

# Morphological descriptors and ISSR molecular markers in the evaluation of genetic variability of *Tectona grandis* genotypes

A.M. Chimello<sup>1</sup>, J.G. Jesus<sup>1</sup>, P.E. Teodoro<sup>2</sup>, A.A.B. Rossi<sup>1</sup>, K.L. Araújo<sup>1</sup>, T.N. Marostega<sup>1</sup>, L.G. Neves<sup>1</sup> and M.A.A. Barelli<sup>1</sup>

<sup>1</sup>Laboratório de Melhoramento de Plantas, Departamento de Agronomia, Universidade do Estado de Mato Grosso, Cáceres, MT, Brasil

<sup>2</sup>Departamento de Agronomia, Universidade Federal do Mato Grosso do Sul, Chapadão do Sul, MS, Brasil

Corresponding author: P.E. Teodoro

E-mail: eduteodoro@hotmail.com

Genet. Mol. Res. 16 (2): gmr16029665

Received March 9, 2017

Accepted April 17, 2017

Published May 25, 2017

DOI <http://dx.doi.org/10.4238/gmr16029665>

Copyright © 2017 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

**ABSTRACT.** This study aimed to evaluate the genetic variability of the teak germplasm bank, using morphological traits and inter-simple sequence repeat molecular markers. Thirty clones were evaluated in a randomized complete block design with three replicates, and each plot was composed of three plants. A joint analysis of quantitative and qualitative variables was performed using the Gower algorithm. Quantitative, qualitative, and molecular variables were analyzed simultaneously using the Ward-MLM procedure. There is genetic variability among the 30 teak genotypes studied, considering the quantitative, qualitative, and molecular variables by the Ward-MLM statistical procedure. Morphological traits used proved to be efficient for the study of genetic variability; however, it was not possible to compose a descriptor table for clonal teak genotypes based on the traits evaluated. The Gower method was efficient in discriminating

the groups, demonstrating that the simultaneous analysis of qualitative and quantitative data is feasible and can allow greater efficiency in the knowledge of the variability among teak genotypes. The genotype 22 showed to be the most divergent compared to the other genotypes, except for the cluster of genotypes by the UPGMA method based on the Gower distance obtained by the Ward-MLM procedure, which formed a group with genotypes 9 and 30, in the morphological and molecular analyses and was grouped alone.

**Key words:** Joint analysis; Gower algorithm; Polymorphic index content

## INTRODUCTION

Wood is one of the most used materials by man and with increasing demand for raw materials for the various sectors, so fast growing species play an important role in the supply of wood, due to the high rates of deforestation at world level (Food and Agriculture Organization of the United Nations, 2009). One species that has stood out in this sector is the teak (*Tectona grandis* L.f), which is one of the most cultivated in the world, and is characterized by being one of the most valuable woods, due to its excellent physical properties and diverse uses such as shipbuilding, furniture manufacturing, among others (Cáceres Florestal S/A, 1997).

According to Cruz and Carneiro (2006), studying genetic variability is an activity of great relevance for the improvement of plants and the conservation of many species. Through this knowledge, it is possible to identify different genotypes, with characteristics of interest, as sources of disease resistance and high productivity and use them in breeding programs. Genetic variability can be verified by the use of several types of descriptors, such as morphological, agronomic, biochemical, and molecular descriptors.

The use of multivariate techniques is one of the factors that has driven the increase in the studies on genetic divergence between accesses of active germplasm banks, where we can highlight the agglomerative methods that have as principle to gather the genotypes into groups, in such a way that there is homogeneity within these and heterogeneity among them. This methodology depends on the calculation of dissimilarity measures from quantitative and qualitative variables (Crossa and Franco, 2004).

A technique that allows the simultaneous analysis of quantitative and qualitative data was proposed by Gower (1971). This method allows values of the distance matrix to be between 0 and 1, being necessary the standardization of the quantitative and qualitative variables. Several studies were carried out with *Brassica napus* L. (Rodríguez et al., 2005), *Triticum aestivum* L. (Vieira et al., 2007), and *Solanum lycopersicum* (Gonçalves et al., 2008) using the Gower algorithm. Given the above, this study aimed to evaluate the genetic variability of the teak germplasm bank through morphological and molecular inter-simple sequence repeat (ISSR) markers.

## MATERIAL AND METHODS

This research was developed at the University Campus of Cáceres of the University of the State of Mato Grosso - UNEMAT (16°11'42"S and 57°40'51"W). The evaluated genotypes are selected clones, coming from PROTECA Biotecnologia Florestal Company

(Table 1). Experimental design was a randomized block with three replicates, and each plot was composed of three plants; 30 different selected clonal teak genotypes were evaluated in a greenhouse (Table 1).

**Table 1.** Identification of the 30 teak genotypes selected from PROTECA Biotecnologia Florestal Company.

Treatment	Identification	Origin	Treatment	Identification	Origin	
1	A2	Solomon Islands	16	B17	Brazil	
2	A3		17	C1	Malaysia	
3	A4		18	C2		
4	A6		19	C5		
5	A7		20	D2	India	
6	A8		21	D3		
7	A9		22	D4		
8	A10		23	E2	Indonesia	
9	A11		24	E4		
10	A12		25	G1	Ivory Coast	
11	B2		Brazil	26	G2	
12	B6			27	J1	Ghana
13	B8	28		J2		
14	B14	29		T4	Tanzania	
15	B16	30		T8		

Seedlings were initially planted with a height of 10 cm in 2-L pots containing a mixture of soil and sand at a ratio 3:1. Fertilization was performed once a week using 100 mL nutrient solution (Clark, 1975). Irrigation was performed manually with nebulization of 3 s every 10 min.

### Morphological descriptors

To determine the morphological characteristics of teak seedlings, at 120 days after transplanting, samples composed of young, developing and developed leaves were collected, totaling 6 leaves from each plant: two from the upper third, two from the middle third, and two from the lower third. The evaluations were separated into quantitative and qualitative data. Quantitative data were analyzed by plant height, stem diameter, fresh and dry mass of the aerial part, fresh and dry mass of the root, length and width of the leaf blade, and abaxial and adaxial hairiness. For characterizing the morphological expression, we used the studies carried out by Vidal and Vidal (2003) and the International Union for Protection of New Varieties of Plants (UPOV, 2010).

Qualitative data were analyzed by the shape of the leaf blade, the shape of the base, the shape of the leaf apex, leaf margin, the color of adaxial leaf, and presence or absence of petiole. The evaluations and observations were accompanied by photographic records.

The determination of the morphological traits was adapted from Miranda (2013) who worked with the morphological characterization and evaluation of the initial development of teak (*T. grandis* L.f.) clones. The choice of the traits to be evaluated was based on the descriptor table for *Eucalyptus* and *Hevea* (BRASIL, 2011); on the studies carried out by Lyngdoh et al. (2007), and Alcantara and Souza (2007); and in similar studies carried out with other species of agronomic interest (Nascimento, 2008; Andrade et al., 2009; Gomes Filho et al., 2010; Pinto et al., 2010).

