, in centimeters. These measurements consisted of the distance between the soil surface and the main stem apex of each plant at the R8 reproductive growth stage, in which 95% of the pods would have reached their mature color.

c) Number of days to flowering (NDF): this corresponds to the number of days from seedling emergence to flowering when 50% of the plants have an open flower on the main stem (R1).

d) Number of days to maturity (NDM): this considers the number of days from emergence to physiological grain maturity in the field when 95% of pods are mature (R8).

e) Insertion Height of the first pod (IHFP): this is the distance (cm) from the soil surface to the insertion point of the first pod on the plant's main stem, measured in five plants at the R8 stage.

f) Total number of pods (TNP): five plants were measured at random in each plot. The number of pods per plant was counted and averaged;

g) Grain yield (GY) ( $\text{kg}\cdot\text{ha}^{-1}$ ): the grain yield was obtained from the harvest of each plot. Harvested material was threshed, and the grains were cleaned and weighed. The data (grams per plot) was converted to  $\text{kg}\cdot\text{ha}^{-1}$ , and yield was corrected to 13% moisture. The molecular analyses were carried out in the laboratory.

Ten leaves of each genotype were randomly selected, placed in paper bags and transported to the laboratory, where they were stored in silica gel. DNA extraction from soybean leaves was performed following the protocol described by Doyle and Doyle (1990), with some modifications as described below. An aliquot of 200 mg of leaf tissue was ground in liquid  $\text{N}_2$ , 1 mL of CTAB extraction buffer and 20 mg of proteinase K were added the ground tissue at 65° (Tris-HCl 50 mM pH 8.0; EDTA 50 mM pH 8.0; NaCl 0.7 M,  $\beta$ -mercaptoetanol 1% (v/v); CTAB 1% (w/v) and PVP). Samples were transferred to 2 mL microtubes, homogenized, and incubated in a 65°C water bath for 45 min. The tubes were inverted every 15 min.

After incubation, 1 mL of isoamyl alcohol:chloroform solution (24:1) was added to the samples. The resulting suspension was homogenized and centrifuged for ten minutes at 5,000 rpm. The supernatant was collected and transferred to microtubes. A solution of 95% ethyl alcohol: 7.5 M ammonium acetate (6: 1) was added in a volume that corresponded to three times the volume of the supernatant, and the product kept in a freezer (-20°C) overnight.

The material was centrifuged again for 10 min at 5,000 rpm and the supernatant was discarded. 200  $\mu\text{L}$  of TE buffer (Tris-EDTA) was added to the pellet. The DNA was precipitated again with the addition of a solution of 95% ethyl alcohol: 3M sodium acetate (20: 1) in the amount of three times the suspension volume and kept in a freezer (-20°C) for one hour. Then, the tubes were centrifuged for five minutes at 5,000 rpm. The supernatant was discarded and the microtubes were inverted until the pellet was dry. 200  $\mu\text{L}$  TE was added to the DNA. The samples were stored in freezer. DNA was quantified by spectrophotometric analysis (Nanodrop) and diluted to a final concentration of 10  $\text{ng}\cdot\mu\text{L}^{-1}$ . The integrity of DNA was checked on an agarose gel (0.8% w/v), stained with ethidium bromide.

The microsatellite markers were selected based on similar studies highlighting the most polymorphic markers as well as the distribution in different linking groups and soybean chromosomes. The sequences are available at <http://soybase.com>. The microsatellite markers used in this study were: Satt 317, Satt 436, Satt 233, Satt 191, Satt 197, Satt 487, Satt 180, Satt 309 and Satt 178 (Liu et al., 2011, Oda et al., 2015 and Tantasawat et al., 2011).

The PCR was performed in a final volume of 15  $\mu$ L. The reagents were: 1X reaction buffer containing  $MgCl_2$ , 0.15 mM dNTPs, 0.4  $\mu$ M of each primer; 20 ng genomic DNA; 1 U Taq polymerase (Code NB-NT-500D, Neotaq) and Milli-Q Water. The thermocycling consisted of an initial DNA denaturation for 4 min at 94°C, followed by 30 cycles each including the steps: DNA denaturation for 30 s at 94°C, primer annealing for 30 s at 50°C, and extension by the enzyme Taq DNA polymerase for 45 s at 72°C. After the cycles, a final 10 min extension period at 72°C was included.

The amplified fragments were separated on agarose gels (2% w/v) using 0.5 X TBE buffer (Tris-borate-EDTA) at 120 V for 2 h. The gels were stained with ethidium bromide, visualized under ultraviolet light, and the image capture was made using a photodocumentation system.

### Statistical analysis for agronomic data

Genetic dissimilarity among all pairs of genotypes was evaluated through the generalized Mahalanobis Distance ( $D^2_{ii'}$ ) according to the following estimator:

$$D^2_{ii'} = \delta' \Psi^{-1} \delta \quad (\text{Eq. 1})$$

Where:

$D^2_{ii'}$ : generalized distance of Mahalanobis between genotypes i and i';

$\Psi$ : Matrix of residual variances and covariance;

$\delta'$ : [ $d_1$   $d_2$ ...  $d_v$ ] where  $d_j = Y_{ij} - Y_{i'j}$ ;

$Y_{ij}$ : average of i-th genotype in relation to the j-th variable.

The genotypes were grouped by the hierarchical method of average linkage among clusters (UPGMA) and Tocher's optimization method (Rao, 1962) after the dissimilarity matrix between genotypes was prepared.

