

## Heterotic groups in tropical maize germplasm by test crosses and simple sequence repeat markers

C.G. Aguiar<sup>1</sup>, I. Schuster<sup>1</sup>, A.T. Amaral Júnior<sup>2</sup>, C.A. Scapim<sup>3</sup>  
and E.S.N. Vieira<sup>1</sup>

<sup>1</sup>Cooperativa Central de Pesquisa Agrícola,  
Setor de Biotecnologia, Cascavel, PR, Brasil

<sup>2</sup>Laboratório de Melhoramento Genético Vegetal,  
Universidade Estadual do Norte Fluminense Darcy Ribeiro,  
Campos dos Goytacazes, RJ, Brasil

<sup>3</sup>Departamento de Agronomia, Universidade Estadual de Maringá,  
Maringá, PR, Brasil

Corresponding author: A.T. Amaral Júnior  
E-mail: amaraljr@uenf.br

Genet. Mol. Res. 7 (4): 1233-1244 (2008)

Received August 15, 2008

Accepted September 4, 2008

Published November 11, 2008

**ABSTRACT.** The objectives of the present study were to determine heterotic groups of germplasm lines of tropical maize by test crosses and by simple sequence repeat (SSR) markers and to compare five grouping methods of heterogeneous maize. Sixteen lines of nine populations in the S<sub>5</sub> generation were evaluated in test crosses with three testers. The results of four experimental trials over two years were used to group the lines by five methods: evaluation based on the hybrid mean in top-cross tests, hybrid index, genetic diversity by the Mahalanobis distance, genetic diversity by the Euclidean distance, and genetic diversity by SSR markers. The concordance of grouping by the Mahalanobis and Euclidean distance amounted to 87.50%, and the concordance of these methods and grouping by SSR markers was 56.25%. Grouping by SSR markers was consistent with

the genealogy of the lines and is a useful procedure for the formation of heterotic groups of tropical maize lines.

**Key words:** Heterotic group; Tropical heterotic maize; *Zea mays* L.; Simple sequence repeat markers; Heterotic grouping methods; Combining ability

## INTRODUCTION

The constitution of heterotic groups is one of the foundation pillars of maize breeding. The grouping of lines in different clusters would avoid the establishment and evaluation of unnecessary hybrids. The formation of such groups of tropical maize lines and the underlying methodological procedures are not yet completely understood.

Among investigators, it is commonly assumed that the combination of lines of different heterotic groups originates hybrids with higher chances of genetic expression of the target effects of hybridization (Troyer, 1999; Austin et al., 2000; Birchler et al., 2003; Tollenaar et al., 2004; Ricci et al., 2007).

Some authors have demonstrated the efficiency of the identification of heterotic groups of maize lines by using molecular procedures such as restriction fragment length polymorphisms (RFLPs; Ajmone-Marsan et al., 1998; Benchimol et al., 2000; Pinto et al., 2003; Warburton et al., 2005), amplified fragment length polymorphisms (AFLPs; Oliveira et al., 2004; Legesse et al., 2007) and simple sequence repeat (SSR) markers (Reif et al., 2003; Barata and Carena, 2006). An advantage over conventional methods is that few divergent lines are not discriminated, and consequently, heterotic groups are formed that contain genotypes, which unequivocally represent the differences in the allele frequency of the populations.

The above studies are consistent, particularly because biometric information of known robustness, e.g., the estimates of combining ability or even of knowledge about the germplasm genealogy evaluated, is not jettisoned for the identification of promising lines in crosses. For example, Reif et al. (2003) demonstrated that SSR markers formed heterotic groups in concordance with the genealogy of subtropical flint and dent composites, derived from germplasm of the CIMMYT.

Studies on tropical germplasm, although rare, have also been conducted. Among these, Oliveira et al. (2004) stated that the use of AFLP markers is a robust procedure for the formation of heterotic groups in maize genotypes adapted to tropical conditions.

The scarcity of studies that deal with tropical germplasm is a constraint to the development of superior hybrids adapted to a great part of the production areas of South America, Africa and Oceania.

