Original Article

DISSIMILARITY MEASURES AND HIERARCHICAL METHODS FOR THE STUDY OF GENETIC DIVERSITY ON SOYBEAN

MEDIDAS DE DISSIMILARIDADE E MÉTODOS HIERÁRQUICOS PARA ESTUDO DA DIVERSIDADE GENÉTICA EM SOJA

Bárbara RODRIGUES[1]; Ana Paula Rodrigues GOMES[1]; Josiane Dias GOMES[1]; Fábio SERAFIM[2]; Ana Paula Oliveira NOGUEIRA[1]; Cristiane Divina Lemes HAMAWAKI[3]; Raphael Lemes HAMAWAKI[4]; Osvaldo Toshiyuki HAMAWAKI[1]

1. Universidade Federal de Uberlândia, Uberlândia, MG, Brasil. barbara_2730@msn.com; 2. Universidade Federal de Lavras, Lavras, MG, Brasil, 3. Instituto Master de Ensino Presidente Antônio Carlos, Araguari, MG, Brasil; 4. Southern Illinois University Carbondale, Carbondale, Illinois, United States.

ABSTRACT: In analysis of the genetic diversity on soybean can be used agronomic, morphological and molecular traits, which are subjected to multivariate biometrical analysis. There are different multivariate methodologies available such as Euclidean distance, Mahalanobis distance and different hierarchical methods. However, studies that may assist in the choice of such methods are lacking. The aim of this paper was to evaluate the clustering standards of soybean genotypes using Euclidean and Mahalanobis distances, following different hierarchical methods. The experiment was conducted in "Capim Branco" farm which belongs to the Federal University of Uberlândia and were used a complete randomized block design composed of 15 soybean genotypes (nine breeding lines and six cultivars) and four replications. The agronomic traits evaluated were: number of days to flowering and to maturity, height of the plant at flowering and at maturity, height of the insertion of the first pod, number of nodes on the main stalk in flowering and at maturity, number of grains per pod, total number of pods, severity of Asian rust, number of pustules and yield. The data were submitted to multivariate analysis in GENES program. The Mahalanobis distance or the Euclidean distance obtained by agronomic traits allows the determination of soybean genetic diversity. The use of the Euclidean distance in hierarchical methods allows a greater group differentiation. The UPGMA method and the nearest neighbor method shows a greater accuracy using the Mahalanobis distance and Euclidean distance.

KEYWORDS: *Glycine max*. Genetic variability. Euclidean distance. Mahalanobis distance.

INTRODUCTION

The soybean (*Glycine max* L. Merrill) is a leguminous of great importance in the world scenario, such in the food context as in the economic context (OLIVEIRA et al., 2015).

The main aim of soybean breeding programs, through the selection of plants that shows better characteristics, is to obtain genotypes with high grain productivity and resistance to biotic and abiotic factors, for the development of new cultivars that exceed the commercials ones (NOGUEIRA et al., 2015). For the success of those programs is important the existence of genetic variability (ALMEIDA et al., 2011).

The adoption of strategies which aims to provide a greater selection gain is indispensable, considering the large number of genes involved in the control of quantitative traits.

In this way, breeders have recommended for the formation of a population base, the intercrossing between superior cultivars and divergent ones, resulting in hybrid combinations of a greater heterozygosis, in a way that the segregated generations would have a bigger possibility of obtaining superior genotypes (ALMEIDA et al., 2011).

In genetic diversity studies, agronomic, morphological and molecular traits can be used in multivariate biometric analysis. Among the existing techniques for the genetic divergence study, the most applied are dissimilarity measures, clustering methods, principal components and canonic variables, allowing to unify multiple information from a set of traits resulting in a bigger opportunity of choice between divergent parents in breeding programs (CRUZ et al., 2011).

The multivariate statistic technics are based in measures of dissimilarity, such as Euclidean distance, Mahalanobis distance and other different measures for molecular characters, being the application of such measures useful in genetic breeding programs by allowing the gathering of information about the degree of resemblance or difference between two or more genotypes (CRUZ et al., 2014).

To quantify the genetic difference between pairs of genotypes, can be used different measures of dissimilarity, such as Euclidean distance and the Mahalanobis distance. Both emphasize variations in

Received: 21/03/17
Accepted: 05/07/17

morphological, agronomic and physiological characteristics. The Mahalanobis distance (D^2) can be used as genetic diversity estimation when several traits are measured in distinct genotypes (Elias et al., 2007). However, the Euclidean distance can be determined just with averages, while the Mahalanobis distance adopts the averages, considering correlated traits and the residual covariance matrix, being needed an experiment with repetitions (CRUZ et al., 2014).

The use of methods that allow grouping genotypes based on some similarity or dissimilarity measures has the purpose of separate an original group of observation in several subgroups, in a way to obtain homogeneity inside such subgroups and heterogeneity among them. Thus, they are also alternatives to analyze and understand the data (CRUZ et al., 2014).

The hierarchical methods are applied in a widespread range by breeders. They determine the genetic diversity among the genotypes, in which the genotypes are grouped by a process that repeats in various levels according to the shorter distances. In that way, is established a dendrogram, where the groups are formed from the delimitation of a cutting line, that is made in a subjective way, without worrying about the optimal number of groups.

There are different hierarchical methods available that allows to obtain dendrograms from a distance matrix such as average linkage among groups (UPGMA), method of the nearest neighbor or single bond, method of the farther neighbor or complete connection and Ward's minimum variance method (CRUZ et al., 2011).

Therefore, there are several methodologies to study the genetic divergence. They are based on hierarchical methods and it is up to the researcher to choose the one that best fits to the set of data. Nowadays, many studies of genetic diversity have been conducted in soybean, and in most of them the UPGMA method is used (MATSUO et al., 2011; MIRANDA et al., 2007; PRIOLLI et al., 2010; SANTOS et al., 2013; SINGH et al., 2010; VAL et al., 2014; YAMANAKA et al., 2007). In this way, there is a lack of studies about concordance pattern of genotype grouping when distance measures and different grouping methods are adopted. Therefore, the objective of this work was to evaluate the genetic divergence of soybean cultivars from Euclidean and Mahalanobis distance with different hierarchical methods.

MATERIAL AND METHODS

Implantation and conduction of the experiment

The trial was carried out in the experimental area of the Soybean Breeding Program, situated on the Capim Branco Farm (18° 52' S; 48° 20' W and 805m of Altitude), belonging to the Federal University of Uberlândia, located in the city of Uberlândia - MG, on the December 12, 2013.