The hierarchical method of average linkage among clusters (Unweighted Pair Group Method with Arithmetic Average - UPGMA), was utilized to establish a dendrogram for the genotypes with high similarity. In this dendrogram, the distance between the genotype and the group of i and j is given by the equation:

$$d_{(ij)k} = \frac{d_{ik} + d_{jk}}{2} \quad (\text{Eq. 2})$$

The dissimilarity matrix was used, and then Tocher's optimization grouping method was applied (Rao, 1962). The first group consisted of genotypes whose dissimilarity was lower; later other genotypes were included in the group by comparing the increase in average value of distance in the group and a pre-established maximum permissible level of dissimilarity measure found in the set of smaller distances involving each genotype. The inclusion or not of each genotype was determined by:

$$\frac{d_{(\text{group})k}}{n} \leq \theta \text{ genotype } k \text{ was included in the group} \quad (\text{Eq. 3})$$

$$\frac{d_{(\text{group})k}}{n} > \theta \text{ genotype } k \text{ is not included} \quad (\text{Eq. 4})$$

Where:

$n$  = genotype number in the original group.

The distance between genotype  $k$  and the group formed by genotypes  $i$  and  $j$  was given by:

$$d_{(ij)k} = d_{ik} + d_{jk} \quad (\text{Eq. 5})$$

The relative contribution of each trait to genotypes dissimilarity was determined by the Methodology of Singh (1981) according to the  $S_{ij}$  statistic

$$D_{ii}^2 = \delta^{-1} \Psi^{-1} \delta = \sum_{j=1}^n \sum_{j'=1}^n \omega_{jj'} d_j d_{j'} \quad (\text{Eq. 6})$$

Where:

$\omega_{jj'}$  = the element of  $j$ -th inverse column of variance and residual covariance matrix.

### Statistical analysis of molecular data

The number of alleles per locus was determined from the gel profile analysis. The frequency of alleles was evaluated from the occurrence in the different genotype classes.

For the molecular data analysis, the polymorphic information content (PIC) of each microsatellite loci was calculated and evaluated through the frequency of alleles average using the equation:

$$\text{PIC} = 1 - \sum_{i=1}^a p_i^2 - \sum_{i,j=1}^a \sum_{i \neq j}^a p_i^2 p_j^2 \quad (\text{Eq. 7})$$

Where:

$A$  is number of locus alleles studied;

$p_i$  is the  $i$ -th frequency of locus allele studied;

$p_j$  is the  $j$ -th frequency of locus allele studied.

A locus was considered polymorphic when  $\text{PIC} \geq 0.1$  (corresponding approximately to the situation in which the most frequent allele has a frequency lower than 0.95) and highly polymorphic when  $\text{PIC} \geq 0.7$  (Cruz et al., 2011).

The number of alleles evaluated the genetic distance between pairs of genotypes using the index weighted complement ( $D = 1 - S$ ) as a dissimilarity measure is given by:

$$S_{ii} = \frac{1}{2} \sum_{j=1}^L p_j c_j \quad (\text{Eq. 8})$$

Where:

$p_j = \frac{a_j}{A}$  : Weight associated with locus  $j$  given by:

$C_j$ : total number of locus alleles  $j$ ;

$L$ : total number of studied alleles.

$$\sum_{j=1}^L p_j = 1 \quad (\text{Eq. 9})$$

Where:

$p_j$ : number of common alleles among pairs of genotypes  $i$  and  $i'$

$L$ : total number of studied alleles.

The genotypes were grouped by the hierarchical method of average linkage among clusters (UPGMA) and Tocher's optimization method (Rao, 1962), in the same way as the phenotypic data evaluation, and after obtaining the dissimilarity matrix between genotypes. The co-optic correlation coefficient was evaluated to determine the degree of fit between the dissimilarity matrix and the respective dendrogram for both analyses (phenotypic and molecular). The analyses were carried out with the Software for Experimental Statistics in Genetics (GENES) (Cruz, 2016).

## RESULTS

Significant differences ( $P \leq 0.05$ ) were detected by the F test for all traits, except for insertion height of the first pod. This was an indicator of the genetic variability among the genotypes. The different groups formed by the Scott-Knott test also highlighted the genetic variability for the traits (Table 1).

The number of generated groups varied from one in insertion height of the first pod (to four in number of days to maturity). Similar results were found by Almeida et al. (2013), who evaluated the agronomic performance of 30 soybean genotypes according to eight traits; they obtained from one to five groups.

Two groups were formed with the highest average (84.77) for G9 considering the total number of pods, followed by G28 (79.22). The shortest vegetative stage was the genotype G31 with 41 days to flowering; G3 had a delayed start of the reproductive stage at 53 days, but this was less than the highest average NDF found by Ferreira Júnior et al. (2015). These authors reported a value of 56.6 days to flowering in their evaluation of the genetic diversity and agronomic performance of advanced and superior soybean groups from biparental, quadruple, and octuple crosses.

The greatest plant height at flowering and maturity was found for G18 (52.40 cm) and G33 (78.20 cm), respectively. The plant height at flowering and maturity influenced plant lodging. According to Nogueira et al. (2009) plants that grow tall with thin stems tend to lodge easily.

