Full Length Research Paper

Identification of new descriptors for differentiation of soybean genotypes by Gower algorithm

Francisco Charles dos Santos Silva[1]*, Tuneo Sediyama[1], Amilton Ferreira da Silva[1], André Ricardo Gomes Bezerra[1], Daniele Piano Rosa[1], Lucas Vieira Ferreira[1] and Cosme Damião Cruz[2]

[1]Departament of Plant Science, Universidade Federal de Viçosa, Zip Code 36570-900, Minas Gerais, Brazil.
[2]Departament of Plant Biology, Universidade Federal de Viçosa, Zip Code 36570-900, Minas Gerais, Brazil.

The descriptors used to differentiate new soybean cultivars which ensure their protection, has become insufficient. This makes evident the need to expand the list of that already used. The morphological traits that may be used as descriptors can be quantitative or qualitative; however, in the evaluation of their efficiency, they are not always analyzed simultaneously. Thus, this study aimed to identify new morphological characteristics for purposes of discrimination of soybean cultivars with the aid of Gower algorithm, which allows evaluation of quantitative and qualitative characters simultaneously. Ten soybean cultivars were used in two experiments, one in the summer and another in the winter. They were conducted in a greenhouse. The experiments were carried out in an experimental randomized complete block design with four replications. Each experimental unit consisted of one pot with two plants. Five quantitative and seventeen qualitative characteristics were evaluated and the discrimination analysis was performed by Gower algorithm. It was concluded that there are differences among cultivars for all traits analyzed in both sowing dates, except for the opening of the petioles of the unifoliolate leaves. The use of Gower algorithm was efficient in the evaluation of discriminatory capacity of quantitative and qualitative characteristics simultaneously, demonstrating that they are useful as soybean descriptors.

Key words: *Glycine max* (L.), law on plant variety protection, productivity, qualitative variables.

INTRODUCTION

The expansion and establishment of agricultural frontiers was only possible due to the development of cultivars with high yield, broad adaptation to different environmental conditions, and resistant to pests and diseases (EMBRAPA, 2011). In Brazil, the protection of intellectual property rights on the varieties is performed

by obtaining the protection certificate. To get this, it is necessary to prove the new plant variety is distinct, uniform, and stable. The distinction is made by a minimum margin of descriptors, specific to each species (Neto et al., 2005).

Currently, 38 descriptors are used among the mandatory and additional to prove the distinction of soybean cultivars required for protection. However, these are still insufficient, mainly due to the narrow genetic base of this specie in Brazil (Boldt et al., 2007; Nogueira et al., 2008). There is therefore a clear need for studies aimed at identifying new descriptors for soybeans in order to expand the list of already used in the differentiation of soybean cultivars.

There are reports relating to effects of genetic differences for the hypocotyl length (Knittle and Burris, 1979; Nóbrega and Vieira, 1995). Costa et al. (1999) also reported that the variation in hypocotyl length is assigned to genetic differences among cultivars. In order to separate seeds that have the same characters, Dorchester (1945) used as auxiliary trait, the base shape of unifoliate leaf (acuminate, straight and auriculate), concluding that the shape of the leaf has well-defined characteristics, although there is variation within the same cultivar.

Nogueira et al. (2008) identified the hypocotyl and epicotyl length as being useful in soybean genotypes distinction. To evaluate the significance of a variable to descriptor, several statistical methods can be used. However, most of these take into account only quantitative variables, although the joint analysis of quantitative and qualitative variables potentially raises the chances of new varieties distinction. A technique that allows the simultaneous analysis of quantitative and qualitative data was proposed by Gower (1971) by means of an algorithm that estimates the similarity between two individuals using continuous data and discrete distributions.

The technique proposed by Gower (1971) allows that values of distance matrix stood between 0 and 1. Having the distance matrix is performed in a single cluster analysis for the different variables. This grouping has in principle to join the accesses in groups, so that there is uniformity within and heterogeneous between groups (Gonçalves et al., 2009).