## ISSR molecular markers

For the study of the genetic variability of the 30 clonal teak genotypes, we used the ISSR molecular marker, because it is a simple, quick, inexpensive, and efficient technique that does not require prior information on the DNA sequence of the organism being studied.

## Collection, extraction, and quantification of DNA

Young, expanded, and healthy leaves from each selected genotype were collected. Each leaf, alone, was wrapped in aluminum foil to be transported from the greenhouse to the Laboratory of Genetics of the University Campus of Alta Floresta - UNEMAT, where the molecular analyses were carried out.

At the Laboratory, genomic DNA was extracted according to the CTAB method (3% CTAB, 1.4 mM NaCl, 0.02 mM EDTA, 0.1 mM Tris-HCl, pH 8.0, and 3%  $\beta$ -mercaptoethanol) described by Doyle and Doyle (1990) and modified based on Alcântara et al. (2008). The quality and concentration of the extracted DNA were confirmed by 0.8% agarose gel electrophoresis. The concentration was compared to molecular weights of standard DNA ( $\lambda$ ) with variation amplitude of 10, 20, 50, and 100 ng. Quantified DNA was diluted to obtain the working solutions.

## Selection of ISSR primers and amplification reactions

Initially, amplification tests were performed on three teak genotypes in 38 ISSR primers developed by University of British Columbia (UBC) (Table 2). Based on the banding patterns of intensity, polymorphism, and repeatability, 15 primers were selected for the PCRs.

**Table 2.** List of primers tested for the molecular characterization of the 30 teak genotypes.

Name	Sequence (5'-3')	AT (°C)	Name	Sequence (5'-3')	AT (°C)
UBC 807*	(AG) <sub>8</sub> T	47	UBC 861*	(ACC) <sub>6</sub>	60.6
UBC 808	(AG) <sub>8</sub> C	48.8	DiCA(YG)	(CA) <sub>8</sub> YG	58.8
UBC 810*	(GA) <sub>8</sub> T	45.4	DiCA(CR)	(CA) <sub>8</sub> CR	58.8
UBC 811*	(GA) <sub>8</sub> C	48	DiCA(CY)	(CA) <sub>8</sub> CY	58.8
UBC 812	(GA) <sub>8</sub> A	48	DiCA(C)	(GA) <sub>8</sub> C	57.2
UBC 816	(CA) <sub>8</sub> T	48	DiCA(RC)	(GA) <sub>8</sub> RC	58.8
UBC 817*	(CA) <sub>8</sub> A	50.3	TriCAC(RC)	(CAC) <sub>8</sub> RC	63.2
UBC 818*	(CA) <sub>8</sub> G	51	TriGTG	(GTG) <sub>8</sub>	58.9
UBC 826*	(AC) <sub>8</sub> C	48	TriGTG(YC)	(GTG) <sub>8</sub> YC	63.2
UBC 827*	(AC) <sub>8</sub> G	53	TriGTG(CY)	(GTG) <sub>8</sub> CY	63.2
UBC 828	(TG) <sub>8</sub> A	51.3	TriAAG(RC)	(AAG) <sub>8</sub> RC	51.2
UBC 830*	(TG) <sub>8</sub> G	52.7	TriACA(RC)	(ACA) <sub>8</sub> RC	51.2
UBC 835*	(AG) <sub>8</sub> YC	50.2	TriAGG(RC)	(AGG) <sub>8</sub> RC	63.2
UBC 840*	(GA) <sub>8</sub> YT	47.4	TriTCA(RC)	(TCA) <sub>8</sub> RC	51.2
UBC 841*	(GA) <sub>8</sub> YC	43	TriTCC(RC)	(TCC) <sub>8</sub> RC	63.2
UBC 842*	(GA) <sub>8</sub> YG	48.8	TriTGA(RC)	(TGA) <sub>8</sub> RC	51.2
UBC 855*	(AC) <sub>8</sub> YT	49	TriCAT(RC)	(CAT) <sub>8</sub> RC	63.2
UBC 856	(AC) <sub>8</sub> YA	51	TriCGA(RC)	(CGA) <sub>8</sub> RC	51.2
UBC 857*	(AC) <sub>8</sub> YG	52	TriGAA(RC)	(GAA) <sub>8</sub> RC	51.2

\*Primers used; Y = C or T; R = A or G; V = A, C, or G; H = A, C, or T; AT = annealing temperature of the primer.

Amplification reactions via PCR were performed in Biocycler thermocycler with a final volume of 20  $\mu$ L, being 1  $\mu$ L DNA ( $\pm$  20 ng), 2  $\mu$ L buffer 10X (1 M KCl; 1 M Tris, pH 8.3; 1 M MgCl<sub>2</sub>; 10% Tween 20), 2  $\mu$ L MgCl<sub>2</sub> (25 mM), 3  $\mu$ L of each primer (0.2 mM),

4  $\mu\text{L}$  dNTP (0.1 mM of each dNTP), 1  $\mu\text{L}$  DMSO and 0.2  $\mu\text{L}$  Taq polymerase (5 U/ $\mu\text{L}$ ). Amplification program was proposed by Narayanan et al. (2006), which was composed by a denaturation cycle of 94°C for 3 min; 30 cycles of 94°C for 30 s, 45°-60.3°C (depending on the temperature of the primer) for 30 s and 72°C for 1 min, and a final extension cycle of 72°C for 10 min.

The amplification products were separated by 1.5% agarose gel electrophoresis in 1X TBE run buffer, with a constant voltage of 85 V for approximately 6 h. The sizes of the amplified fragments were estimated by comparison with the molecular marker of 100-bp DNA ladder. Gel staining was performed with 0.6 ng/mL ethidium bromide. Then, the gels were visualized in transilluminator UV and photo-documented with a digital camera.

### Data analysis

For the morphological data, a joint analysis of the quantitative and qualitative variables was carried out based on Gower's algorithm (1971) to estimate the genetic distance matrix, expressed by the Equation 1:

$$S_{ij} = \frac{\sum_{k=1}^p W_{ijk} \cdot S_{ijk}}{\sum_{k=1}^p W_{ijk}} \quad (\text{Equation 1})$$

where  $k$  is the number of variables ( $k = 1, 2, \dots; p = \text{total number of traits evaluated}$ );  $i$  and  $j$ , to any individual;  $W_{ijk}$  is a weight given to the comparison  $ijk$ , assigning value 1 for valid comparisons and value 0 for invalid comparisons (when the value of the variable is absent in one or both individuals);  $S_{ijk}$  is the contribution of the  $k$  variable in the similarity between individuals  $i$  and  $j$ , having values between 0 and 1. For a nominal variable, if the value of  $k$  variable is the same for both individuals ( $i$  and  $j$ ),  $S_{ijk} = 1$ , otherwise, it is equal to 0. A continuous variable is given by Equation 2:

$$S_{ijk} = 1 - |x_{ik} - x_{jk}| / R_k \quad (\text{Equation 2})$$

where  $x_{ik}$  and  $x_{jk}$  are the values of the  $k$  variable for the individuals  $i$  and  $j$ , respectively, and  $R_k$  is the amplitude of variation of the  $k$  variable in the sample. The division by  $R_k$  eliminates the differences between scales of the variables, producing a value within the interval  $[0, 1]$  and equal weights.