The treatments were composed of 15 soybean genotypes, nine lines developed by the Soybean Breeding Program of the Federal University of Uberlândia, and six cultivars.

The complete randomized block with four replications experimental design was used. The plots were composed of four 5 m long rows of soybean plants, spaced 0.50 m between rows, and the usable area represented by the two center rows discarding 0.50 m of each extremity.

The experiment was installed using the conventional sowing system, in an area of dystrophic dark red latosol, which has been under soybean cultivation for several years. The soil was prepared with a plowing, two gradations, followed by the groove prior to sowing. Fertilization was performed with formula NPK 02-20-10.

Sowing was performed manually, using a density of 12 plants per linear meter. After the sowing, seed treatment was carried out with the application of a directed jet of fungicide of carboxanilide chemical composition (Carboxin) and Dimethyldithiocarbamate -trade name Vitavax - Thiran 200 SC, at a dose of 250 ml of commercial product per 100 kg of seed and insecticide Thiamethoxam, trade name Cruiser 350FS, in a dose of 200 mL per 100 kg of seed, as well as the inoculation with the bacteria *Bradyrhizobium japonicum* using the liquid inoculant Masterfix.

For weed control, pre-emergent applications were made using 2.0 L/ha^{-1} of the herbicide Dual Gold (S-metolachlor) and a post-emergent application for weed control with Cobra (lactofen) (0.4 L/ha^{-1}) + Classic® (Chlorimuron-ethyl) (40 g/ha^{-1}), to control invasive plants with wide leaves and narrow leaves, respectively.

Traits Evaluated

The stages of development in soybean were identified according to the scale of Fehr and Caviness (1977). All the evaluations were conducted on five plants, randomly sampled from each plot, and the following traits were evaluated:

a) Number of days to flowering (NDF): the

period that corresponds to the number of days elapsed between emergence and flowering (stage R1).

b) Number of days to maturity (NDM): the period that corresponds to the number of days elapsed from emergence to the date that 95% of the pods mature (stage R8).

c) Height of the plant at flowering (HPF): measured in centimeters from the surface of the soil to the last node on the main stalk in stage R1.

d) Height of the plant at maturity (HPM): measured in centimeters from the surface of the soil to the last node of the main stalk in stage R8.

e) Height of the insertion of the 1st pod (HIP): measured in centimeters from the surface of the soil to the insertion of the first pod on the main stalk.

f) Number of nodes on the main stalk in flowering (NNF): direct counting of all the visible nodes on the main stalk in stage R1.

g) Number of nodes on the main stalk at maturity (NNM): Direct counting of all the visible nodes on the main stalk in stage R8.

h) Number of grains per pod: after harvesting, the total number of pods on each plant was counted. They were then ranked as NP1G (Number of pods with one grain), NP2G (Number of pods with two grains), and NP3G (Number of pods with three grains) and subsequently the TNP (Total number of pods) per plant was calculated.

i) Severity of Asian rust (SEV) (%): evaluations were carried out on five central trefoil leaflets per plot from the appearance of the first pustules until total defoliation, and the average severity was an estimate of the disease average in the plot. The diagrammatic scale shown below was adopted (GODOY et al., 2006).

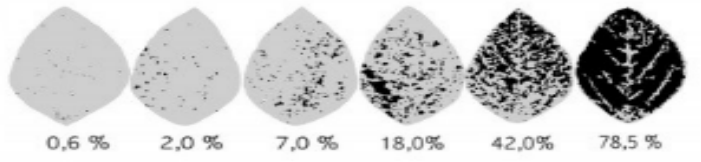


Figure 1. Diagrammatic scale to evaluate the severity of Asian rust (GODOY et al., 2006).

j) Number of Pustules (NP): the number of pustules per cm^2 on the central leaflet of the middle third of the plant in a 1 cm^2 area on the right side and center of the leaf were counted with the aid of a magnifying glass.

k) Yield (Y): The data obtained (grams per plot) was transformed into kg/ ha^{-1}, this being the yield corrected for 13% moisture, according to the equation:

$$FW = IW \times \frac{100 - IM}{100 - FM} \quad \text{(Equation 1)}$$

where FW is the final corrected weight of the sample, IW is the initial weight of the sample, IM is the initial moisture of the sample, and FM is the final moisture of the sample (13%).

Data were subjected to analysis of variance (ANOVA). The genetic distance between all pairs of genotypes were estimated by Mahalanobis distance and Euclidean distance. Optimization clustering methods were applied, such as: method of the nearest neighbor or single bond; method of the furthest neighbor or complete connection; method of linking groups mean (UPGMA); Ward's minimum variance method, performed by the computer program Genes (CRUZ, 2013).

RESULTS AND DISCUSSION

The dissimilarities estimates provide useful information to breeders by quantifying and informing about the degree of resemblance or difference presented between two genotypes (CRUZ; FERREIRA; PESSONI, 2011). The dissimilarities measures estimated by Mahalanobis distance oscillated from 8,9 (BRSGO 7560 and UFUS 7401) to 335,7 (BRSGO 7560 and UFUS 27) indicating a large genetic diversity among the studied genotypes (Table 1). These dissimilarity results were lower than the ones obtained by Oliveira et al. (2014), who in a study with 22 soybean genotypes, also including lineages developed by UFU soybean breeding program, observed amplitudes from 0,00013 to 25,00 for Mahalanobis distance. The Euclidean distance oscillated from 0,43 (UFUS 15 and UFUS 36) to 2,91 (M-Soy 6101 and UFUS 139) (Table 2), being this amplitude higher than the one obtained by Torres et al. (2015), who in a study with 5 soybean genotypes, found an Euclidean distance oscillating between 0,96 to 2,91.

Table 1. Measures of the higher and lower genetic dissimilarity by the Mahalanobis distance in 15 soybean genotypes evaluated on agronomic traits in Uberlândia – MG.