Although this analysis have been proposed more than 40 years ago. Only in recent years, it has been the most frequently used to quantify the genetic dissimilarity in *Brassica napus* L. (Rodriguez et al., 2005), *Triticum aestivum* L. (Vieira et al., 2007), *Solanum lycopersicum* (Gonçalves et al., 2008), and *Capsicum* spp. (Moura et al., 2010) for example. However, their use in the identification of morphological descriptors was not reported. This may be due to lack of knowledge of statistical techniques that allow this approach and the lack of free computer programs that can analyze such a procedure (Rocha et al., 2010).

Thus, this study aimed to identify and evaluate new morphological traits for purposes of differentiation of soybean cultivars using Gower algorithm.

MATERIALS AND METHODS

The experiment was carried out in soybean breeding program, in greenhouse in the Department of Plant Science at the Federal University of Viçosa, Minas Gerais, Brazil. Two experiments were performed, one in the summer (December) and another in the winter (May).

To express a possible influence of environmental variations in the discrimination capacity of the genotypes through of the evaluated characteristics, the control of temperature or brightness were not performed, since by stability criterion, the discrimination ability of a descriptor can not be influenced by environmental variations.

The treatments consisted of 10 soybean cultivars: BRS Carnaúba, BRS Candeia, BRS 278 RR, BRS 271 RR, BRS Tracajá, UFVTN 105 AP, TMG 401, BRSMG 68 (Vencedora), TMG 801, and FMT Tucunaré, which were arranged in a randomized block design with four replications and each experimental unit consists of a pot with two plants.

In total, 22 characteristics were evaluated, which were not used as soybean descriptors, being five qualitative [base shape of unifoliate leaf (BSUL), phyllotaxy of the first pair of trifoliate leaves (PFTL), depression in the pod (DV), hypocotyl color (HC), and size of pod apiculus (SPA)] and seventeen quantitative traits [hypocotyl length (HL), plant height (PH), epicotyl length (EL), length of the first internode (LFI), petiole length of the first trifoliate leaf (PLFTL), petiole length of the unifoliate leaf (PLUL), rachis length of the first trifoliate leaf (RLFTL), angle formed by the insertion of petiole of the unifoliate leaf (APUL), opening angle of the petioles of the unifoliolate leaves (OAPUL), stem diameter (SD), being these measured only once when the plants were in stage V3 (Fehr and Caviness, 1977).

The measurements made from pods were: length (PL), diameter (PD), thickness (PT), thickness of bandaging (TBP), and curvature (PC), which are measured in ten pods of three seeds from each experimental unit. Finally, the width (HW) and the length of the heel (HL) obtained in a seed pod each were evaluated. With the information collected, an analysis of variance was performed to verify the existence of genetic variability among the materials evaluated for the traits studied.

The ability of joint discrimination of traits that have proven useful in distinguishing cultivars was estimated by genetic distance matrix obtained by Gower algorithm (1971), expressed as:

$$S_{ijk} = \frac{\sum_{k=1}^{p} W_{ijk} \cdot S_{ijk}}{\sum_{k=1}^{p} W_{ijk}}$$

where k is the number of variables (k = 1, 2, ..., p); i and j are two individuals representing the access; W_{ijk} is the weight given to ijk comparison, assigning value 1 for valid comparisons and 0 for invalid comparisons (when the variable value is missing in one or both individuals); S_{ijk} is the variable contribution k in the similarity between i and j individuals, with values between 0 and 1. For a qualitative variable (nominal), if the value of the variable k is the same for both individuals i and j, then $S_{ijk} = 1$, otherwise it is equal to 0; for a quantitative variable (continuous) $S_{ijk} = 1 - |x_{ik} - x_{jk}| / R_k$ where x_{ik} and x_{jk} are values to variable k for individuals i and j, respectively, and R_k is the range (maximum minus minimum), the

variable k in the sample. The division by Rk eliminates the differences between the variables scales, producing a value within the range [0, 1] and equal weights.

The graphical representation of genetic distances was obtained by Unweighted Pair-Group Method Using an Arithmetic Average (UPGMA) method, the cut-off point in dendrogram was obtained by Mojema method. The data were analyzed by the program genes: Biometrics (Cruz, 2013).