Cluster technique used was the mean linkage between groups (UPGMA), and the adjustment between the distance matrix and the dendrogram was estimated by the cophenetic correlation coefficient (CCC) (Sokal and Rohlf, 1962) and using the Tocher optimization technique. For the molecular traits, the ISSR fragments (amplification products) were analyzed and coded as binary traits: presence (1) or absence (0) of bands.

To obtain the dissimilarity matrix, the arithmetic complement of the Jaccard index was used. From the dissimilarity matrix, clustering genotypes by the Tocher optimization method and building a dendrogram using the linkage mean between groups (UPGMA) were

possible, and the adjustment between the distance matrix and the dendrogram was estimated by the CCC (Sokal and Rohlf, 1962).

Genetic diversity of the locus was estimated based on the polymorphic information content (PIC), which is an estimate used to evaluate the discriminatory power of a locus. The information of pi locus is the frequency of the allele p in locus pi, calculated by Equation 3:

$$PIC = 1 - \sum_i \sum_j p_{ij}^2 \quad (\text{Equation 3})$$

where the information of the primer pij is the frequency of the allele p of the locus i, in the j primer (Rezende et al., 2009). PIC represents the existence of variability, which will be higher with higher values of PIC. Thus, ISSR markers that have this feature are the most suitable for variability studies (Ribeiro, 2011). All statistical analyses of morphological and molecular data were performed in the Genes computational program (Cruz, 2013).

### Joint analysis

Quantitative, qualitative, and molecular variables were simultaneously analyzed together using the Ward-MLM procedure for the composition of the accession groups through the CLUSTER and IML procedure of the SAS program (SAS Institute, 2009). For the use of the Ward clustering method, the distance matrix was obtained by the Gower algorithm (Gower, 1971).

The definition of the ideal number of groups was performed according to pseudo-F and pseudo-t<sup>2</sup> criteria, combined with the likelihood profile, associated with the likelihood ratio test. The Ward-MLM strategy consisted of two phases, in the first phase the groups were defined by the Ward clustering method (Ward, 1963), using the Gower dissimilarity matrix (Gower, 1971), and the second phase, the vector mean of the quantitative variables for each group, regardless of the qualitative variable values, were estimated by the MLM procedure.

The distance adapted by Franco et al. (1998), for the distribution of the joint variables (quantitative, qualitative, and molecular), was used to determine the dissimilarity between the groups formed. To express the variability among the genotypes, the canonical variables and the UPGMA clustering method were used.

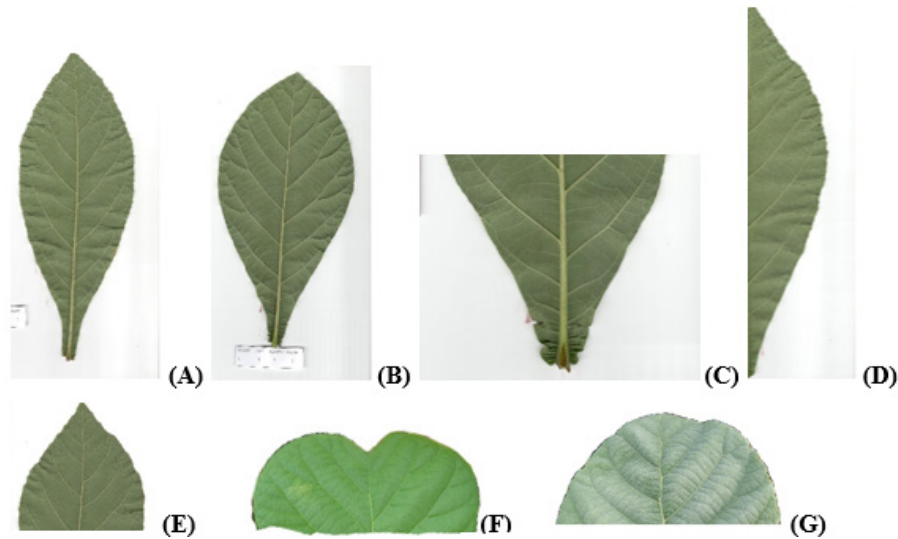
## RESULTS AND DISCUSSION

### Morphological characterization

Among the qualitative variables, the ones that showed variation were: the shape of the leaf blade, the shape of the blade apex, and the presence or absence of petiole. However, the shape of the base, the leaf margin, and the color of the adaxial leaf did not show dissimilarity between the genotypes.

Among the 30 teak genotypes evaluated, we observed that the shape of the leaf blade was between elitist and obovate (Figure 1); the shape of the leaf base was only the attenuated type; for the leaf margin, there was no difference between the genotypes, and the shape found was the corrugated; the color of the adaxial leaf was dark green, and the variable that most diverged was the shape of the apex, whose forms were cuspidate, retuse, and obtuse (Vidal and Vidal, 2003).

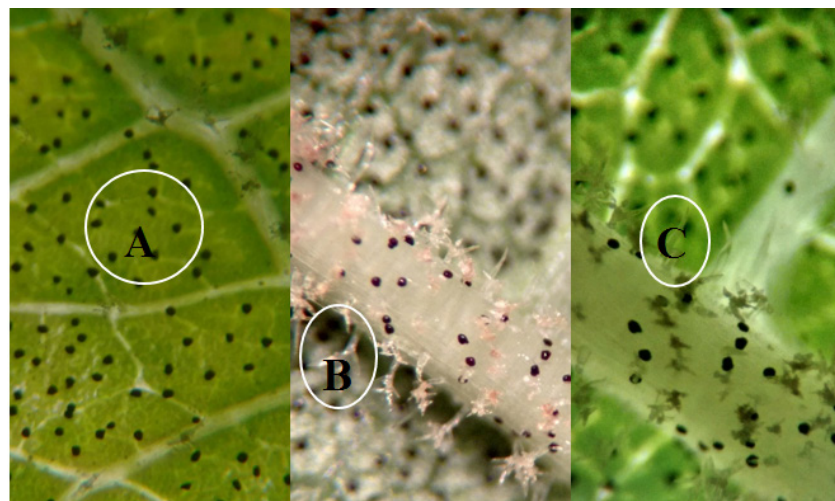




**Figure 1.** Shapes of the leaves of *Tectona grandis* L.f. found in this study. **A.** Elliptical shape of limb; **B.** oval shape of limb; **C.** attenuated shape of the leaf base; **D.** corrugated leaf margin; **E.** cuspidate shape of the apex; **F.** retuse shape of the apex; and **G.** obtuse shape of the apex.