Genotypes	Shortest distance	Longest distance
(G1) M-Soy 6101	64,8 (G11)	248,2 (G8)
(G2) BRSGO 7560	8,9 (G14)	335,7 (G6)
(G3) UFUS 15	15,8 (G7)	168,5 (G2)
(G4) UFUS 110	14,7 (G11)	176,6 (G6)
(G5) UFUS 24	15,0 (G13)	245,5 (G2)
(G6) UFUS 27	56,3 (G5)	335,7 (G2)
(G7) UFUS 26	9,3 (G13)	185,7 (G2)
(G8) UFUS 139	37,7 (G12)	270,3 (G10)
(G9) UFUS 54	12,9 (G7)	154,0 (G1)
(G10) UFUS Riqueza	41,0 (G5)	312,7 (G2)
(G11) UFUS 11	14,7 (G4)	156,3 (G6)
(G12) TMG 801	19,1 (G14)	225,1 (G6)
(G13) UFUS 36	9,3 (G7)	176,4 (G2)
(G14) UFUS 7401	9,0 (G2)	272,9 (G6)
(G15) UFUS 6901	27,5 (G4)	170,9 (G10)

Table 2. Measures of the lower and higher genetic dissimilarity by the Euclidean distance in 15 soybean genotypes evaluated on agronomic traits in Uberlândia – MG

Genotypes	Shortest distance	Longest distance
(G1) M-Soy 6101	1,70 (G11)	2,91 (G8)
(G2) BRSGO 7560	0,68 (G14)	2,61 (G1)
(G3) UFUS 15	0,43 (G13)	2,04 (G1)
(G4) UFUS 110	0,78 (G3)	2,28 (G1)
(G5) UFUS 24	0,48 (G13)	1,94 (G1)
(G6) UFUS 27	0,92 (G13)	2,27 (G1)
(G7) UFUS 26	0,71 (G5)	1,85(G1)
(G8) UFUS 139	1,0 (G9)	2,91 (G1)
(G9) UFUS 54	0,74 (G7)	2,31 (G1)
(G10) UFUS Riqueza	1,01 (G7)	2,12 (G2)
(G11) UFUS 11	0,57 (G3)	1,74 (G8)
(G12) TMG 801	0,67 (G14)	2,50 (G1)
(G13) UFUS 36	0,43 (G3)	2,09 (G1)
(G14) UFUS 7401	0,67 (G12)	2,37 (G1)
(G15) UFUS 6901	0,85 (G4)	2,03 (G1)

The Mahalanobis distance obtained for the cultivars BRSGO 7560 and UFUS 7401 (Table 1) were the smaller distance (8,9), because of that being considered the most similar. On the other hand, the BRSGO 7560 and UFUS 27 genotypes were considered the most divergent by presenting a bigger difference (335,7). Rigon et al. (2012), analyzed the genetic divergence in 18 soybean cultivars and obtained values between 0,008 to 0,53**.** Almeida et al. (2011), worked with 11 soybean cultivars and observed an elevated magnitude of the Mahalanobis distance which was from 2,65 to 374,06 indicating a high genetic variability according to what was found in this study.

By observing the values of the Euclidean distance of the UFUS 15 and UFUS 36 genotypes (Table 2), it was noticed that they showed the shortest distance (0,43), in this way, being considered as the most similar genotypes. On the other hand, the M-Soy 6101 and UFUS 139 cultivars were the most distinctive by showing a longer distance (2, 91).

In the obtained measures by the Mahalanobis distance (Table 1) there was a high frequency of pairs with the greatest distances when one of the components was the UFUS 27 lineage. On the other side, when the BRSGO 7560 and UFUS 7401 took part as one of the genotypes, the shortest distances were observed, meaning the smaller genetic dissimilarity.

In the measures obtained by the Euclidean distance (Table 2) it can be observed that the M-Soy 6101 genotype has a high dissimilarity with the most genotypes, except for the UFUS 26.

The cophenetic correlation is a coeficient that enable evaluate the adequancy betweeen the distance matrix and the dendrogram, allowing to increase the conclusions reliability in face of the dendrogram interpretation (Kopp et al., 2007). Besides that, with values of the correlation higher than 0,75 lower will be the distortion caused by the grouping (CRUZ et al., 2011). Cophenetic correlations higher than 0,75 can be considered high and indicate a good fitness of the distances showed in the dendrogram (MC GARIGAL et al., 2000).

In the method of the nearest neighbor, using the Mahalanobis distance matrix, was found a value of 0,53 in the cophenetic correlation (Table 3). Rigon et al. (2012), in a study with 18 soybean cultivars found a cophenetic correlation of 0,75 being, in this way higher than the one obtained in this study.

Table 3. Cophenetic correlation of different hierarchical methods in 15 soybean genotypes. *Meaningful to 1% of probability by t test.

Methods	Cophenetic Correlation **Mahalanobis distance**	Cophenetic Correlation **Euclidean distance**
Nearest neighbor	0,53*	0,82*
Furthest neighbor	0,58*	0,77*
UPGMA	0,67*	0,86*
Ward	0,62*	0,56*

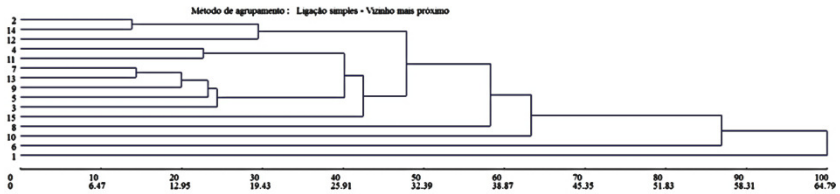
The cophenetic correlation estimates obtained in the method of the furthest neighbor were 0,58 and 0,77 respectively to the Mahalanobis and Euclidean distances (Table 3). These results were partially higher than the ones obtained by Cargnelutti Filho et al. (2010) in studies using beanstalk, as those authors found values from 0,62 and 0,67 for the Mahalanobis and Euclidiana distances respectively.

In the UPGMA method, by using the Mahalanobis distance, it was observed a cophenetic correlation estimate of 0,67. Sousa et al. (2015), in a study using 110 soybean genotypes, reported the value of 0,70. Bertan et al. (2007) considered 17 morphologic traits in soybean genetic diversity studies, perceived a cophenetic correlation coefficient of 0,80. Still in the UPGMA method, when the Euclidean distance is considered, the cophenetic correlation was 0,86. Gonçalves et al. (2014), in a study using 65 bean genotypes, found a cophenetic correlation of 0,81 which come close to the estimates reached in this study.

In the grouping method of Ward, the values found for the cophenetic correlations were 0,62 (Mahalanobis) and 0,56 (Euclidean). Cargnelutti et al. (2008), by the analysis of 14 bean cultivars, reported estimates which were higher, as they observed a correlation of 0,90 either for the Mahalanobis distance, or for the Euclidean distance.