RESULTS AND DISCUSSION

A necessary condition for a trait to be useful in differentiation of the cultivars is the existence of genetic variability. For most quantitative variables, there were significant differences by the F test, except for the AAFU characteristic (Table 1). In addition to distinctness, stability can be observed by the uniformity of variability in two sowing dates.

Genetic variability among soybean cultivars for the characteristics hypocotyl length, epicotyl length, petiole length of unifoliate leaf, petiole length of the first trifoliate leaf, rachis length of terminal leaflet in the first trifoliate leaf and shape of unifoliate leaf had reported itself by Nogueira et al. (2008), demonstrating that they can be useful as additional soybean descriptors.

Table 1 also presented the genotypic determination coefficient (H^2) for quantitative traits evaluated in two sowing dates. Large genetic influence was found for most traits at different times, with values ranging between 0.716 and 0.939, showing little environmental effect, however, the variables OAPUL, APUL, and HL showed minimum values of 4.5, 55.8, and 64.9%, respectively, showing significant environmental effect.

The coefficients of variation of all variables showed values of low to high magnitude, ranging from 2.94% to PT trait in the experiment conducted in the summer up 24.04% for RLFTL also in the summer (Table 1). High values of coefficient of variation were also obtained by Nogueira et al. (2008) for RLFTL. The high CV values can be associated with non-homogenization of traits during the cultivars development process.

In Figures 1, 2 and 3, the discrimination capacity of the traits are measured by the amount of cultivars that are isolated below the cutting line, so that the more genotypes are isolated, the greater the possibility of using characteristics as descriptors.

Results for discriminatory ability of the traits, quantitative, and qualitative, considered together are as shown in Figure 1, in which the groups formed by the cultivars in summer and winter conditions are demonstrated.

Excluding the trait OAPUL, which did not show variability among genotypes analyzed, the remaining 21 morphological traits did not allow a complete distinction of all varieties. However, in the summer, sowing was possible to differentiate the varieties to each other, except for the BRS 271 RR, BRS, and BRS Tracajá

Carnauba which formed a distinct group of the others and BRS 278 RR and BRS MG 68 (Vencedora), which were also together in one group (Figure 1A).

In winter sowing, the traits provided a distinction of 60% of the studied varieties. However, it is not enough to make differentiation of BRS Carnaúba of the BRS 271 RR and of BRS Candeia of the BRS 278 RR (Figure 1B). The efficiency of the use of quantitative and qualitative morphological characteristics together for discriminating cultivars by Gower algorithm becomes clear when compared with their effectiveness in distinguishing materials with the results which were obtained when they were used only quantitative or qualitative variables.

In summer sowing, the use of variables together provided a distinction of 50% of all cultivars (Figure 1A), while only the use of quantitative variables (Figure 2A) did not provide the differentiation of any cultivar individually, forming only two major groups. In turn, the use of only qualitative variables (Figure 3A) allowed the differentiation of 20% of the evaluated materials.

In winter, the simultaneous use of quantitative and qualitative variables resulted in the differentiation of 60% of the cultivars (Figure 1B), while when none of the quality characteristics (Figure 2B) was being considered, there was no distinction of individual cultivar, with only the formation of three groups of similar cultivars to each other. The use of qualitative variables only (Figure 3B), allowed the distinction of 40% of the cultivars.

These results indicate the possibility of using the Gower algorithm in soybean genotypes distinction when quantitative and qualitative traits are evaluated together, resulting in a more efficient discrimination since a greater number of descriptors will be considered.

It should be noted that cultivars that were not distinguished in the discrimination analysis of the two sowing periods (Figure 1) does not imply the futility of variables to differentiate cultivars, because the number of adopted traits was limited. Thus, by including traits already used as descriptors with the new identified, can improve the efficiency of the discriminatory capacity.

Since it was not used in this analysis, none of the 38 descriptors already used in soybean analysis, differentiation of 60% of 10 cultivars strengthens the possibility of using traits as descriptors. Since the analysis of a larger quantity of cultivars possibly would result in higher levels of discrimination. With regard to lack of consistency in the classification of cultivars between the different sowing dates, a likely explanation is the lack of homogeneity of cultivars for the traits evaluated that have not yet been considered in the process of selection and breeding of the species. The same effect was observed by Nogueira et al. (2008) who studied the efficiency of discriminant analysis in the differentiation of soybean cultivars for new descriptors.