Similar results were found by Miranda (2013) who, working with the morphological characterization and evaluation of the initial development of teak clones (*T. grandis* L. f.), found plants with the shape of the leaf between elliptical in younger leaves and oval in major and mature leaves, apex with acute caudate type, and corrugated margin.

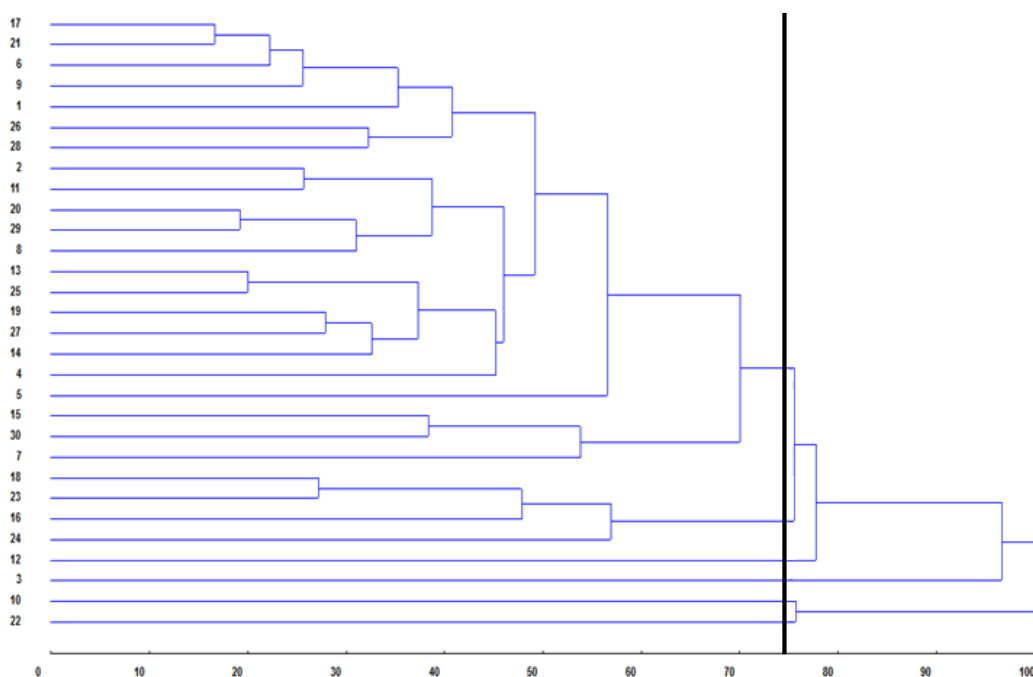
All genotypes presented serosal, glandular, and intense hypoid hairiness on the abaxial face and glandular and hispid hairiness on the adaxial face (Figure 2).



**Figure 2.** Hairiness observed in leaves of *Tectona grandis* L.f. **A.** Glandular; **B.** serice, and **C.** hispid.

A greater occurrence of glandular and hypoid trichomes on the abaxial face of teak leaves was observed. On the adaxial face, besides the glandular trichomes, Fermino Jr and Scherwinski-Pereira (2009) observed the occurrence of pelleted glandular trichomes and tectonic trichomes in leaves of *T. grandis* L.f.

Through the multivariate analysis, based on Gower (1971), we verified high genetic variability among the 30 evaluated teak genotypes and by using the UPGMA clustering method, the formation of six distinct groups was obtained (Figure 3). The first group represented the highest number of genotypes (22 genotypes), demonstrating that these materials have similar morphological characteristics and that this result reduces the perspectives of using these materials combined among themselves in breeding programs of this crop. Group II was formed by genotypes 18, 23, 16, and 24, which showed several common features, such as the shape of the apex in the upper and middle leaves, the absence of petiole and the shape of the limb in the lower leaves.



**Figure 3.** Dendrogram representative of the genetic variability among the 30 teak genotypes obtained by the UPGMA clustering method, using the Gower algorithm as a measure of dissimilarity.

The groups III, IV, V, and VI were formed by a genotype each, which were genotypes 12, 03, 10, and 22, respectively. This result showed that these varieties had a high dissimilarity about other genotypes, showing to be the most promising for use in breeding programs because they are phenotypically divergent to the other genotypes. Similar results were observed by Gomes Filho (2009) who, working with genetic diversity in guava (*Psidium guajava* L.) accessions coming from Bom Jesus do Itabapoana-RJ, found the formation of four groups by the joint analysis, and the stated that the most reliable methodology for assessing the genetic dissimilarity of guava accessions was the joint analysis. The cluster by the UPGMA method



obtained an adjustment with the original distances of the order of 78%; this means that for the UPGMA group, a CCC of 0.78 was obtained, which is considered acceptable, demonstrating that the cluster method is in agreement with the original distances, so there are no major distortions in the results (Sokal and Rohlf, 1962).

In the cluster assessed by the Tocher method, the formation of seven groups was verified, and a result different from that presented by the UPGMA method was found (Table 3). The first group formed by the Tocher method was composed by 23 genotypes, similar to the UPGMA method; however, the genotypes 18, 23, and 24 of group I, in the UPGMA method were allocated in group II. The group II was composed of two genotypes (07 and 30). These genotypes were assigned to group I by the UPGMA method. Groups III, IV, V, VI, and VII were formed by only one genotype, a result similar to that found by the UPGMA method, except genotype 16, which was in the group II by the UPGMA method.

**Table 3.** Cluster of the 30 teak genotypes by the Tocher method, based on the dissimilarity expressed by the Gower algorithm.

Groups	Genotypes
I	17, 21, 06, 09, 26, 20, 02, 11, 27, 29, 25, 01, 28, 19, 14, 13, 04, 08, 15, 05, 18, 23, and 24
II	07 and 30
III	10
IV	22
V	12
VI	16
VII	03

The distance proposed by Gower was also used by Rodríguez et al. (2005) to study together qualitative and quantitative variables, originating from 28 morphological and agronomic traits in *Brassica napus* L. With the study, the authors determined the adequacy of the germplasm studied for the summer crop and estimated the genetic divergence among the local populations. With the joint analysis of the data, the authors were able to demonstrate the diversity and value of the accessions studied for crop improvement.