Hamawaki et al. (2010) evaluated the agronomic performance of 24 soybean genotypes for the conditions of Minas Gerais and found an average plant height at flowering of 86.8 cm, higher than the one found in our study. The insertion height of the first pod ranged from 7.20 (G11) to 16.53 cm (G18), which is lower than that found by Amorim et al. (2011), who evaluated seven cultivars of soybean in found sowing stages and reported values between 18.05 and 15.73 cm.

Regarding the number of days to maturity, the earliest genotype was G11 at 110 days and the later, G3 at 139 days. G11 was superior to G3, since it presented precocity, which is one of breeding programs' goals, as well as a higher grain yield ( $5608.89 \text{ kg}\cdot\text{ha}^{-1}$ ), 23.93%. Genotype G22 had the highest yield,  $5884.44 \text{ kg}\cdot\text{ha}^{-1}$ , 38.41% higher than the total average obtained,  $4251.30 \text{ kg}\cdot\text{ha}^{-1}$ , and higher than the results obtained by Amorim et al. (2011),  $2953.71 \text{ kg}\cdot\text{ha}^{-1}$ .



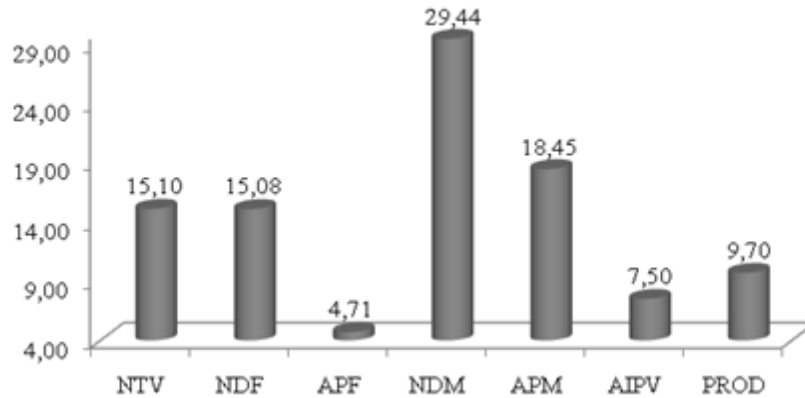
**Table 1.** Averages of seven agronomic traits evaluated in 35 cultivated genotypes of soybean in Uberlândia, MG\*.

GENOTYPES	TNP	PDF	PHF	NDM	PHM	IHFP	GY							
G1	54.80	b	48.33	A	45.27	a	126.00	C	62.13	b	12.27	a	4400.00	b
G2	58.53	b	48.67	A	39.40	b	131.33	B	69.47	a	12.40	a	4173.33	b
G3	41.65	b	52.33	A	44.67	a	139.00	A	72.92	a	12.58	a	4364.44	b
G4	51.00	b	47.00	A	41.47	b	127.33	C	54.73	c	11.33	a	3688.89	b
G5	77.23	a	46.67	A	39.20	b	125.00	C	54.47	c	10.00	a	3733.33	b
G6	55.47	b	49.33	A	48.33	a	137.00	A	59.80	b	9.13	a	4195.56	b
G7	41.67	b	47.67	A	45.40	a	126.67	C	62.20	b	10.67	a	4826.67	a
G8	56.80	b	44.33	B	35.27	b	117.33	d	43.73	c	8.87	a	3324.44	b
G9	84.77	a	50.33	A	51.73	a	126.67	C	61.07	b	11.13	a	4133.33	b
G10	48.27	b	48.00	A	48.53	a	123.67	C	68.67	a	11.13	a	4222.22	b
G11	58.03	b	42.67	B	39.53	b	110.00	d	53.00	c	7.20	a	5608.89	a
G12	63.95	a	49.00	A	45.53	a	128.00	C	59.53	b	8.07	a	5288.89	a
G13	69.03	a	46.33	A	43.07	b	130.00	B	68.27	a	9.87	a	2186.67	b
G14	52.67	b	47.00	A	41.93	b	123.67	C	60.33	b	10.60	a	4106.67	b
G15	61.63	a	49.33	A	42.47	b	123.67	c	61.40	b	12.27	a	3777.78	b
G16	68.30	a	47.00	A	51.37	a	122.00	c	62.00	b	9.20	a	5493.33	a
G17	67.00	a	48.33	A	52.33	a	126.67	c	59.87	b	11.60	a	4488.89	b
G18	69.00	a	51.67	A	52.40	a	126.00	c	64.93	a	16.53	a	4577.78	b
G19	73.60	a	44.00	B	32.53	b	121.33	c	57.07	b	9.27	a	3733.33	b
G20	58.17	b	48.33	A	43.80	a	126.00	c	50.57	c	11.20	a	4195.56	b
G21	34.53	b	47.00	A	39.33	b	126.00	c	58.93	b	11.53	a	5404.45	a
G22	45.07	b	46.33	A	44.80	a	126.00	c	57.80	b	13.07	a	5884.44	a
G23	61.33	a	48.33	A	47.07	a	127.33	c	63.00	b	10.60	a	4257.78	b
G24	68.62	a	45.33	B	35.13	b	132.00	b	55.00	c	9.87	a	4035.55	b
G25	39.67	b	48.00	A	43.87	a	128.00	c	58.47	b	11.87	a	4097.78	b
G26	61.27	a	49.33	A	48.47	a	132.00	b	59.33	b	12.13	a	5653.33	a
G27	75.07	a	48.00	A	47.07	a	126.00	c	61.67	b	9.53	a	4577.78	b
G28	79.22	a	47.67	A	41.53	b	122.00	c	54.53	c	10.53	a	3911.11	b
G29	55.87	b	48.00	A	48.53	a	123.33	c	66.07	a	12.27	a	4417.78	b
G30	67.30	a	46.67	A	35.67	b	126.00	c	46.13	c	9.93	a	3342.22	b
G31	56.23	b	41.00	B	36.80	b	112.00	d	55.60	c	10.60	a	4488.89	b
G32	64.17	a	43.00	B	40.60	b	112.00	d	56.67	b	11.00	a	2613.33	b
G33	67.63	a	47.67	A	50.00	a	127.33	c	78.20	a	14.07	a	5084.45	a
G34	50.82	b	46.67	A	43.13	b	126.67	c	58.00	b	9.93	a	3475.55	b
G35	67.33	a	45.67	B	39.20	b	114.67	d	53.00	c	9.20	a	3031.11	b
Average	60.16		47.29		43.58		125.10		56.67		10.90		4251.30	