Particularly for Brazil, correlated studies are of fundamental importance for the increase in yield of hybrids used by public and private institutions. This becomes even more relevant in view of the possibility that part of the genotypes cultivated by farmers may have been derived from germplasm groups with a high parentage degree.

The present study was undertaken to form heterotic groups of germplasm of tropical maize lines by SSR markers and to compare the accuracy of five grouping methods for the formation of heterogeneous maize groups.

## MATERIAL AND METHODS

Sixteen S<sub>5</sub> maize lines of the Maize Breeding Program of the Cooperativa Central de Pesquisa Agrícola, Paraná State, Brazil (COODETEC), were used in crosses with three testers (Table 1).

**Table 1.** Lines used with the codification and the origin.

Line	Codification	Origin
LIN 03	CD 32-241-1	GU 04-0216 S2 / F1(USA)
LIN 06	CP 66-0009-1	CARGILL 606
LIN 08	CP 71A-0341-1	CARGILL 701
LIN 11	CR 81-012-1	ZENICA 8501
LIN 12	CR 81-013-1	ZENICA 8501
LIN 13	GU 11-022-1	AGROMEN EXP. 11
LIN 17	IMI 238-1	F1 (USA)
LIN 18	LD 18-031-1	AGROCERES 6018
LIN 19	LD 84-010-1	AGROCERES 8014
LIN 25	OR 34-040-1	COLORADO 34 (CO 34)
LIN 26	OR 34-041-1	COLORADO 34 (CO 34)
LIN 27	OR 34-042-1	COLORADO 34 (CO 34)
LIN 29	OR 34-083-1	COLORADO 34 (CO 34)
LIN 30	OR 60-071-1	COLORADO 9560 (CO 9560)
LIN 33	PJ SP-151-1	GU 12-0372-S 3 / CP 53F
LIN 42	UB TK-017-1	TRAKTOR

The crosses were performed in the field, in the late growing season of 2005. The trials were conducted in experimental areas of COODETEC in Palotina, PR, and in Cascavel, PR, in the growing season of 2005/2006. The experiment in Palotina-1st season (P1) was begun on October 19, 2005, Palotina-2nd season (P2) on November 12, 2005, Cascavel-1st season (C1) on October 24, 2005, and Cascavel-2nd season (C2) on December 1, 2005. The experiment was arranged in a complete block design with four replications, with randomized treatments and four controls. The experimental units consisted of two 5-m long rows, with a spacing of 0.8 m between rows.

The grain yield of each hybrid was evaluated. The grain yield data per plot were corrected to kg/ha at 13% moisture and transformed to the index in relation to the hybrid means obtained with the same tester:

$$I = \frac{MH}{MT}, \quad (\text{Equation 1})$$

where  $I$  is the hybrid index;  $MH$  is the hybrid mean, in kg/ha, and  $MT$  is the mean of all hybrids evaluated with the same tester, in kg/ha.

To obtain SSR markers, DNA samples were first taken from plantlet leaves. DNA was extracted as described by Doyle and Doyle (1990). The SSR fragments were amplified in a Thermo Hybaid PCR Express thermal cycler, using a touch-down program. The fragments were separated by horizontal electrophoresis on gels containing 0.7% agarose, 2.15% synergel and 0.2 mg/mL ethidium bromide. Gel images were revealed under UV and recorded for further analysis in a gel documentation system (Vilber Lourmat).

To select the most informative loci, 99 SSR loci were amplified in 10 maize lines. The most informative loci with best amplification quality were amplified for a set of 16 lines.

The genetic information content of each SSR locus was evaluated by allele frequency, using the expression:

$$PIC = 1 - \sum_{j=1}^n p_{ij}^2, \quad (\text{Equation 2})$$

where  $p_{ij}$  is the frequency of the  $j^{\text{th}}$  allele for the  $i^{\text{th}}$  primer (Anderson et al., 1993).

The genetic diversity in the lines was estimated by three approaches: a) average Euclidean distance to index  $I$  of the hybrid; b) Mahalanobis' generalized distance to index  $I$  of the hybrid, and c) distance to SSR loci.