### Molecular characterization

The ISSR molecular markers revealed a high genetic variability among the 30 selected teak genotypes, and for the 15 selected ISSR primers, a total amplification of 111 fragments was possible, being 99 polymorphic bands (89.19%), with an average of bands amplified by primers of 7.4 (Table 4). Among the 15 primers used, five of them showed 100% polymorphism (UBC 811, UBC 817, UBC 826, UBC 835, and UBC 857). UBC 818 was the primer with the lowest percentage of polymorphism (66.66%).

These results were lower than the ones shown by Ansari et al. (2012) who, studying 29 natural populations of teak based on ISSR markers, found an average of 8.6 bands per primer; however, it was higher than the average verified by Giustina et al. (2013) who, working with genetic diversity of teak genotypes (*T. grandis* Linn f. - Lamiaceae) using ISSR markers, found an average of 4.66 bands per primer.

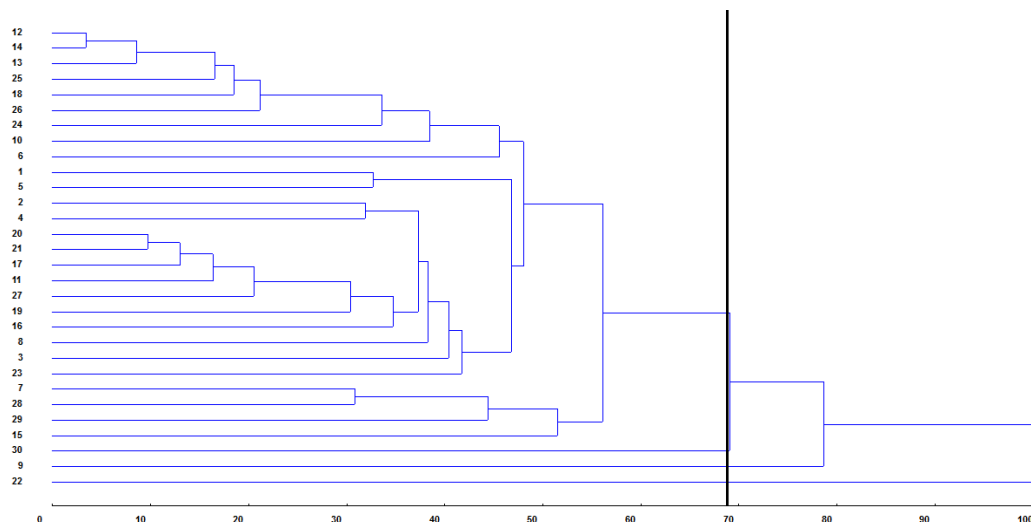
PIC, which according to Varshney et al. (2007) take into account the total number of fragments detected for each locus of a given primer and the frequency of these alleles in the set of genotypes investigated, in this study ranged from 0.70 (UBC 835) to 0.1468 (UBC 827), with mean of 0.4294.

**Table 4.** Total number of bands, percentage of polymorphism, and polymorphic information content (PIC) value of the 30 teak genotypes analyzed.

Primers	Number of bands	% polymorphism	PIC
UBC 807	8	87.5	0.3788
UBC 810	10	90	0.3345
UBC 811	6	100	0.4818
UBC 817	5	100	0.5308
UBC 818	3	66.66	0.2370
UBC 826	12	100	0.3679
UBC 827	8	75	0.1468
UBC 830	10	90	0.2861
UBC 835	4	100	0.7000
UBC 840	4	75	0.4497
UBC 841	8	87.5	0.5647
UBC 842	11	81.81	0.4189
UBC 855	7	85.71	0.4401
UBC 857	10	100	0.5576
UBC 861	5	80	0.5460
Mean	7.4	89.19	0.4294

According to the Botstein et al. (1980) classification, markers with PCI values greater than 0.5 are considered very informative, values between 0.25 and 0.50 are moderately informative, and values lower than 0.25 are not very informative. Thus, in the present study five primers considered very informative were found (UBC 817, UBC 835, UBC 841, UBC 857, and UBC 861), and ISSR markers that present this characteristic are the most indicated for variability studies (Ribeiro, 2011). Eight primers were moderately informative: UBC 807, UBC 810, UBC 811, UBC 826, UBC 830, UBC 840, UBC 842, and UBC 855 and only two primers were considered uninformative: UBC 818 and UBC 827.

Based on the dissimilarity generated by the arithmetic complement of the Jaccard index, it was possible to make a dendrogram by the UPGMA clustering method, and it formed 4 groups (Figure 4). The first group consisted of 27 genotypes, the group with the least different genotypes.

**Figure 4.** Dendrogram obtained by the UPGMA clustering method based on the ISSR markers.

The groups II, III, and IV were formed by only one genotype each, which were the genotypes 30, 9, and 22 respectively; these genotypes are considered the most divergent by the UPGMA clustering method. The CCC obtained between a distance matrix and a dendrogram ( $r = 0.93$ ) revealed a good fit between a graphic representation of distances and an original matrix (Rohlf, 2000), demonstrating that the cluster method is in accordance with the original distances, and there are no large distortions in results, thus revealing to be a safety cluster formed (Sokal and Rohlf, 1962). This result was different from that obtained by the Tocher optimization method, which provided the formation of only 2 groups (Table 5). Group II was formed by genotype 22, which by the UPGMA clustering method also formed a separate group, showing to be the most different genotype among the 30 evaluated teak genotypes. Group I was formed by the other genotypes.

**Table 5.** Cluster of 30 teak genotypes by the Tocher method, from the molecular analysis using the ISSR markers.

Groups	Genotypes
I	12, 14, 13, 25, 18, 26, 24, 19, 10, 21, 11, 20, 17, 27, 03, 29, 05, 02, 16, 08, 01, 04, 23, 06, 28, 07, 15, 30, and 09
II	22

The analysis of the genetic diversity for the knowledge of the variability between materials of interest becomes very advantageous in identifying new sources of genes (Amaral Jr and Thiébaud, 1999). For breeding, which seeks genetic variability in the progeny, aiming at selecting superior cultivars, the most different accessions are the most appropriate (Cruz and Carneiro, 2004).

### Joint data analysis

By analyzing the dissimilarity matrix obtained by the Gower algorithm (1971), it was possible to make a heat map showing the distance between the 30 teak genotypes (Figure 5). The heat map shows that the greater the degree of redness, the greater the similarity between the genotypes, whereas the greater dissimilarities are represented by lighter colors; therefore, we can observe that the greatest similarities are among the genotypes 13, 14, 12, 25, 18, and 26, and the greatest dissimilarities are between genotype 22 and genotypes 3, 5, 27, 11, 21, 17, 23, 2, 19, 24, 18, 25, 12, 14, and 13.

By the procedure of the likelihood function, the ideal number of groups was equal to three, with an increment value of 34.49 (Figure 6). The analysis of the likelihood function allows defining more precise criteria in the formation of the groups, resulting in the establishment of less individual groups (Gonçalves et al., 2009; Barbé et al., 2010; Cabral et al., 2010).