The relative contribution of each agronomic trait from the genetic divergence evaluation can be shown by the multivariate analysis. The relative contribution of each trait is shown in Figure 1, highlighting number of days to maturity. This was the most important trait and it contributed with more than 29.44% of the divergence, followed by plant height at maturity, total number of pods, and number of days to flowering (18.45, 15.10 and 15.08%, respectively). The other traits had a small magnitude, indicating that they had relatively little participation in the divergence discrimination.

Rington et al. (2012) found similar results for some traits and different results for others when evaluating genetic divergence among 18 soybean cultivars based on six morphoagronomic traits. They concluded that the most important trait was insertion height at the first pod, with 23.02% importance, followed by number of days for maturity and plant height at maturity, with 23.28 and 18.80%, respectively. Bharadwaj et al. (2009) studied the genetic divergence among 85 soybean genotypes and two controls (commercial cultivars) concluded that the two most important trait for genotype divergence contributing 43% with all genetic dissimilarity; similar results were found in our study for two of the most important traits with 47.89%.

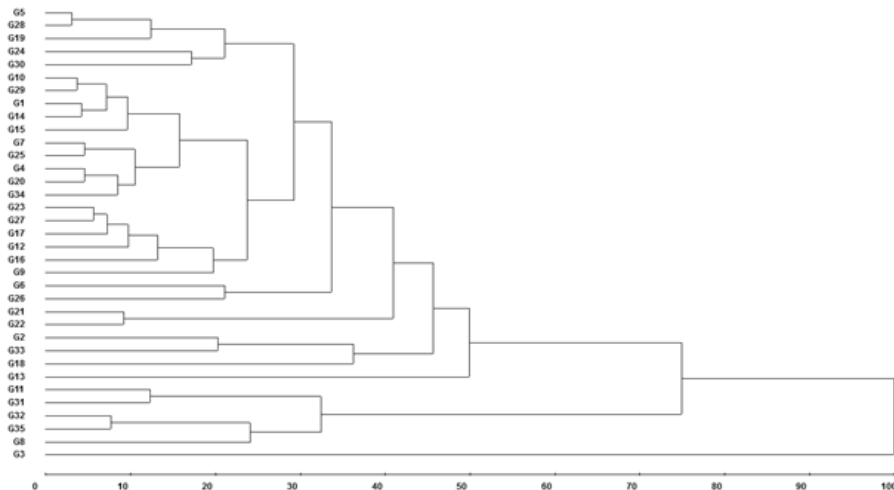




**Figure 1.** Relative contribution of the seven traits: total number of pods (TNP); number of days to flowering (NDF); number of days to maturity (NDM); plant height at maturity (PHM - cm), insertion height at the first pod (IHFP - cm) and grain yield (GY - kg.ha<sup>-1</sup>), to quantify the genetic divergence among 35 soybean genotypes.

Grain yield had a contribution of 9.70%. This indicates that the higher the productivity variation among genotypes, the greater genetic divergence between genotypes. This result was higher than what Pelúzio et al. (2009) found, who evaluated genetic divergence among 14 soybean cultivars and obtained a relative contribution to grain yield of 3.74%. Similar results were found by Bharadwaj et al. (2009) (3.4%).

Figure 2 is a dendrogram showing dissimilarity of genotypes based on agronomic traits. The cophenetic correlation coefficient was 0.72 indicating that the graphical representation is in agreement with dissimilarities among the pairs of genotypes.



**Figure 2.** Dendrogram of genetic dissimilarity among 35 soybean genotypes resulting from the cluster analysis by UPGMA method and according to the generalized distance of Mahalanobis ( $D^2$ ). Cophenetic correlation coefficient ( $r$ ): 0.72\*\*. \*\* Significant at 1% probability by the T test.

According to Barroso and Artes (2003), a cophenetic correlation coefficient equal or higher than 0.70 shows a good fit between graphical representation of genetic distances, the genotypes, and their original matrix. Accordingly, the results in this study (0.72) exceeded the established limit and were similar to those of other studies that evaluated divergence in soybeans (Barroso and Artes, 2003; Bharadwaj et al., 2009; Pelúzio et al., 2009; Mannan et al., 2010).

The UPGMA criterion established an initial group of similar genotypes. Then, the distances were calculated considering the first group formed (Cruz and Carneiro, 2006). Group delimitation is usually subjective, but considers points of high-level change in the dendrogram (Cruz et al., 2011).