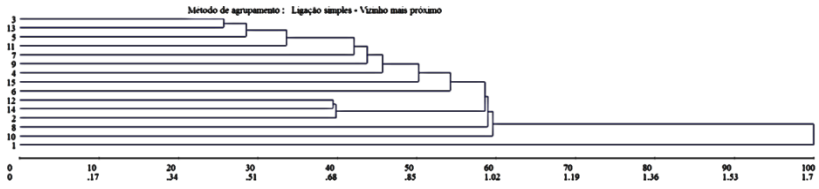
In the dendrogram generated from the Mahalanobis and the Euclidiana distances, the cuts were made in a level of 50% of dissimilarity, agreeing with Santos et al. (2013) and Torres et al. (2015), who used the same percentage of dissimilarity to observe the group formation in soybean genetic diversity study.

In the dendrogram of the nearest neighbor for the Mahalanobis distance (Picture A) there were the formation of 5 groups in which 73,3% of the genotypes were gathered in the first group (BRSGO 7560; UFUS 7401; TMG 801; UFUS 110; UFUS 11; UFUS 26; UFUS 36; UFUS 54; UFUS 24; UFUS 15; UFUS 6901). The remaining groups II, III, IV e V were formed by only one genotype in each (6, 7%) being them UFUS 139; UFUS Riqueza; UFUS 27 and M-Soy 6101 respectively. In the study of Almeida et al. (2011), with 12 soybean cultivars, if they performed a cut in 50%, the result would be like the one reported in this study, as most of the genotypes studied also would agglomerate in just one group (91,6%). Peluzio et al. (2012), in an analysis of 12 soybean genotypes, obtained two groups and which one of them were formed by all the genotypes with an exception of just one, which was spotted in the second group.



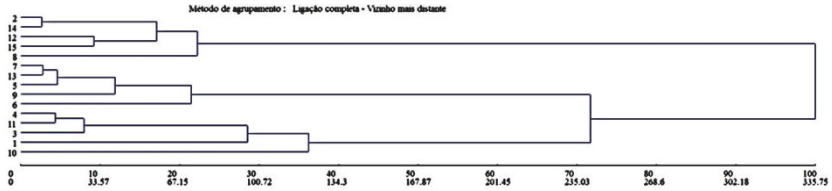
Picture A. Dendrogram of 15 soybean genotypes by the simple linkage method (the nearest neighbor) obtained from the Mahalanobis distance based in 14 agronomical characters 1- M-Soy 6101; 2 - BRSGO 7560; 3 - UFUS 15; 4 - UFUS 110; 5 - UFUS 24; 6 - UFUS 27; 7 - UFUS 26; 8 - UFUS 139; 9 - UFUS 54; 10 - UFUS Riqueza; 11 - UFUS 11; 12 - TMG 801; 13 - UFUS 36; 14 - UFUS 7401; 15 - UFUS 6901.

Considering the Euclidean distance to generate the nearest neighbor dendrogram (Picture B) was observed the formation of seven groups, with 46,7% of the genotypes joined in group I (UFUS 15; UFUS 36; UFUS 24; UFUS 11; UFUS 26; UFUS 54; UFUS 110). The groups II, III, V, VI and VII were composed by only one genotype (6, 7%) being them UFUS 6901, UFUS 27, UFUS 139, UFUS Riqueza and M-Soy 6101 respectively. However, the group IV was formed by 20% of the studied genotypes (TMG 801, UFUS 7401, and BRSGO 7560). Torres et al. (2015) highlighted in his study with six soybean genotypes, the development of three groups being one of them formed by 66,7% of the genotypes and the others formed by just 16,7%.



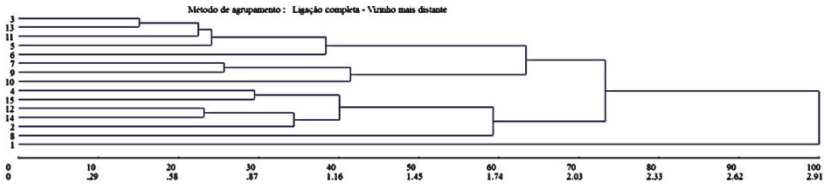
Picture B. Dendrogram of 15 soybean genotypes the simple linkage method (the nearest neighbor) obtained from the Euclidean distance based on 14 agronomical characters. 1- M-Soy 6101; 2 - BRSGO 7560; 3 - UFUS 15; 4 - UFUS 110; 5 - UFUS 24; 6 - UFUS 27; 7 - UFUS 26; 8 - UFUS 139; 9 - UFUS 54; 10 - UFUS Riqueza; 11 - UFUS 11; 12 - TMG 801; 13 - UFUS 36; 14 - UFUS 7401; 15 - UFUS 6901.

In the furthest neighbor method with the Mahalanobis distance (Picture C) three distinctive groups were formed being each one of them composed equally by 33,3% of the studied genotypes. The genotypes gathered in the first group were BRSGO 7560; UFUS 7401; TMG 801; UFUS 6901; UFUS 139. The second group formed gathered the genotypes: UFUS 26; UFUS 36; UFUS 24; UFUS 54; UFUS 27, and finally the third grouping was composed by the genotypes UFUS 110; UFUS 11; UFUS 15; M-Soy 6101; UFUS Riqueza.



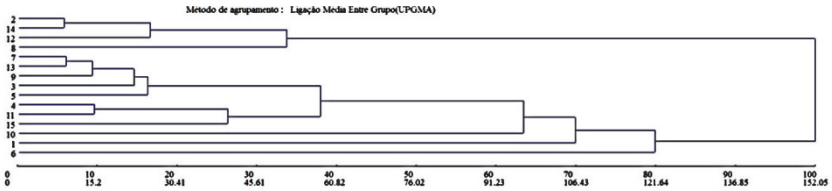
Picture C. Dendrogram of 15 soybean genotypes by the complete linkage method (the furthest neighbor) obtained from the Mahalanobis distance based on 14 agronomical characters. 1- M-Soy 6101; 2 - BRSGO 7560; 3 - UFUS 15; 4 - UFUS 110; 5 - UFUS 24; 6 - UFUS 27; 7 - UFUS 26; 8 - UFUS 139; 9 - UFUS 54; 10 - UFUS Riqueza; 11 - UFUS 11; 12 - TMG 801; 13 - UFUS 36; 14 - UFUS 7401; 15 - UFUS 6901.