The average Euclidean distance was estimated using the expression:

$$d_{ii'} = \sqrt{\frac{1}{n} \sum_j (X_{ij} - X_{i'j})^2}, \quad (\text{Equation 3})$$

assuming  $X_{ij}$  as observation of the  $j^{\text{th}}$  trait ( $j = 1, 2, \dots, n$ ) evaluated in the  $i^{\text{th}}$  line ( $i = 1, 2, \dots, p$ ) (Cruz and Carneiro, 2003; Mohammadi and Prasanna, 2003).

The Mahalanobis distance was estimated by:

$$D^2 = \sum_{j=1}^n (Z_{ij} - Z_{i'j})^2, \quad (\text{Equation 4})$$

where  $Z_j$  represents the means of the non-correlated and standardized variables, based on the pivotal condensation process (Cruz and Carneiro, 2003).

The genetic distances between the lines to the SSR markers were obtained by a similarity matrix based on the principle of the simple match coefficient for codominant and multiallelic data (Schuster and Cruz, 2003), expressed by:

$$S = \frac{AC}{AT}, \quad (\text{Equation 5})$$

where  $S$  is the similarity index;  $AC$  the number of common alleles, and  $AT$  the total number of alleles evaluated in each plant. For this purpose, the SSR markers were labeled as follows: 2 for alleles in homozygous loci, 1 for each allele in heterozygous loci, and 0 for no alleles. The similarity index was then calculated by the expression:

$$S = \frac{2a + b + c + d}{2L}, \quad (\text{Equation 6})$$

where  $a$  = number of combinations 2 and 2;  $b$  = number of combinations 2 and 1;  $c$  = number of combinations 1 and 2;  $d$  = number of combinations 1 and 1, and  $L$  = total number of evaluated loci. The dissimilarity index ( $D$ ) was obtained by:  $D = 1 - S$ .

The genetic distances based on the average Euclidean distance, Mahalanobis' general-

ized distance and the simple match coefficient for codominant and multiallelic data were estimated using the software package GENES (Cruz, 2006).

With the estimates of genetic distances, the lines were grouped based on the minimum variance method of Ward (1963), using the program Statistica (Statsoft Inc., 1999).

Five methods were used to cluster the heterotic groups: a) hybrid mean, using the comparison of the mean results of each top-cross hybrid with the tester, considering the data of the four environments in joint analysis; b) hybrid index ( $I$ ), where the information of the top crosses was considered based on the mean of the four environments, including lines with hybrid index values below 1.05 in the same group; c) hybrid index ( $I$ ), using Mahalanobis' generalized distance and Ward's grouping; to calculate Mahalanobis' generalized distance, each environment was considered a replication, and the variables were the means of the indices ( $I$ ) of each tester; d) hybrid index ( $I$ ), considering each hybrid in each environment as variable; the genetic distances were estimated by the average Euclidean distance, considering each environment as individual data, and using Ward's method of grouping, and e) grouping based on SSR markers, using Ward's method.

Of the five methods used for the formation of heterotic groups, four were based on test crosses, and for these, the number of testers was different. The classification was, therefore, based on groups of lines evaluated by the same testers, to allow a systematic comparison. A set of 16 lines was considered, which were hybridized with the same three testers.

The concordance index of the four grouping methods was obtained by counting the total coincidences of all lines evaluated in each method.

## RESULTS AND DISCUSSION

Ninety-nine SSR primers, distributed across 99 bins of the maize genome, were amplified in a set of 10 genetically unrelated lines. Of these, 59.6% were polymorphic and 28 were selected in view of their information degree and amplification quality, for the characterization of the 44 lines.