When 27% dissimilarity was used as a cutoff in the dendrogram, which is a significant threshold by the Mojema test, the genotypes were separated into 10 groups. Three of the groups, G18, G13 and G3, remained alone, thus presenting high divergence in relation to the others.

The genotypes in each group:

- Group 1: G5, G28, G19, G24 and G30;
- Group 2: G10, G29, G1, G14, G15, G7, G25, G4, G20, G34, G23, G27, G17, G12, G16 and G9 (group 2 consisted of the greatest proportion of genotypes studied, 43%);
- Group 3: G6 and G26;
- Group 4: G21 and G22;
- Group 5: G2 and G33;
- Group 6: G18;
- Group 7: G13;
- Group 8: G11 and G31;
- Group 9: G32, G35 and G8;
- Group 10: G3.

Group 1 comprised the genotypes with yields above 3300 kg.ha<sup>-1</sup>, medium to late cycle, over 120 days, and lower insertion height of first pod. Group 2 was composed of several genotypes for the traits, from high yield genotypes, for example G16 (5493.33 kg.ha<sup>-1</sup>), to genotypes with below-average yield of 3475.55 kg.ha<sup>-1</sup>, for example G34.

Group 3 allocated genotypes later than those in group 1, with 134.5 days of NDM average and PHM of 59 cm. Group 4 had the highest grain yield average, 5644.45 kg.ha<sup>-1</sup>, and same parental lines. Both groups 6 and 7 were formed by one representative genotype each and comprised the genotypes with the highest IHFP (16.53 cm) and lowest GY (2186.67 kg.ha<sup>-1</sup>), respectively. Genotypes of group 9 had the second lowest average among the groups, 2989.62 kg.ha<sup>-1</sup>, and the genotypes of group 10 presented a yield of 4364.44 kg.ha<sup>-1</sup>.

The Mulato et al. (2010) study used the UPGMA method and obtained five groups; the group with the largest number of genotypes had only 29.1% of the genotypes, which was lower than the results found in our study. These group formations through the UPGMA method are useful for parents selection in future crosses since the new hybrid combinations must be based on the magnitude of their dissimilarities and potential per se (Almeida et al., 2011). Thus, genotypes in different groups indicate higher dissimilarity and are promising for artificial hybridizations. However, in addition to divergence, parents need to associate high averages and variability in traits (Sousa et al., 2015).

According to grouping by Tocher's method (Rao, 1962) based on the dissimilarity matrix expressed by the Mahalanobis distances ( $D^2$ ), the distribution of genotypes gave 12 groups, two more than the grouping by the UPGMA method (Table 2).

**Table 2.** Grouping of 35 soybean genotypes by Tocher's method and generalized distance of Mahalanobis as a measure of genetic distance based on the seven agronomic traits.

Group	Number of genotypes	Genotypes
1	17	G5, G28, G27, G23, G17, G1, G29, G14, G15, G12, G34, G4, G20, G7, G10, G25, G16
2	3	G32, G35, G8
3	2	G21, G22
4	2	G11, G31
5	2	G24, G30
6	2	G2, G33
7	2	G6, G26
8	1	G9
9	1	G3
10	1	G18
11	1	G13
12	1	G19

\* Averages followed by equal letters belong to the same group according to the Scott-Knott test at 10% of probability. TNP: total number of pods (TNP); NDF: number of days to flowering (NDF); PHF: plant height at flowering (PHF) in cm; NDM: number of days to maturity (NDM); PHM: plant height at maturity (PHM) in cm; IHFP: insertion height of the first pod (IHFP) in cm and GY: grain yield (GY) in kg/ha.

Group 1 was of major importance because it included almost half of the genotypes. Group constitution had a similarity between the methodology of intra-group linkage average and Tocher's optimization method. Genotypes classification in groups based on the two methodologies was predominantly coincident, but presented some exceptions. Group G28 was in same group as G24 and G30 according to the UPGMA, but G28 was in separate group considering Tocher's method. G19 was isolated by Tocher's methodology, but in the UPGMA it was in the same group as G5 and G28 genotypes, which agrees with what was found by Mulato *et al.* (2010) and Sousa *et al.* (2015).

Table 3 shows microsatellite loci distributed across eight linking groups of soybean genome (H, D1a, A2, G, B1, C1 and K) according to genetic divergence evaluation among genotypes by molecular analysis using microsatellite markers.

**Table 3.** Polymorphic primers used in evaluation of 35 soybean genotypes, specifying genome localization (GL), chromosome (Cr), allele frequency and polymorphic information content (PIC) values.

Primers <sup>1</sup>	GL	Cr.	Frequency of allele				Alleles number	PIC
			1	2	3	4		
Satt 317	H	12	0.14	0.77	0.08		3	0.34
Satt436	D1a	1	0.20	0.28	0.51		3	0.54
Satt 233	A2	8	0.26	0.74			2	0.30
Satt 191	G	18	0.20	0.57	0.23		3	0.51
Satt 197	B1	11	0.11	0.23	0.34	0.31	4	0.66
Satt 487	O	10	0.08	0.80	0.11		3	0.31
Satt180	C1	4	0.14	0.25	0.60		3	0.48
Satt 309	G	18	0.23	0.54	0.23		3	0.53
Satt 178	K	9	0.23	0.77	0.08		2	0.29
Total							26	
Average							2.88	0.44