Many studies use the procedure of the likelihood function to determine the ideal number of groups, such as Ortiz et al. (2008) who, studying maize breeds of high altitudes in Peru, observed that the largest increases in probability function were in the constitution of four and eight groups. In the same way, Padilla et al. (2005), when evaluating the diversity of 120 accessions of *Brassica rapa* subsp *rapa* L. found that the greatest increase in probability function was achieved when five groups were considered.

Based on the Gower algorithm, we verified through the UPGMA cluster method the formation of three groups (Figure 7). The first group was formed by 14 genotypes, which are 1, 2, 3, 4, 5, 8, 11, 16, 17, 19, 20, 21, 23, and 27, group II was formed by 13 genotypes (6, 7, 10, 12, 13, 14, 15, 18, 24, 25, 26, 28, and 29), and group III was formed only by genotypes 9, 22, and 30.

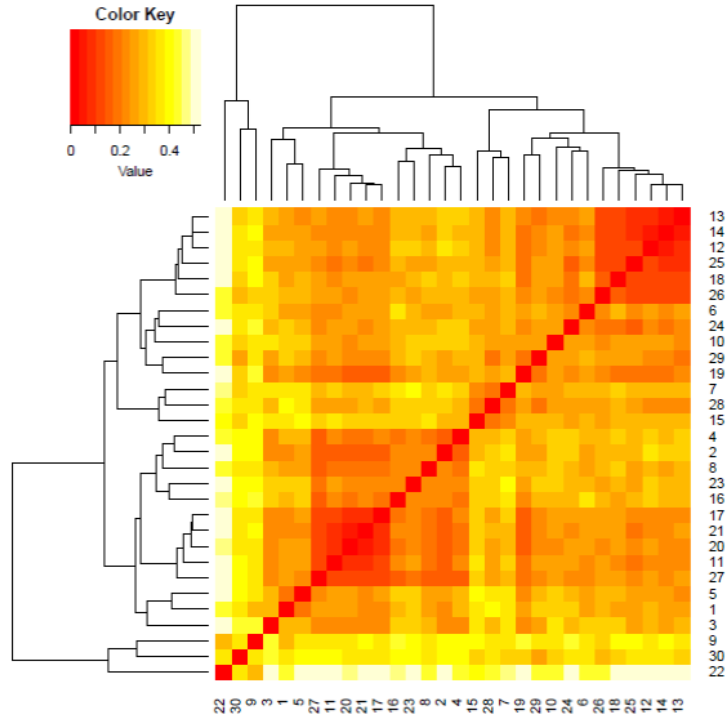


Figure 5. Distances among the 30 teak genotypes by the heat map from the dissimilarity matrix.

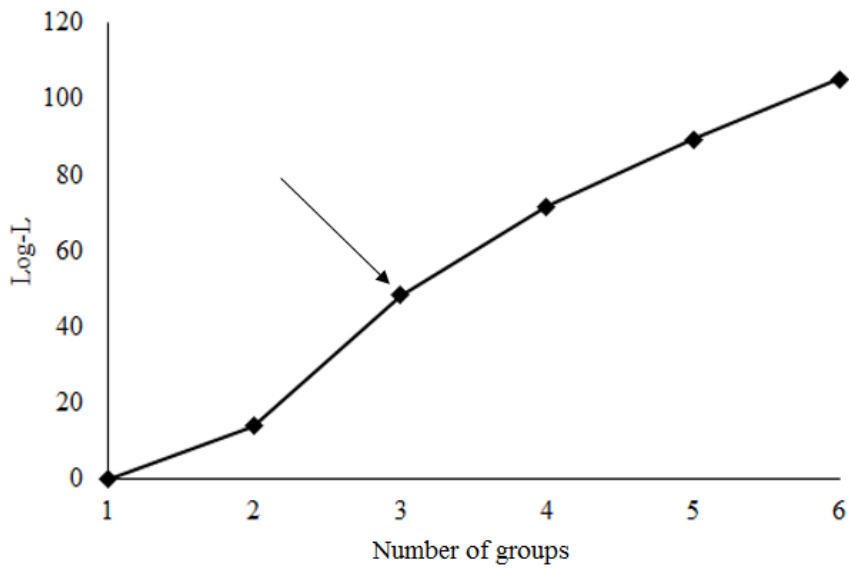
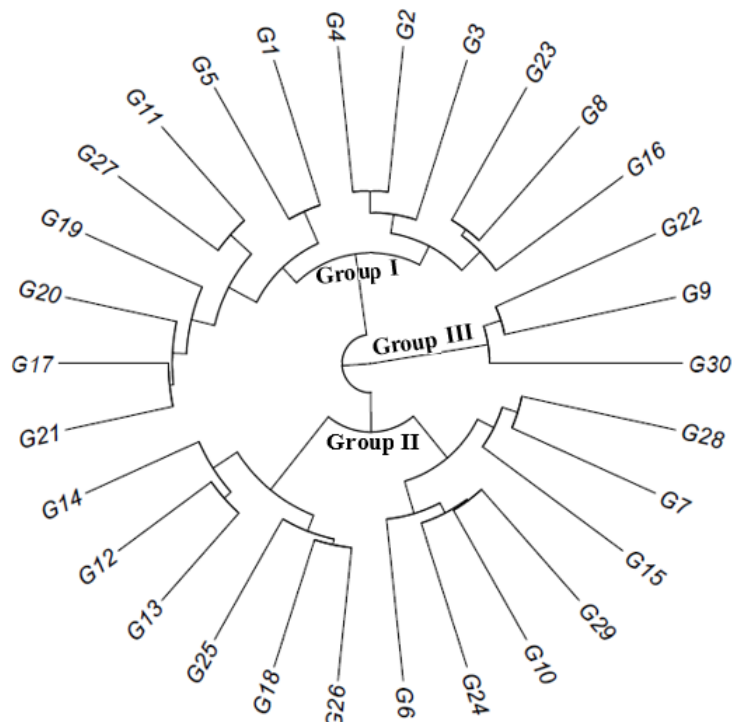


Figure 6. Graph of the logarithmic likelihood function (Log-L) about the number of groups.