Considering the Euclidean distance to obtain the dendrogram by the furthest neighbor method (Picture D), was possible to observe the constitution of five groups, being the first one of them composed by 33,3% of the genotypes (UFUS 15; UFUS 36; UFUS 11; UFUS 24; UFUS 27). The group II was composed by 20% of the genotypes, being them UFUS 26; UFUS 54; UFUS Riqueza. The group III gathered the genotypes, UFUS 110; UFUS 6901; TMG 801; UFUS 7401; BRSGO 7560, representing 33,3%. The groups IV and V condensed only one genotype in each one (6,7%), which were UFUS 139 and M-Soy 6101, respectively.



Picture D. Dendrogram of 15 soybean genotypes by the complete linkage method (the furthest neighbor) obtained from the Euclidean distance base on 14 agronomical characters. 1- M-Soy 6101; 2 - BRSGO 7560; 3 - UFUS 15; 4 - UFUS 110; 5 - UFUS 24; 6 - UFUS 27; 7 - UFUS 26; 8 - UFUS 139; 9 - UFUS 54; 10 - UFUS Riqueza; 11 - UFUS 11; 12 - TMG 801; 13 - UFUS 36; 14 - UFUS 7401; 15 - UFUS 6901.

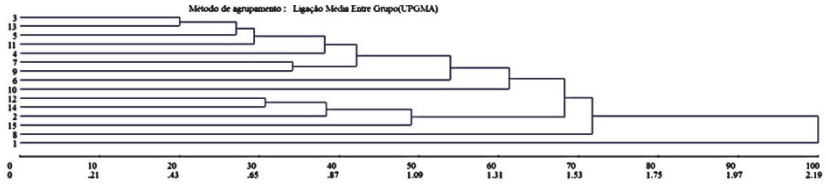
In the UPGMA method obtained with the Mahalanobis distance (Picture E) it was verified the formation of five groups being the first one of them formed by four genotypes (BRSGO 7560; UFUS 7401; TMG 801; UFUS 139), representing 26,7%. The second developed group gathered most of the genotypes (UFUS 26; UFUS 36; UFUS 54; UFUS 15; UFUS 24; UFUS 110; UFUS 11; UFUS 6901), representing, 53,3%. The three remaining groups (III, IV, V) were established by only one genotype in each (6, 7%), being them UFUS Riqueza, M-Soy 6101 and UFUS 27, respectively. Santos et al. (2013), in a study using 18 soybean genotypes, reached the formation of four distinctive groups being that approximately 50% of the studied genotypes gathered in a same group.



Picture E. Dendrogram of 15 soybean genotypes by the UPGMA method obtained from the Mahalanobis distance based in 14 agronomical characters. 1- M-Soy 6101; 2 - BRSGO 7560; 3 - UFUS 15; 4 - UFUS 110; 5 - UFUS 24; 6 - UFUS 27; 7 - UFUS 26; 8 - UFUS 139; 9 - UFUS 54; 10 - UFUS Riqueza; 11 - UFUS 11; 12 - TMG 801; 13 - UFUS 36; 14 - UFUS 7401; 15 - UFUS 6901.

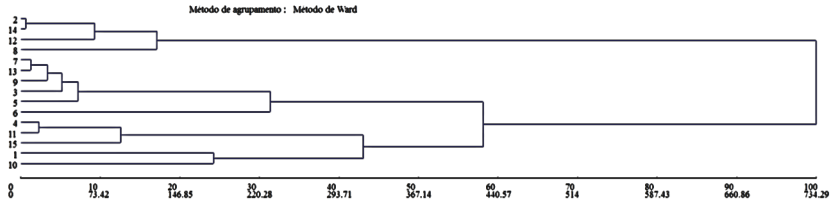
From the Euclidean distance adopting the UPGMA (Picture F) the genotypes were distributed in six groups. The first group was composed by 46,7% of the genotypes (UFUS 15; UFUS 36; UFUS 24; UFUS 11; UFUS 110; UFUS 26; UFUS 54). The fourth group represents 20% of the studied genotypes (TMG 801; UFUS 7401; BRSGO 7560, UFUS 6901). The other groups (II, III, V, VI) were

composed by one genotype, UFUS 27, UFUS Riqueza, UFUS 139, M-Soy 6101, respectively. Polizel; Juliatti; Juliatti (2010), when observing 111 soybean genotypes in a dendrogram with a cut of 50% of distance, got the formation of six groups being that 64% of the studied genotypes were grouped in the same group, corroborating with the results of this study.



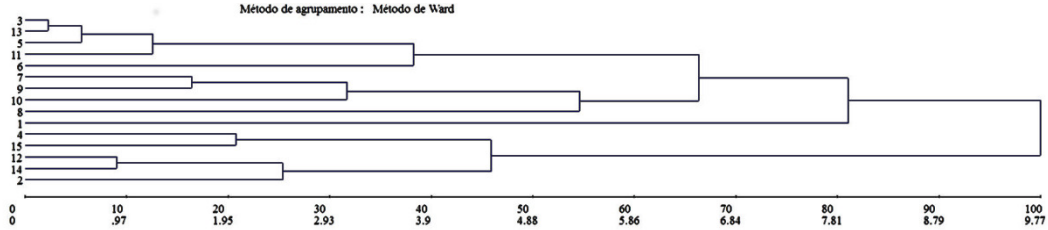
Picture F. Dendrogram of 15 soybean genotypes by the UPGMA method obtained from the Euclidean distance based in 14 agronomical characters. 1- M-Soy 6101; 2 - BRSGO 7560; 3 - UFUS 15; 4 - UFUS 110; 5 - UFUS 24; 6 - UFUS 27; 7 - UFUS 26; 8 - UFUS 139; 9 - UFUS 54; 10 - UFUS Riqueza; 11 - UFUS 11; 12 - TMG 801; 13 - UFUS 36; 14 - UFUS 7401; 15 - UFUS 6901.

In the Ward method with the Mahalanobis distance (Picture G) were formed three groups composed respectively by 26,7% (BRSGO 7560; UFUS 7401; TMG 801; UFUS 139), 40% (UFUS 26; UFUS 36; UFUS 54; UFUS 15; UFUS 24; UFUS 27) e 33,33% (UFUS 110; UFUS 11; UFUS 6901; M-Soy 6101; UFUS Riqueza) of the studied genotypes. Júnior et al. (2015), in a research with 41 soybean lineages obtained the formation of six groups, being the biggest one of them formed by 12 genotypes (29%).