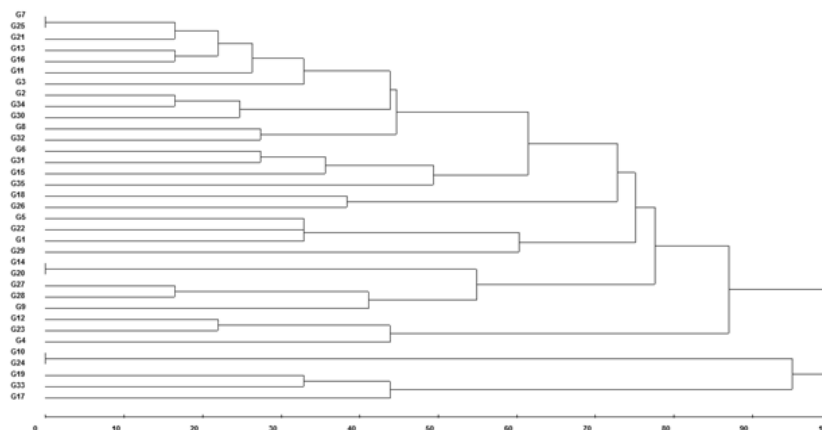
Nine microsatellite markers amplified 26 alleles, ranging from 2 to 4 and averaging 2.88 per marker. Microsatellites - (Satt 233), (Satt 317, Satt 436, Satt 191, Satt 487, Satt 180, 309 Satt and Satt 178) and (Satt 197) - had two, three and four alleles, respectively, similar to the results obtained by Vieira et al. (2009). They analyzed 53 soybean cultivars with 111 polymorphic primers and found one band of 2 to 4 alleles per SSR locus, with an average of 2.20.

Zhang et al. (2013) evaluated the genetic divergence among 48 soybean genotypes using 22 SSR markers and found 71 alleles with an average of 3.23 alleles per locus. Ribeiro et al. (2013) genotyped 30 soybean cultivars and found 2 to 7 alleles per locus, with an average of 3.85, which is 34% higher than the average found our study. Evaluating the genetic divergence through microsatellite markers of 138 soybean cultivars, Priolli et al. (2010) found from 2 to 8 alleles per SSR locus.

Allele frequencies data calculates the probability of sibling identity and probability of exclusion, thus indicating whether two samples have the same genotype (Schuster et al., 2006). The allele frequency varied from 0.08 to 0.80 when primers 178 Satt and Satt 487 were used.

Polymorphic information content (PIC), which reflects the frequency and allelic diversity among the genotypes, ranged from 0.29 (Satt 178) to 0.66 (Satt 197), with an average of 0.44. This value was higher than that found by Zhang *et al.* (2013), who reported values of PIC between 0.14 and 0.63, with an average of 0.38. Liu et al. (2011) found values between 0.11 and 0.60 with an average of 0.26 when evaluating 91 soybean cultivars. Differences in the content data of microsatellite markers occurred because of the number of genotypes involved in each evaluation since there is a greater chance to find alleles when the number of genotypes is also high. The highest PIC value was 0.66 for the primer Satt 197, which was the most polymorphic. To be considered polymorphic, the locus should have a value higher than or equal to 0.7 (Cruz et al., 2011).

Genetic dissimilarity between genotypes utilized the complement of the index weighted by the number of alleles according to the dendrogram based on the dissimilarity matrix (Figure 3). This dissimilarity showed that the cophenetic correlation coefficient was 0.74, significant at 1% probability by the *T*-test, which indicates a good fit between graphical representation of genetic distances between genotypes and their original matrix. Other studies had similar results when they evaluated genetic divergence using microsatellite markers (Singh et al., 2010; Liu et al., 2011; Jun et al., 2011; Tantasawat et al., 2011; Ribeiro et al., 2013; Zhang et al., 2013).



**Figure 3** Dendrogram generated through the average linkage grouping (UPGMA) methodology based on genetic dissimilarity matrix composed of 35 soybean genotypes and evaluated by utilizing nine microsatellite markers. Cophenetic correlation coefficient ( $r$ ): 0.74, which is significant at 1% probability by the *T* test.

Eleven groups were formed with different numbers of genotypes each when using 48% of dissimilarity as a cutoff in the dendrogram. Group 1 had 12 genotypes (Figure 3). The clustering of genotypes by Tocher's methodology (Table 4) and the molecular analysis was in accordance with the dendrogram. However, G9 was observed as an additional group of isolated genotypes. G7 was not in the same group as G8 in the UPGMA.

**Table 4.** Grouping of 35 soybean genotypes by Tocher's method and the weighted dissimilarity for the number of alleles matrix based on nine microsatellites markers.

Group	Number of genotypes	Genotypes
1	11	G7, G25, G13, G16, G21, G11, G32, G34, G3, G2, G30
2	2	G10, G24
3	4	G14, G20, G27, G28
4	3	G12, G23, G4
5	3	G6, G31, G15
6	3	G1, G5, G22
7	2	G8, G17
8	2	G19, G33
9	2	G18, G26
10	1	G35
11	1	G29
12	1	G9

Sequence of primers available at <http://soybase.org/resource/ssr.php>

The genetic divergence study aims to identify parents who present ideal and divergent traits. In this context, considering that a fundamental trait in soybean breeding is grain yield, seven potential parents are displayed in Table 5, which had averages higher than 5000 kg.ha<sup>-1</sup>. Among the seven genotypes, some were in the same group according to groupings through the UPGMA and Tocher's method, obtained from the matrix of phenotypic and molecular data.