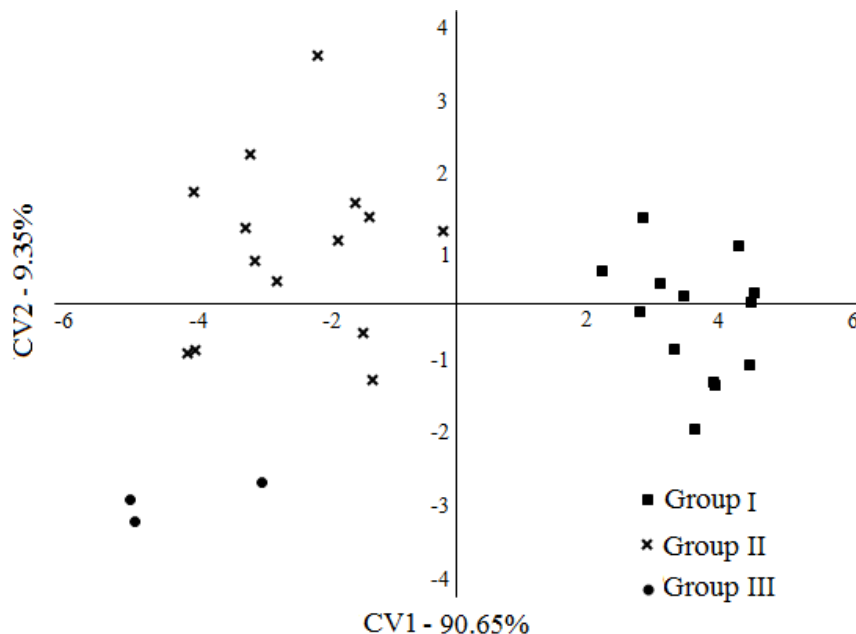
**Figure 7.** Cluster of genotypes by the UPGMA method based on Gower distance obtained by the Ward-MLM procedure.

The main qualitative characteristic in common among the genotypes of group I was the absence of petiole. The quantitative characteristics, which showed higher values, were plant height, fresh and dry mass of aerial part and root, length and width of the median leaf, width of the lower leaf, and glandular hairiness of the adaxial part.

Among the genotypes of the group II, the main qualitative characteristics in common were the absence of petiole and the shape of the limb in the middle leaf of the elastic type. The quantitative characteristics showed emphasis in the diameter of the stem, length of the upper leaf, length of the inferior leaf, pilosity of the hypsid form in the adaxial and abaxial parts, and hairiness in siricius form in the abaxial part.

The group III, formed by the smallest number of genotypes, had a common shape of the limb of elliptical type and shape of the apex of cuspid type in the qualitative characteristics, according to Vidal and Vidal (2003) classification. In the quantitative traits, only the characteristic pilosity of the glandular shape in the abaxial part was superior to the other groups. The CCC was 0.86 revealing a good fit between the graphic distance representation and its original matrix (Rohlf, 2000).

Based on the analysis of canonical variables, it was verified that the first two variables were responsible for 90.65%. Thus, a two-dimensional plot was sufficient to show a good proportion of the genetic variability (Figure 8).



**Figure 8.** Cluster of the genotypes by the first three canonical variables (CV) by the Ward-MLM analysis.

A similar result was found by Gonçalves et al. (2009) who, working with tomatoes, verified that the first two canonical variables explained about 90% of the variability between the groups, and the two-dimensional graph was adequate for the visualization of the relation between the groups. By the canonical analysis of the variables, we can observe a distancing of group I from the other groups formed by the Ward-MLM procedure and the approximation of groups I and II. These results were consistent with those obtained by the distance proposed by Franco et al. (1998) (Table 6), in which we can verify that groups I and III were the most distant (66.02), while groups I and II were the closest (15.24).

**Table 6.** Distance between the groups formed by the Ward-MLM procedure, proposed by Franco et al. (1998).

Group	I	II	III
I	0.00	35.92	66.02
II		0.00	15.24
III			0.00

### Conflicts of interest

The authors declare no conflict of interest.

### ACKNOWLEDGMENTS

We thank the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES) and the National Council for Scientific and Technological Development (CNPq) for financial support.



## REFERENCES

- Alcantara BK and Souza VC (2007). Identificação dos descritores morfológicos para teca (*Tectona grandis*). In: Simpósio Internacional de Iniciação Científica da USP. USP, Pirassununga.
- Alcântara BK, Siqueira MVBM and Veasey EA (2008). Comparação de métodos de extração de DNA de teca (*Tectona grandis* L. f.) para estudos de caracterização genética usando SSR. In: 54º Congresso Brasileiro de Genética. SBG, Salvador.
- Amaral Jr AT and Thiébaud JTL (1999). Análise multivariada na avaliação da diversidade em recursos genéticos vegetais. Universidade Estadual do Norte Fluminense, UENF/CCTA, Campos dos Goytacazes.
- Andrade RA, Lemos EGM, Martins ABG and Paula RC (2009). Caracterização morfológica de rambutan. *Acta Sci. Agron.* 31: 613-619. <https://doi.org/10.4025/actasciagron.v31i4.890>
- Ansari SA, Narayanan C, Wali AS, Kumar R, et al. (2012). ISSR markers for analysis of molecular diversity and genetic structure of Indian teak (*Tectona grandis* L.f.) populations. *Ann. For. Res.* 55: 11-23.
- Barbé TC, Amaral Júnior AT, Gonçalves LSA, Rodrigues R, et al. (2010). Association between advanced generations and genealogy in inbred lines of snap bean by the Ward-Modified Location Model. *Euphytica* 173: 337-343. <https://doi.org/10.1007/s10681-009-0089-z>
- Botstein D, White RL, Skolnick M and Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32: 314-331.
- BRASIL Ministério da Agricultura, Pecuária e Abastecimento (2011). Proteção de Cultivares no Brasil. Secretaria de Desenvolvimento Agropecuário e Cooperativismo. Mapa/ACS, Brasília, 202.
- Cabral PDS, Soares TCB, Gonçalves LSA, Amaral Júnior AT, et al. (2010). Quantification of the diversity among common bean accessions using Ward-MLM strategy. *Pesqui. Agropecu. Bras.* 45: 1124-1132. <https://doi.org/10.1590/S0100-204X2010001000011>
- Cáceres Florestal S/A (1997). Manual do reflorestamento da teca. Cáceres, 30.
- Clark RB (1975). Characterization of phosphatase of intact maize roots. *J. Agric. Food Chem.* 23: 458-460. <https://doi.org/10.1021/jf60199a002>
- Crossa J and Franco J (2004). Statistical methods for classifying genotypes. *Euphytica* 137: 19-37. <https://doi.org/10.1023/B:EUPH.0000040500.86428.e8>
- Cruz CD (2013). Genes: a software package for analysis in experimental statistics and quantitative genetics. *Acta Sci. Agron.* 35: 271-276. <https://doi.org/10.4025/actasciagron.v35i3.21251>
- Cruz CD and Carneiro PCS (2004). Modelos biométricos aplicados ao melhoramento genético II. UFV, Viçosa.
- Cruz CD and Carneiro PCS (2006). Modelos Biométricos Aplicados ao Melhoramento Genético. 2ª ed. UFV, Viçosa.
- Doyle JJ and Doyle JL (1990). Isolation of plant DNA from fresh tissue. *Focus* 13: 13-15.
- Fermino Jr PCP and Scherwinski-Pereira JE (2009). Características anatômicas de folhas de Teca (*Tectona grandis* L.) desenvolvidas sob condições de cultivo *in vitro* e *ex vitro*. *Evidencia* 9: 17-28.
- Food and Agriculture Organization of the United Nations (2009). Future of teak and the high-grade tropical hardwood sector: planted forests and trees working paper FP/44E. Rome. Available at [http://www.fao.org]. Accessed August 23, 2010.
- Franco J, Crossa J, Villaseñor J, Taba S, et al. (1998). Classifying genetic resources by categorical and continuous variables. *Crop Sci.* 38: 1688-1696. <https://doi.org/10.2135/cropsci1998.0011183X003800060045x>
- Giustina LD, Rossi AAB, Passos AF, Barelli MAA, et al. (2013). Diversidade genética de genótipos de teca (*Tectona grandis* LINN F.-Lamiaceae) utilizando marcadores ISSR. In: 7º Congresso Brasileiro de Melhoramento de Plantas, Uberlândia, 212-216.
- Gomes Filho A (2009). Diversidade genética em acessos de goiabeiras (*Psidium guajava* L.) provenientes de Bom Jesus do Itabapoana, RJ. Tese (Doutorado em Genética e Melhoramento de Plantas), Campos dos Goytacazes, RJ.
- Gomes Filho A, Oliveira JG, Viana AP, Siqueira APO, et al. (2010). Marcadores moleculares RAPD e descritores morfológicos na avaliação da diversidade genética de goiabeiras (*Psidium guajava* L.). *Acta Sci. Agron.* 32: 627-633. <https://doi.org/10.4025/actasciagron.v32i4.4720>
- Gonçalves LSA, Rodrigues R, Amaral Jr AT, Karasawa M, et al. (2008). Comparison of multivariate statistical algorithms to cluster tomato heirloom accessions. *Genet. Mol. Res.* 7: 1289-1297. <https://doi.org/10.4238/vol7-4gmr526>
- Gonçalves LSA, Rodrigues R, do Amaral Júnior AT, Karasawa M, et al. (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: 364-374. <https://doi.org/10.4238/vol8-1gmr549>
- Gower JC (1971). A general coefficient of similarity and some of its properties. *Biometrics* 27: 857-874. <https://doi.org/10.2307/2528823>