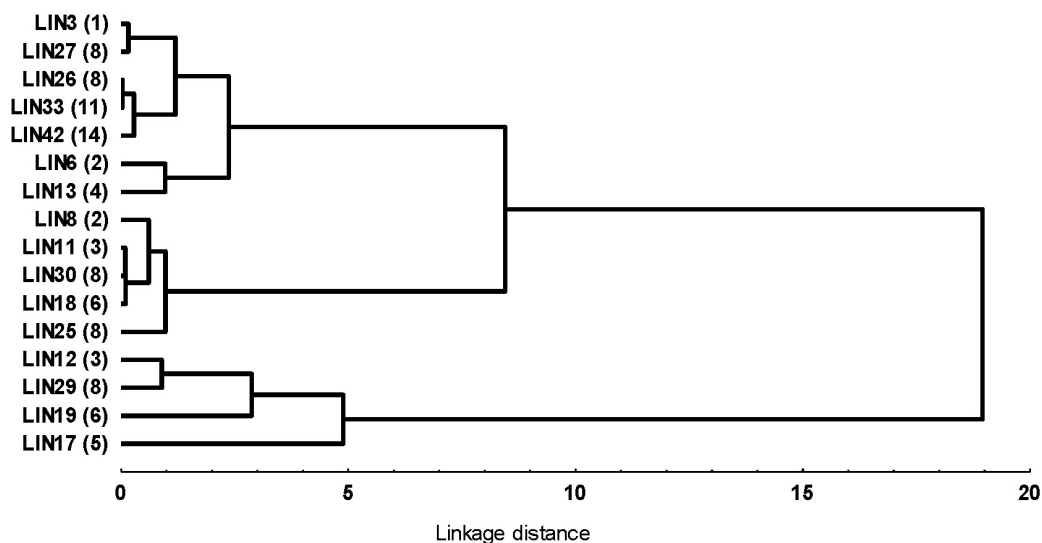
For the selected primers, 96 alleles were identified, with an average of 3.4 alleles per locus and a variation of 2 to 5 alleles per locus. The polymorphism index (PIC) varied from 0.05 to 0.66 (mean of 0.51).

Sênior et al. (1998), using the same separation procedure of SSR fragments for the characterization of maize lines, identified 70 polymorphic primers with adequate amplification quality in a set of 120 tested primers, which accounted for 58% of the total polymorphism, with a variation of 2 to 23 alleles per locus.

The comparatively greater range of variation of alleles per locus observed by Sênior et al. (1998) can be explained by the fact that a larger set of lines had been used. In this case, a greater number of alleles would be expected to be detected. It must, however, be emphasized that the majority of these alleles were of rare occurrence. Consequently, considering only the alleles found in more than two lines, the number of alleles per locus varied from 2 to 9, in agreement with the results of the cited study. Besides, the estimates of the PIC values obtained by Sênior et al. (1998) were also not in disagreement, with a variation between 0.17 and 0.94.

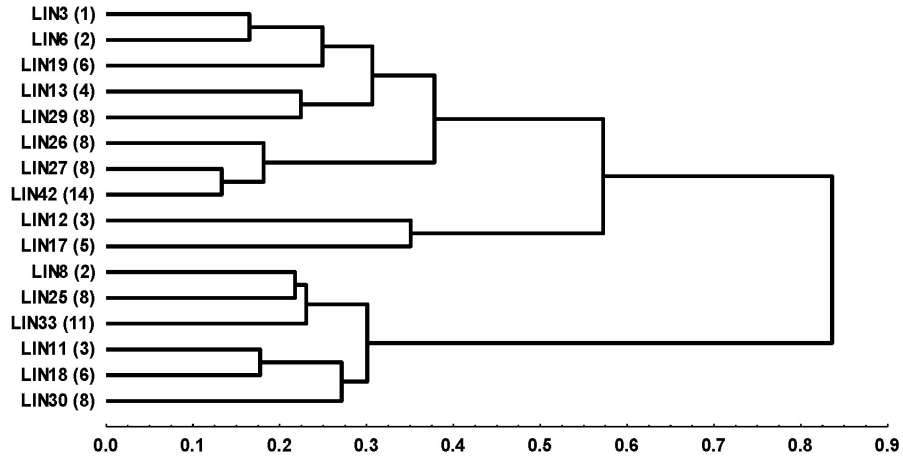
The grouping of these 16 lines using Mahalanobis' generalized distance based on

the hybrid indices in relation to the mean of the respective testers is shown in Figure 1. The definition of the cut-off point took into account the number of groups obtained by the method conventionally used by breeding programs, which is based on the hybrid means in each environment. When considering the cut-off at a distance of 10, two groups were formed. Group 1 contained only lines unrelated with the tester. In group 2, although a separation in two subgroups was possible, the two contained progenies of tester-related lines, so that the lines 25, 26, 27, and 30 together, originated by a single tester, composed one and the same heterotic group. It is interesting to observe that these lines came from the same genealogy, i.e., a tropical maize produced by COLORADO Company, located in São Paulo, Brazil. Furthermore, the grouping of lines 11, 18 and 30 with little genetic distance reveals an interesting situation. Considering that the lines came respectively from the companies ZENICA, AGROCERES and COLORADO, this indicates that the materials provided by different breeding programs in Brazil can be of high genetic similarity. Consequently, this favors a reduction in the genetic base of materials that had been worked from release to the farmers in the tropically edaphoclimatic conditions.