Picture G. Dendrogram of 15 soybean genotypes by Ward method obtained from the Mahalanobis distance based in 14 agronomical characters. 1- M-Soy 6101; 2 - BRSGO 7560; 3 - UFUS 15; 4 - UFUS 110; 5 - UFUS 24; 6 - UFUS 27; 7 - UFUS 26; 8 - UFUS 139; 9 - UFUS 54; 10 - UFUS Riqueza; 11 - UFUS 11; 12 - TMG 801; 13 - UFUS 36; 14 - UFUS 7401; 15 - UFUS 6901.

Still in the Ward method with the Euclidean distance (Picture H) were verified five groups from which the first gathered 33,3% of cultivars (UFUS 15; UFUS 36; UFUS 24; UFUS 11; UFUS 27). The second group was composed by 20% of the genotypes (UFUS 26; UFUS 54; UFUS Riqueza). The groups III and IV were composed by the genotypes UFUS 139 and M-Soy 6101, respectively. Finally, the last group gathered 33,3% of the genotypes in the study (UFUS 110; UFUS 6901; TMG 801; UFUS 7401; BRSGO 7560). Arshad; Ali; Ghafoor (2006), in research using 33 soybean genotypes, got three groups, in which 42,4% of the genotypes joined in the same group, and the others go with 24,2% (group 2) and 33,3% (group 3).



Picture H. Dendrogram of 15 soybean genotypes by Ward method obtained from the Euclidean distance based in 14 agronomical characters. 1- M-Soy 6101; 2 - BRSGO 7560; 3 - UFUS 15; 4 - UFUS 110; 5 - UFUS 24; 6 - UFUS 27; 7 - UFUS 26; 8 - UFUS 139; 9 - UFUS 54; 10 - UFUS Riqueza; 11 - UFUS 11; 12 - TMG 801; 13 - UFUS 36; 14 - UFUS 7401; 15 - UFUS 6901.

When comparing the hierarchical methods from the Mahalanobis distance and the Euclidean distance it was possible to observe that there was a coincidence of grouping of some genotypes. In the complete linkage method (the furthest neighbor) and the UPGMA method, the coincidence among the genotypes which constituted the same group was equal or higher than 50%. In relation to the number of groups formed it was possible to find that the Euclidean distance showed a greater capacity of differentiation in a group in comparison to the Mahalanobis distance. Araújo et al. (2014) comparing the hierarchical methods with 11 cotton cultivars obtained that the nearest neighbor, the furthest neighbor, UPGMA and Ward grouping method also resulted in 90% or more of similarity among the grouped genotypes.

Independently of the dissimilarity measure, Mahalanobis or Euclidean, it was verified that the nearest neighbor method and UPGMA made possible the identification of a greater number of groups. Arriel et al. (2006) also commented a better differentiation of genotypes by the UPGMA method.

The BRSGO 7560 and TMG 801 genotypes remained grouped in all used methods. It is important to highlight that both cultivars have tolerance to soybean Asian rust (MT foundation, 2011; GLASSENAP et al., 2015; POLIZEL et al., 2010), and it can, consequently be potential parents in soybean breeding programs focused in resistance to *Phakopsora pachyrhizi*, because besides the tolerance to the fungi they also present productive potential (PASSOS et al., 2014).

The genotypes UFUS 139 and M-Soy 6101 remained isolated in distinctive groups in all analyzed groupings except for the furthest neighbor and Ward grouping methods of the Mahalanobis distance. This isolation shows that those cultivars can be potential parents, as emphasized by Arriel et al. (2006).

Although it has been showed in several studies that the soybean has a narrow genetic base (HIROMOTO; VELLO, 1986; PRIOLLI et al., 2004) it was possible to verify in this study the existence of a genetic diversity even among enhanced genotypes, agreeing in this way with Oda et al. (2015) and Glassenap et al. (2015), who still stated that there is substantial genetic variability in soybean.

In the view of the plant breeding, the data processing by several grouping methods and based in diverse dissimilarity measures, considering the particularities of each one, it was possible to determine the most divergent genotype as in this way useful in the breeding programs.

CONCLUSIONS

The generalized Mahalanobis distance or the Euclidean distance obtained with agronomic traits allow determining the genetic diversity in soybean. The use of Euclidean measure in hierarchical methods allows a greater differentiation of groups in soybean comparing with Mahalanobis distance.

The UPGMA method and the nearest neighbor showed a greater agreement in genotype grouping of soybeans using the Mahalanobis distance and the Euclidean distance.

ACKNOWLEDGMENTS

The authors thank CNPq for the financial support (postgraduate scholarship) and FAPEMIG (Foundation for Research Support of the State of Minas Gerais) for providing financial assistance.

RESUMO: Em estudos de diversidade genética de soja são utilizados caracteres agronômicos, morfológicos e moleculares que, por sua vez, são submetidos às análises biométricas multivariadas. Encontram-se disponíveis diferentes metodologias multivariadas, tais como as a distância Euclidiana, a distância de Mahalanobis e diferentes métodos hierárquicos. No entanto, são escassos os estudos que orientam para uma melhor escolha de tais análises em pesquisas com soja. O objetivo deste trabalho foi avaliar o padrão de agrupamento de genótipos de soja utilizando distância Euclidiana e Mahalanobis, seguindo diferentes métodos hierárquicos. O experimento foi realizado na Fazenda Capim Branco, da Universidade Federal de Uberlândia. Os tratamentos consistiram de 15 genótipos de soja (nove linhagens e seis cultivares) avaliados em delineamento de blocos completos casualizados com quatro repetições. Avaliaram-se os caracteres número de dias para o florescimento e maturidade, altura da planta no florescimento e na maturidade, altura de inserção da primeira vagem, número de nós na haste principal no florescimento e na maturidade, número de vagens com um, dois e três grãos, número total de vagens, produtividade de grãos, severidade da ferrugem asiática e número de pústulas. Os dados foram submetidos a análises multivariadas utilizando o Programa Genes. Tanto a distância generalizada de Mahalanobis como a distância Euclidiana, obtidas com caracteres agronômicos, permitem determinar a diversidade genética em soja. O uso da distância Euclidiana em métodos hierárquicos permite maior diferenciação de grupos. O

método UPGMA e o métodos do vizinho mais próximo apresentam maior concordância no agrupamento de genótipos utilizando a distância de Mahalanobis e a distância Euclidiana.