According to Vieira et al. (2009), the morphological traits evaluated at the seedling stage and that evaluated

Table 1. Mean square of the genotypes (MSG), coefficients of variation (CV%) and genotypic determination coefficient (H2) of quantitative characteristics; hypocotyl length (HC), plant height (PH), epicotyl length (EL), length of the first internode (LFI), petiole length of the unifoliate leaf (PLUL) petiole length of the first trifoliate leaf (PLTL), rachis length of the first trifoliate leaf (RLFTL) angle formed by the insertion of petiole of the unifoliate leaf (APUL), opening angle of the petioles of the unifoliolate leaves (OAPUL), stem diameter (SD), pod length (PL), pod diameter (PD), pod thickness (TP), thickness of the pod bandaging (TPB), pod curvature (PC), heel width (HW) and the heel length (HL) in soybean cultivars in the summer sowing (December) and winter (May) in Viçosa-MG.

Character	Summer					Winter				
	QMG		H^2	Average	CV%	QMG		H^2	Average	CV%
HC	4.807	**	86.5	3.506	11.845	6.161	**	88.5	3.564	11.761
PH	339.215	**	91.6	25.691	10.455	166.657	**	83.2	25.691	11.217
EL	7.103	**	85.5	5.130	14.122	14.126	**	89.4	4.882	12.917
LFI	9.439	**	79.6	5.117	14.199	7.702	**	76.2	4.693	15.458
PLUL	268.396	**	92.3	13.768	17.668	320.244	**	94.5	14.503	14.276
PLTL	7.450	**	78.4	8.216	8.211	12.325	**	81.8	7.7752	10.188
RLFTL	59.330	**	82.3	7.790	21.798	42.423	**	74.2	7.125	24.041
APUL	1184.62	**	79.0	67.075	13.064	533.149	**	55.8	52.843	15.200
OAPUL	318.129		4.5	121.58	17.199	316.718		04.9	144.13	9.299
SD	7.969	**	77.0	6.830	11.035	11.398	**	86.3	5.806	11.669
PL	1.002	**	84.4	4.889	4.360	1.151	**	91.9	5.074	3.077
PD	3.254	**	77.5	6.532	6.886	3.446	**	87.1	6.647	5.112
TP	2.324	**	79.2	3.984	9.475	2.257	**	85.9	9.737	2.944
TPB	1.900	**	72.3	7.695	5.009	2.937	**	90.1	7.825	3.497
HW	0.435	**	82.4	1.549	9.244	0.269	**	64.9	1.481	10.736
HL	0.438	**	71.6	3.553	6.008	1.505	**	83.1	3.576	7.412
PC	23.564	**	89.5	3.973	20.992	29.808	**	93.9	3.997	16.938