**Table 5.** Grouping of potential genotypes according to dendrogram obtained from the UPGMA and Tocher's method.

Group	Grouping of genotypes	
	Phenotypic data	Molecular data
A	G12 e G16	G11, G16 e G21
B	G26	G26
C	G21 e G22	G22
D	G33	G12
E	G11	G33

Considering biparental crosses, 21 combinations would be possible assuming that using divergent and high yielding parents is advisable. Only 11 crosses of the total combinations are possible (Table 6), because some crosses that are indicated by clusters of phenotypic data cannot be found based on molecular data and vice versa.

**Table 6.** Possible biparental crossings of soybean considering seven potential parents.

Crossings			
G12	G16		G26
G12 x G26	G16 x G22		G26 x G33
G12 x G21	G16 x G33		G26 x G11
G12 x G33	G16 x G33		G26 x G22
G12 x G11			
G12 x G22			

## CONCLUSIONS

Agronomic traits and microsatellite markers are effective tools for soybean genetic diversity studies. These tools allowed detecting significant genetic variability among soybean genotypes. Hybrids among G11, G12, G16, G22, G26 and G33 genotypes are promising for developing segregating populations with superior genetic variability.

## ACKNOWLEDGMENTS

The authors thank FAPEMIG (Fundação de Amparo à Pesquisa de Minas Gerais), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and The Cotton Breeding Program of UFU for assistance and financial support.

## CONFLICTS OF INTEREST

The authors declare no conflict of interest.

## REFERENCES

- Almeida FA, Bruske EL, Polizel AC, Petter FA, et al. (2013). Desempenho agrônômico de linhagens e cultivares de soja frente a doenças foliares. *Rev. Ciênc. Agrár.* 56: 88-94.
- Almeida RD, Pelúzio JM and Afférris FS (2011). Divergência genética entre cultivares de soja, sob condições de várzea irrigada, no sul do Estado Tocantins. *Rev. Ciênc. Agron.* 42: 108-115.
- Amorim FA, Hamawaki OT, Sousa LB, Lana RMQ, et al. (2011). Época de semeadura no Potencial produtivo de Soja em Uberlândia-MG. *Semina: Ciênc. Agrár.* 32: 1793-1802.
- Barroso LP, Artes R (2003). Análise multivariada. Lavras: UFLA. 151p.
- Bharadwaj CH, Satyavathi TC, Husain SM, Chauhan GS et al. (2009). Divergence studies in early-maturing soybean [*Glycine max* (L.) Merrill] germplasm accessions in India. *Plant Genet. Resour. Newsl.* 149: 17-21.
- Caixeta ET, Ferrão LFV and Maciel-Zambolim E (2013). Marcadores moleculares. In: Borém, A.; Fritsche-Neto, R. *Biotecnologia aplicada ao melhoramento de plantas*. Visconde do Rio Branco: Suprema, p. 31-68.
- CONAB. Levantamento de Safras. Available at: [https://www.conab.gov.br/info-agro/safras/graos]. Accessed January 2, 2019
- Cruz CD and Carneiro PCS (2006). Modelos biométricos aplicados ao melhoramento genético. Viçosa, MG: UFV. volume 2, 585 p.
- Cruz CD, Ferreira FM and Personi LA (2011). Biometria aplicada ao estudo da diversidade genética. Visconde do Rio Branco, MG: Suprema, 620 p.
- Cruz CD, Regazzi AJ and Carneiro PCS (2012). Modelos biométricos aplicados ao melhoramento genético. v. 1, 4. ed. Viçosa: UFV, 514 p.
- Cruz CD (2016). Genes Software – Extended and integrated with the R, Matlab and Selegen. *Acta Sci., Agron.* 38: 547-552.
- Empresa Brasileira de Pesquisa Agropecuária. Tecnologias de produção de soja – Região Central do Brasil 2014. – Londrina: Embrapa Soja. 265p