PALAVRAS-CHAVE: *Glycine max.* Variabilidade genética. Distância Euclidiana. Distância de Mahalanobis.

REFERENCES

ALMEIDA, R. D.; PELUZIO, J. M.; AFFÉRRI, F. S. Divergência genética entre cultivares de soja, sob condições de várzea irrigada, no sul do Estado Tocantins. Revista Ciência Agronômica, v. 42, n. 1, p. 108-115, 2011. https://doi.org/10.1590/S1806-66902011000100014

ARAÚJO, L. F.; ALMEIDA, W. S.; BERTINI, C. H. C. M.; VIDAL NETO, F. C. Utilização de diferentes métodos de agrupamentos na avaliação da divergência genética em algodoeiro herbáceo. Revista Ciência Agronômica, v. 45, n. 2, p. 312-318, 2014. https://doi.org/10.1590/S1806-66902014000200012

ARRIEL, N. H. C.; MAURO, A. O. D.; MAURO, S. M. Z. D.; BAKKE, A. O. Técnicas multivariadas na determinação da diversidade genética em gergelim usando marcadores RAPD. Pesquisa Agropecuária Brasileira, v. 41, n. 5, p. 801-809, 2006. https://doi.org/10.1590/S0100-204X2006000500012

ARSHAD, M.; ALI, N.; GHAFOOR, A. Character correlation and path coefficient in soybean *Glycine max (l.) Merrill*. Pakistan Journal of botany, v. 38, n. 1, p. 121-130, 2006.

BERTAN, I.; VIEIRA, E. A.; CARVALHO, I. F. F.; COSTA DE OLIVEIRA, A. Variabilidade genética em trigo aferida por meio da distância genealógica e morfológica. Scientia Agraria, v. 8, n. 1, p. 67-74, 2006. https://doi.org/10.5380/rsa.v8i1.8344

CARGNELUTTI FILHO, A.; RIBEIRO, D. N.; DOS REIS, R. C. P.; SOUZA, J. R; JOST, E. Comparação de métodos de agrupamento para o estudo da divergência genética em cultivares de feijão. Ciência Rural, Santa Maria, v. 38, n. 8, p. 2138-2145, 2008. https://doi.org/10.1590/S0103-84782008000800008

CARGNELUTTI FILHO, A.; RIBEIRO, N. D.; BURIN, C. Consistência do padrão de agrupamento de cultivares de feijão conforme medidas de dissimilaridade e métodos de agrupamento. Pesquisa Agropecuária Brasileira, Brasília, v. 45, n. 3, p. 236-243, 2010. https://doi.org/10.1590/S0100-204X2010000300002

CRUZ, C. D.; FERREIRA, F. M., PESSONI, L. A. Biometria aplicada ao estudo da diversidade genética. Viçosa: Suprema Gráfica e editora, 2011. 599 p.

CRUZ, C. D. Programa Genes: Aplicativo computacional em genética e estatística. Viçosa, UFV, 2013. 1 CD-ROM.

CRUZ, C. D.; CARNEIRO, P. C. S.; REGAZZI, A. J. Modelos biométricos aplicados ao melhoramento genético. 3. ed. Viçosa: UFV. 2014. 668 p.

ELIAS, H. T.; VIDIGAL, M. C. G.; GONELA, A; VOGT, G. A. Variabilidade genética em germoplasma tradicional de feijão-preto em Santa Catarina. Pesquisa agropecuária brasileira, Brasília, v. 42, n. 2, p. 1443-144, 2007.

FEHR, W. R; CAVINESS, C. E. Ames: Iowa State University of Science and Technology, 1977. 11p.

GLASENAPP, J. S.; SEDIYAMA, T.; CRUZ, C. D.; MATSUO, E.; BROMMONSCHENKEL, S. H.; OLIVEIRA, R. C. T; HAMAWAKI, O.T Diversidade de características agronômicas e moleculares em cultivares de soja com diferentes graus de resistência à *Phakopsora Pachyrhizi.* Bioscience Journal, v. 31, n. 1, p. 25-36, 2015. https://doi.org/10.14393/BJ-v31n1a2015-17166
https://doi.org/10.14393/BJ-v31n1a2015-171666

GODOY, C. V.; KOGA, L. J.; CANTERI, M. G. Diagrammatic scale for assessment of soybean rust severity. Fitopatologia Brasileira v. 31, p. 63-68, 2006. https://doi.org/10.1590/S0100-41582006000100011

GONÇALVES, D. L.; AMBROZIO, V. C.; BARELLI, M. A. A.; NEVES, L. G.; SOBRINHO, S. P.; LUZ, P. B.; SILVA, C. R. Divergência genética de acessos tradicionais de feijoeiros através de características da semente. Bioscience Journal, Uberlândia, v. 30, n. 6, p. 1671-1681, 2014.

HIROMOTO, D. M.; VELLO, N. A. The genetic base of Brazilian soybean *Glycine max* (L.) Merril) cultivars. Revista Brasileira de Genética v. 9, p. 295-306, 1986.

JÚNIOR, J. A.; UNÊDA-TREVISOLI, S. H.; ESPINDOLA, S. M. C. G.; VIANNA, V. F.; DI MAURO, A. O. Diversidade genética em linhagens avançadas de soja oriundas de cruzamentos biparentais, quádruplos e óctuplos. Revista Ciência Agronômica, v. 46, n. 2, p. 339-351, 2015.

KOPP, M. M.; SOUZA, V. Q.; COIMBRA, J. L. M.; LUZ, V. K.; MARINI, N.; OLIVEIRA, A. C. Melhoria da correlação cofenética pela exclusão de unidades experimentais na construção de dendogramas. Revista da FZVA. Uruguaiana, v. 14, n. 2, p. 46-53, 2007.

MATSUO, E.; SEDYAMA, T.; CRUZ, C. D.; OLIVEIRA, R. D. L.; OLIVEIRA, R. C. T.; NOGUEIRA, A. P. O Genetic diversity in soybean genotypes with resistance to Genetic diversity in soybean genotypes with resistance to *Heterodera glycines.* Crop Breeding and Applied Biotechnology, n. 11, p. 304-312, 2011. https://doi.org/10.1590/S1984-70332011000400003

MC GARIGAL, K.; CUSHMAN, S.; STAFFORD, S. Multivariate statistics for wildlife and ecology research. New York: Springer/Verlag, 2000. 283 p. https://doi.org/10.1007/978-1-4612-1288-1

MIRANDA, Z. F. S.; ARIAS, C. A. A.; PRETE, C. E. C.; KIIHL, R. A. S.; ALMEIDA, L. A; TOLEDO, J. F. F.; DESTRO, D. Caracterização genética de noventa cultivares elites de soja por meio do coeficiente de parentesco. Pesquisa agropecuária brasileira, Brasília, v. 42, n. 3, p. 363-369, 2007.