**,*Significant at 1 and 5% probability by F test, respectively.
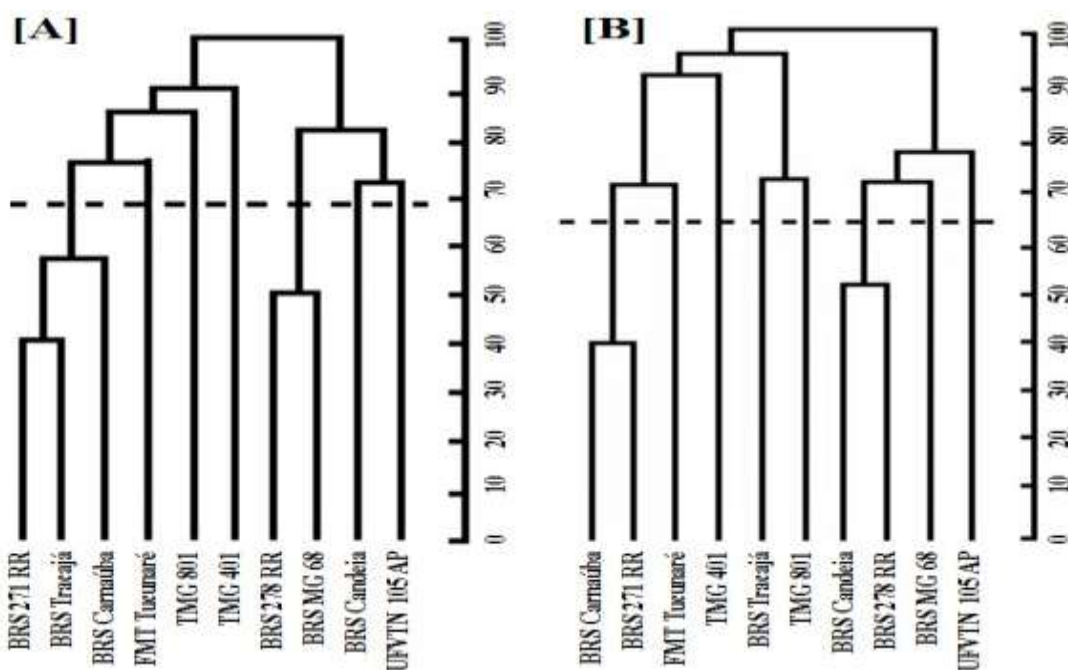

Figure 1. Dendogram of the similarity pattern between 10 soybean cultivars based on qualitative and quantitative traits, obtained by the UPGMA method through Gower algorithm in the summer sowing [A] and Winter [B].
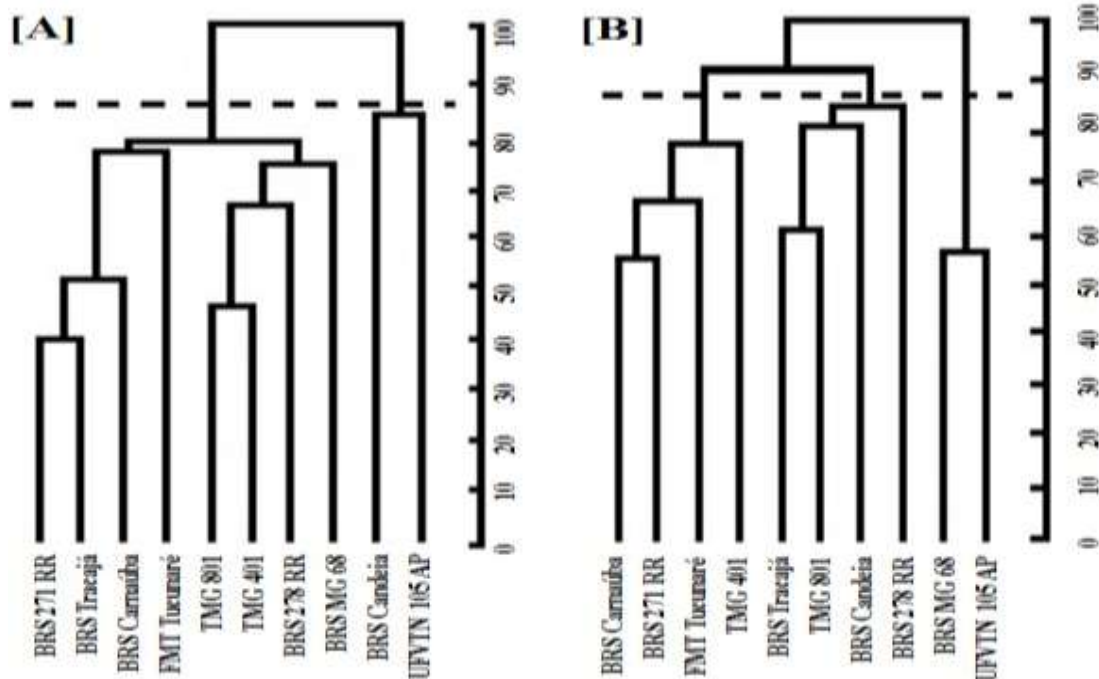
Figure 2. Dendogram of the similarity pattern between 10 soybean cultivars based only on quantitative characteristics, obtained by the UPGMA method through Gower algorithm in the summer sowing [A] and Winter [B].
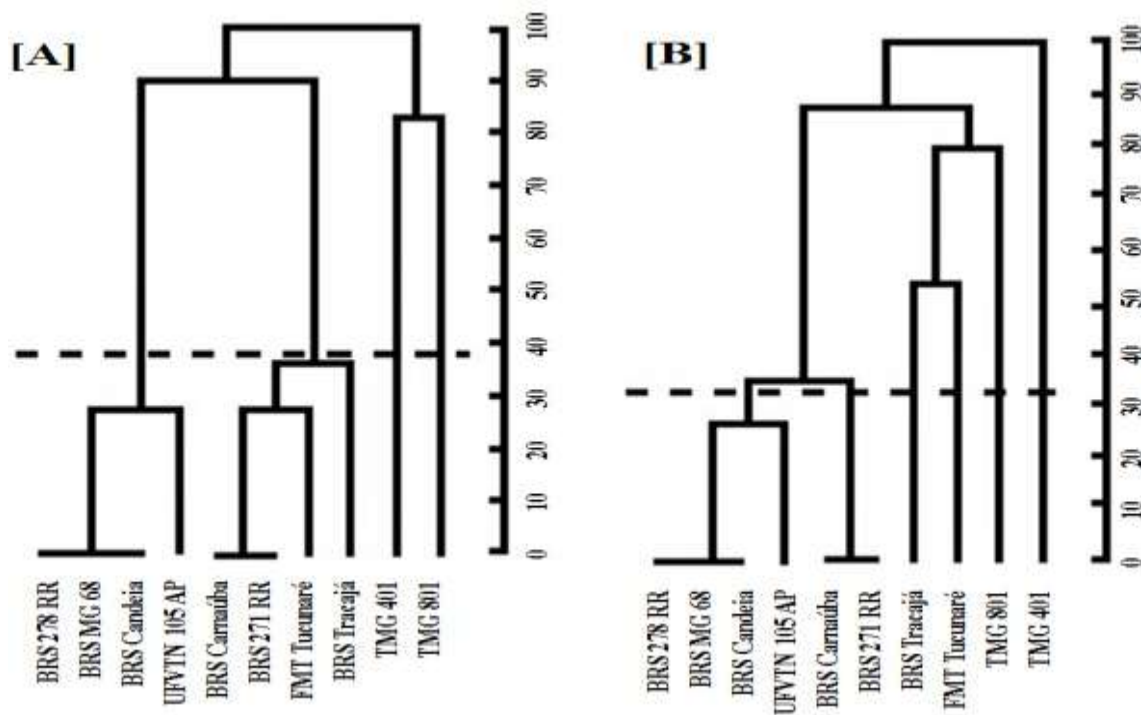