- Fehr WR and Caviness CE (1977). Stages of soybean development. Ames: Iowa State University. Special Report, 80, Iowa Cooperative Extensive Service, Iowa, 12 p.
- Ferreira Júnior JA, Unêda-Trevisoli SH, Espíndola SMG, Vianna VF, et al. (2015). AO. Diversidade genética em linhagens avançadas de soja oriundas de cruzamentos biparentais, quádruplos e ócuplos. *Rev. Ciênc. Agron.* 46: 339-351.
- Hamawaki OT, Sousa LB, Rezende DF and Hamawaki CDL (2010). Yield and agronomic characteristics of soybean breeding lines in Minas Gerais State, Brazil. *Commun. Sci.* 1: 43-47.
- Iqbal Z, Naeem R, Ashraf M, Arshad M, et al. (2015). Genetic diversity of soybean accessions using seed storage proteins. *Pak. J. Bot.* 47: 203-209.
- Jun TH, Michel AP and Mian MAR (2011). Development of soybean aphid genomic SSR markers using next generation sequencing. *Genome.* 54: 360-367.
- Liu M, Zhang M, Jiang W, Sun G, et al. (2011). Genetic diversity of Shaanxi soybean landraces based on agronomic traits and SSR markers. *Afr. J. Biotechnol.* 10: 4823-4837.
- Mannan MA, Karim MA, Khaliq QA, Haque MM, et al. (2010). Assessment of genetic divergence in salt tolerance of soybean (*Glycine max* L.) genotypes. *J. Crop Sci. Biotech.* 13: 33-37.
- Mezzalira I (2017). Ganho genético para produtividade de grãos de soja na região central do Brasil / Itamara Mezzalira. Master's thesis - Universidade Federal de Lavras.
- Mulato BM, Moller M, Zucchi MI, Quecini V, et al. (2010). Genetic diversity in soybean germplasm identified by SSR and EST-SSR markers. *Pesq. Agropec. Bras.* 45: 276-283.
- Narvel JM, Fehr WR, Chu WC, Grant D, et al. (2000). Simple sequence repeat diversity among soybean plant introductions and elite genotypes. *Crop Sci.* 40: 1452-1458.
- Nogueira APO, Sedyama T, Barros HB and Teixeira RC (2009). Morfologia, crescimento e desenvolvimento. In: Sedyama, T. (Ed). Tecnologias de produção e usos da soja. Londrina: Mecenas, 7-16p.
- Oda MC, Sedyama T, Matsuo É, Cruz CD, et al. (2015). Phenotypic and molecular traits diversity in soybean launched in forty years of genetic breeding. *Agron. Sci. Biotechnol.* 1: 1-9.
- Pelúzio JM, Vaz-de-Melo A, Afféri FS, Silva RR, et al. (2009). Variabilidade genética entre cultivares de soja, sob diferentes condições edafoclimáticas. *Appl. Res. Agrotec.* 2(3): 21-29.
- Priolli RHG, Pinheiro JB, Zucchi MI, Bajay MM, et al. (2010). Genetic diversity among Brazilian soybean cultivars based on SSR loci and pedigree data. *Braz. Arch. Biol. Technol.* 53: 519-531.
- Rao RC (1962). Advanced statistical methods in biometric research. New York: John Wiley. 390p.
- Ribeiro CAG, Tanure JPM, Maciel TEF and Barros EG (2013). Molecular characterization of soybean cultivars by microsatellite markers with universal tail sequence. *Pesqu. Agropec. Bras.* 48: 270-279.
- Ringon JPG, Capuani S, Brito Neto JF, Rosa GM, et al. (2012). Dissimilaridade genética e análise de trilha de cultivares de soja avaliada por meio de descritores quantitativos. *Rev. Ceres.* 59: 233-240.
- Santos ER, Barros HB, Ferraz EC, Cella AJS, et al. (2011). Divergência entre genótipos de soja, cultivados em várzea irrigada. *Rev. Ceres.* 58: 755-764.
- Santos ER, Barros HB, Capone A, Melo AV, et al. (2012). Divergência genética entre genótipos de soja com base na qualidade de sementes. *Rev. Bras. Ciênc. Agrár.* 7: 247-254.
- Sedyama T, Silva F and Borm MA (2015). Soja: do plantio à colheita. Viçosa. Editora UFV.
- Sedyama CS, Carneiro JES, Fritsche-Neto R, Sedyama T, et al. (2012). Contribution of the universities to the development of field crop cultivars. *Crop Breed. Appl. Biotechnol.* 12: 121-130.
- Sentelhas PC, Battisti R, Câmara GMS, Farias JRB, et al. (2015). The soybean yield gap in Brazil—magnitude, causes and possible solutions for sustainable production. *J. Agric. Sci.* 153: 1394-1411.
- Schuster I, Vieira ESN and Padilha L (2006). Marcadores moleculares no pós-melhoramento. In: BORÉM, A.; CAIXETA, E. T. (Eds). Marcadores moleculares. Viçosa: UFV, 205-230p.
- Silva AF, Sedyama T, Matsuo E, Silva FCS, et al. (2015). Phenotypic and molecular diversity among soybean cultivars as a function of growing season. *Agron. Sci. Biotechnol.* 1(2): 52-61.
- Singh D (1981). The relative importance of characters affecting genetic divergence. *Indian. J. Genet. Pl. Br.* 41: 237-245.
- Sousa LB, Hamawaki OT, Batista RO, Berton I, et al. (2015). Genetic variability among soybean biparental crosses evaluated by multivariate analysis. *Biosci. J.* 31: 1404-1412.
- Streck EA, Aguiar GA, Júnior Magalhães AM, Facchinello PHK, et al (2017). Variabilidade fenotípica de genótipos de arroz irrigado via análise multivariada. *Rev. Ciênc. Agron.* 48:101-109.
- Tantasawat P, Trongchuen J, Prajongjai T, Jenweerawat S, et al. (2011). SSR analysis of soybean [*Glycine max* (L.) Merrill] genetic relationship and variety identification in Thailand. *Aust. J. Crop Sci.* 5: 283-290.
- Vieira ES, Schuster I, Silva RB and Oliveira MAR (2009). Variabilidade genética em cultivares de soja determinada com marcadores microsatélites em gel de agarose. *Pesq. Agrop. Bras.* 44: 1460-1466.
- Yokomizo GKI (2013). Análise AMMI da produtividade de grãos em linhagens de soja selecionadas para resistência à ferrugem asiática. *Pesq. Agropec. Bras.* 48: 1376-1384.
- Zhang GW, Xu SC, Mao WH, Hu QZ, et al. (2013). Determination of the genetic diversity of vegetable soybean [*Glycine max* (L.) Merr.] using EST-SSR markers. *J. Zhejiang Univ. Sci. B.* 14: 279-288.