- International Union for the Protection of New Varieties of Plants (UPOV) (2010). Glossary of terms used in UPOV documents. Available at [[http://www.upov.int/export/sites/upov/en/publications/tgp/documents/tgp\\_1\\_4\\_1.pdf](http://www.upov.int/export/sites/upov/en/publications/tgp/documents/tgp_1_4_1.pdf)]. Accessed April 23, 2011.
- Lyngdoh N, Gugana RP and Vaduseva R (2007). Delineation of teak (*Tectona grandis* L.f.) clones through leaf descriptors. *Indian J. For.* 30: 21-28.
- Miranda MC (2013). Caracterização morfológica e avaliação do desenvolvimento inicial de clones de teca (*Tectona grandis* L.f.). Dissertação (Mestrado em Ciências Florestais e Ambientais) - Faculdade de Engenharia Florestal da Universidade Federal de Mato Grosso. Cuiabá, MT.
- Narayanan C, Dubey S, Wali SA, Shukla N, et al. (2006). Optimization of DNA extraction for ISSR studies in *Tectona grandis* L. f. - an important forest tree species. *Afr. J. Biotechnol.* 5: 1220-1223.
- Nascimento VE (2008). Caracterização de plantas de mamey. Dissertação (Mestrado em Agronomia) - Universidade Estadual Paulista Júlio de Mesquita Filho, Jaboticabal. Available at [<http://www.fcav.unesp.br/download/pgtrabs/pv/m/3463.pdf>]. Accessed April 11, 2011.
- Ortiz R, Crossa J, Franco J, Sevilla R, et al. (2008). Classification of Peruvian highland maize races using plant traits. *Genet. Resour. Crop Evol.* 55: 151-162. <https://doi.org/10.1007/s10722-007-9224-7>
- Padilla G, Cartea ME, Rodríguez VM and Ordás A (2005). Genetic diversity in a germplasm collection of *Brassica rapa* subsp. *rapa* L. from northwestern Spain. *Euphytica* 145: 171-180. <https://doi.org/10.1007/s10681-005-0895-x>
- Pinto JFN, Reis EF, Faleiro FG, Barbosa ECC, et al. (2010). Seleção de descritores vegetativos para caracterização de acessos de guariroba (*Syagrus oleracea* (Mart.) Becc.). *Rev. Bras. Frutic.* 32: 832-840. <https://doi.org/10.1590/S0100-29452010005000094>
- Rezende RK, Paiva LV, Paiva R, Chalfun-Júnior A, et al. (2009). Divergência genética entre cultivares de gérbera utilizando marcadores RAPD. *Ciência Rural* 39: 2435-2440.
- Ribeiro ICNS (2011). Análise da divergência genética em acessos de *Mangifera indica* com base em descritores agromorfológicos e marcadores microssatélites. Feira de Santana, Universidade Estadual de Feira de Santana, 113. (Dissertação - Mestrado em Recursos Genéticos Florestais).
- Rodríguez VM, Cartea ME, Padilla G, Velasco P, et al. (2005). The nabicol: A horticultural crop in northwestern Spain. *Euphytica* 142: 237-246. <https://doi.org/10.1007/s10681-005-1691-3>
- Rohlf FJ (2000). NTSYS-pc: numerical taxonomy and multivariate analysis system, version 2.1. Exeter Software, New York, 83.
- SAS Institute (2009). Statistical analysis system: user's guide. SAS, Cary.
- Sokal RR and Rohlf FJ (1962). The comparison of dendrograms by objective methods. *Taxonomy* 11: 30-40. <https://doi.org/10.2307/1217208>
- Varshney RK, Thiel T, Sretenovic-Rajicic T, Baum M, et al. (2007). Identification and validation of a core set of informative genic SSR and SNP markers for assaying functional diversity in barley. *Mol. Breed.* 22: 1-13. <https://doi.org/10.1007/s11032-007-9151-5>
- Vidal WN and Vidal MRR (2003). Botânica - Organografia, quadros sinóticos ilustrados de fanerógamos, ed. 4, Viçosa 124.
- Vieira EA, Carvalho FIF, Bertan I, Kopp MM, et al. (2007). Association between genetic distances in wheat (*Triticum aestivum* L.) as estimated by AFLP and morphological markers. *Genet. Mol. Biol.* 30: 392-399. <https://doi.org/10.1590/S1415-47572007000300016>
- Ward JH (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58: 236-244. <https://doi.org/10.1080/01621459.1963.10500845>