Figure 3. Dendogram of the similarity pattern between 10 soybean cultivars based on the qualitative characteristics obtained by the UPGMA method through Gower algorithm in summer seeding [A] and Winter [B].

in the seeds should be preferred, as they allow the rapid achievement of a result, thus, there is no need to wait for adult plants. The BRS Candeia, BRS Tracajá, UFVTN 105 AP, TMG 401, BRSMG 68, GMT 801 and FMT Tucunaré showed morphological traits in seeds and seedlings that were helpful to their separation from other cultivars.

Some qualitative morphological descriptors were specific to a small number of varieties, as the opposite phyllotaxy of the first couple of trifoliate leaves present only in cultivar TMG 801 and FMT Tucunaré; auriculate form of the base of unifoliate leaf in TMG 801 and 401 cultivars; pod depression, which was visible only in the cultivar TMG 401 RR and pod prominent apiculus in UFVTN 105 AP. The quantitative descriptors in turn were also able to differentiate the genotypes as it provided a refinement distinction in two sowing dates (Figures 2 and 3).

Vieira et al. (2009), Nogueira et al. (2008), and Boldt et al. (2007) also suggested new morphological descriptors based on the evaluation of seedlings and pod, for soybeans, since the 38 descriptors recommended by the SNPC have not been sufficient.

The assessed morphological descriptors allowed the characterization and distinction of most cultivars, being useful for the distinctness, uniformity and stability (DUS testing).