MT Foundation. Boletim de Pesquisa de Soja. Cuiabá: Fundação MT, 2011.

NOGUEIRA, A. P. O.; SEDIYAMA, T.; GOMES, J. D. Avanços no melhoramento genético da cultura da soja nas últimas décadas. In: LEMES, E.; CASTRO, L.; ASSIS, R. Doenças da soja melhoramento genético e técnicas de manejo. 1. ed. Millennium editora, cap. 11, 2015. p. 159-172.

ODA, M. C.; SEDIYAMA, T.; MATSUO, E.; CRUZ, C. D.; BARROS, E. G.; FERREIRA, M. F. S. Phenotypic and molecular traits diversity in soybean launched in forty years of genetic breeding. Agronomy Science and Biotechnology, v. 1, n. 1, p. 1 – 9, 2015.

OLIVEIRA, O. C.; LAZARINI, E.; TARSITANIO, M. A. A.; PINTO, C. C.; SÁ, M. E. Custo e lucratividade da produção de sementes de soja enriquecidas com molibdênio. Pesquisa Agropecuária Tropical, Goiânia, n.1, 2015.

OLIVEIRA, S. M.; SOUSA, L. B.; NOGUEIRA, A. P. O.; HAMAWAKI, O. T.; OLIVEIRA, V. M. Caracteres Agronômicos e Divergência Genética Entre Genótipos de Soja. Enciclopédia Biosfera – Centro Científico Conhecer, v. 10, n. 18, p. 759, 2014.

PASSOS, A. M. A.; GODINHO, A. L. M.; BROGIN, R. L.; AKER, A. M. Avaliação de cultivares de soja de ciclo precoce em área de pastagem na região sudoeste da Amazônia. Enciclopédia Biosfera – Centro Científico Conhecer, v. 10, n. 19, p. 319, 2014.

PELUZIO, J. M.; PIRES, L. P. M.; CANCELLIER, L. L.; AFFÉRI, F. S.; COLOMBO, G. A.; JÚNIOR, T. T.; RIBEIRO, G. R. S. Divergência genética entre cultivares de soja em várzea irrigada no Estado do Tocantins. Ciência Rural, v. 42, n. 3, 2012.

POLIZEL, A. C.; JULIATTI, F. C.; JULIATTI, F. C. Resistência parcial de genótipos de soja quanto a ferrugem asiática sob inoculação artificial. Enciclopédia biosfera, Centro Científico Conhecer - Goiânia, v. 6, n. 11, 2010.

PRIOLLI, R. H. G.; PINHEIRO, J. B.; ZUCCHI, M. I.; BAJAY, M. M.; VELLO, N.A. Genetic diversity among Brazilian soybean cultivars based on SSR loci and pedigree data. Brazilian Archives of Biology and Technology, v. 53, n. 3, p. 519-531, 2010. https://doi.org/10.1590/S1516-89132010000300004

PRIOLLI, R. H. G.; MENDES-JUNIOR, C. T.; SOUSA, S. M. B.; SOUSA, N. E. A.; CONTEL, E.P. B Diversidade genética da soja entre períodos e entre programas de melhoramento no Brasil. Pesquisa Agropecuária Brasileira, v. 39 n. 10, p. 967-975, 2004. https://doi.org/10.1590/S0100-204X2004001000004

RIGON, J. P. G.; CAPUANI, S.; BRITO NETO, J.F.; ROSA, G.M.; WASTOWSKI, A. D.; RIGON, C. A. G. Dissimilaridade genética e análise de trilha de cultivares de soja avaliada por meio de descritores quantitativos. Revista Ceres, Viçosa, v. 59, n. 2, p. 233-240, 2012.

SANTOS, E. R.; SANTOS, A. F.; CAPONE, A.; SANTOS, W. R. Dissimilaridade genética entre genótipos de soja cultivados em várzea irrigada no período de entressafra. Journal of Biotechnology and Biodiversity, v. 4, n.3, p. 222-231, 2013.

SINGH, R. K.; BHATIA, V. S.; BHAT, K. V.; MOHAPATRA, T.; SINGH, N. K.; BANSAL, K. C.; KOUNDAL, K. R. SSR and AFLP based genetic diversity of soybean germplasm differing in photoperiod sensitivity. Genetics and Molecular Biology, v. 33, n. 2, p. 319-324, 2010. https://doi.org/10.1590/S1415-47572010005000024

SOUZA L. B.; HAMAWAKI, O. T.; BATISTA, R. O.; BERTAN, I.; NOGUEIRA, A. P. O.; ROMANATO, F. N.; HAMAWAKI, R. L. Genetic variability among soybean biparental crosses evaluated by multivariate analysis. Bioscience Journal, Uberlândia, v. 31, n. 5, p. 1404-1412, 2015. https://doi.org/10.14393/BJ-v31n5a2015-26458

TORRES, F. E.; DAVID, G. V.; TEODORO, P. E.; RIBEIRO, L. P.; CORREA, C. G.; JÚNIOR R. A. L. Desempenho agronômico e dissimilaridade genética entre genótipos de soja. Revista de Ciências Agrárias, v. 38, n. 1, p. 111-117, 2015.

VAL, B. H. P.; JÚNIOR, J. A. F.; BIZARI, E. H.; DI MAURO, A. O.; UNÊDA-TREVISOLI, S.H. Diversidade genética de genótipos de soja por meio de caracteres agromorfológicos. Ciência & Tecnologia: FATEC-JB, Jaboticabal, v. 6, n. 1, p. 72-83, 2014.

YAMANAKA, N.; SATO, H.; YANG, Z.; XU, H.D.; CATELLI, L. L.; BINNECK, E.; ARIAS, C.A. A.; ABDELNOOR, R. V.; NEPOMUCENO, A. L. Genetic relationships between Chinese, Japanese, and Brazilian soybean gene pools revealed by simple sequence repeat (SSR) markers. Genetics and Molecular Biology, v. 30, n. 1, p. 85-88, 2007. https://doi.org/10.1590/S1415-47572007000100016