Conclusions

The evaluated soybean cultivars presented genetic variability to the characteristics basis shape of unifoliate leaf, phyllotaxy of the first pair of trifoliate leaves, pod depression, hypocotyl length, plant height, epicotyl length, length of the first internode, length of the petiole of first trifoliate leaf, length of the petiole of the leaf unifoliate, length of the rachis of the first trifoliate leaf, angle formed by the insertion of petioles of unifoliate leaf, pod length, pod diameter, pod thickness, thickness of bandaging pod, pod curvature, heel width end, and the hilum length .

The use of Gower algorithm was efficient in evaluation of the discriminatory capacity of quantitative and qualitative characteristics simultaneously in soybean.

Conflict of Interests

The authors have not declared any conflict of interests.

REFERENCES

Boldt AS, Sediyama T, Nogueira APO, Matsuo E, Teixeira RC (2007). Influência do tamanho de semente na caracterização de descritores adicionais de soja. In: reunião de pesquisa de soja da região central do brasil, Campo Grande. Anais. Londrina: Embrapa Soja 1:120-122.

Costa JA, Pires, JLF, Thomas AL, Alberton M (1999). Comprimento e índice de expansão radial do hipocótilo de cultivares de soja. Ciência Rural Santa Maria 29(4):609-612.

Cruz CD (2013). GENES - a software package for analysis in experimental statistics and quantitative genetics. Acta Sci. 35(3):271-276.

Dorchester CS (1945). Seed and seedling characters in certain varieties of soybeans. J. Am. Soc. Agron. Madison 37(1):223-232.

EMBRAPA (2011). Sistema de produção 15: tecnologias de produção de soja – região central do Brasil 212 e 2013. Londrina: EMBRAPA SOJA. 261 pp.

Gonçalves LSA, Rodrigues R, Amaral AT, Karasawa M, Sudré CP (2008). Comparison of multivariate statistical algorithms to cluster tomato heirloom accessions. Genet. Mol. Res. 7(4):1289-1297.

Gonçalves LSA, Rodrigues R, Amaral AT, Karasawa M, Sudré CP (2009). Heirloom tomato gene bank: assessing genetic divergence based on morphological, agronomic and molecular data using a Ward-modified location model. Genet. Mol. Res. 8(1):364-374.

Gower JC (1971). A general coefficient of similarity and some of its properties. Biometrics 27(2):857-874.

Knittle KH, Burris JS (1979). Soybean hypocotyl growth under field conditions. Crop Sci.. Madison. 19(1):37-41.

Moura MCCL, Gonçalves LSA, Sudré CP, Rodrigues R, Amaral AT, Pereira TNS (2010). Algoritmo de Gower na estimativa da divergência genética em germoplasma de pimenta. Hortic. Bras. 28(2):155-161.

Neto MOV, Borém A, Portugal RS (2005). Leis de proteção de genótipos. In: Borém, A. (Ed.). Melhoramento de Espécies Cultivadas. Viçosa: UFV. pp. 931-960.

Nóbrega LHP, Vieira RD (1995). Avaliação e classificação de cultivares de soja quanto ao comprimento do hipocótilo, sob condições de laboratório e de casa de vegetação. Revista Brasileira de Sementes, Brasília 17(2):160-164.

Nogueira APO, Sediyama T, Cruz CD, Reis MS, Pereira DG, Jangarelli M (2008). Novas características para diferenciação de cultivares de soja pela análise discriminante. Ciência Rural, Santa Maria. 38(9):2427-2433.

Rodríguez VM, Cartea ME, Padilla G, Velasco P, Ordás A (2005). The nabicol: a horticultural crop in northwestern Spain. Euphytica 142(3):237-246.

Vieira EA, Carvalho FIF, Bertan I, Kopp MM, Zimmer PD, Benin G, Silva JAG, Hartwig I, Malone G, Oliveira AC (2007). Association between genetic distances in wheat (*Triticum aestivum* L.) as estimated by AFLP and morphological markers. Genet. Mol. Biol. 30(2):392-399.

Vieira ESN, Pinho EVRV, Carvalho MGG, Silva PA (2009). Caracterização de cultivares de soja por descritores morfológicos e marcadores bioquímicos de proteínas e isoenzimas. Revista Brasileira de Sementes 31(1):086-094.