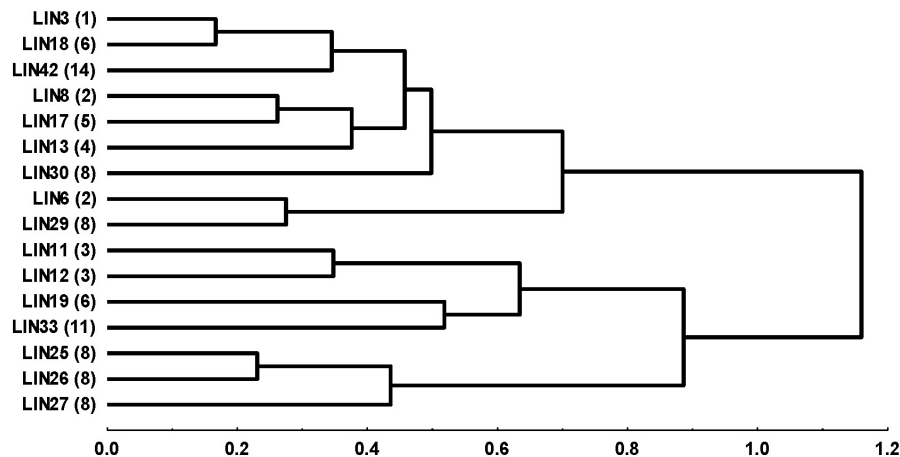
**Figure 1.** Grouping analysis by Ward's hierarchical method of 16 maize lines crossed with three testers, based on the matrix of Mahalanobis' generalized distances and the hybrid yield indices in relation to the means of the respective testers. The same numbers in parentheses after the identification of the line represent lines derived from the same germplasm.

Based on the average Euclidean distance, the grouping of lines (Figure 2), where the indices of each environment were weighed individually, two groups were formed for all testers. Similar to the grouping by Mahalanobis' generalized distance, a relationship between group formation and origin of the lines was observed, although weaker, since in this case the group with highest concordance contained three lines derived from the same tester: LIN 26, LIN 27 and LIN 29.



**Figure 2.** Grouping analysis by Ward's hierarchical method of 16 maize lines crossed with three testers, based on the matrix of average Euclidean distances, predicated on the indices of hybrid yield compared to the mean of the respective testers. The same numbers in parentheses after the identification of the line represent lines derived from the same germplasm.

The estimates of grouping based on genetic divergence for the SSR markers clearly separated the lines into two groups (Figure 3). Again, the number of progenies originated by the same tester, which composed a similar group, was at most three: LIN 25, LIN 26 and LIN 27. However, contrary to the previous groupings, the progenies of tester 3 - LIN 11 and LIN 12, respectively - were not clustered in dissimilar groups, expressing a higher concordance of grouping and origin. In fact, lines 11 and 12 came from the ZENICA Company and originated the same parental, designated ZENICA 8501.



**Figure 3.** Grouping analysis by Ward's hierarchical method of 16 maize lines crossed with three testers, predicated on the matrix of genetic distances, based on molecular data. The same numbers in parentheses after the identification of the line represent lines derived from the same germplasm.

The grouping of lines 3 and 27, 6 and 18 (Figures 1, 2, and 3) is not obscured despite line 3 coming from the United States and 27, 6 and 18 from COLORADO, CARGILL and AGROCERES, all Brazilian Companies. In this case, it was observed that there were few progenies with similarity among these lines. The major similarity occurred between lines 3 and 27 with the values 1 and 8, respectively (Figure 1).

The formation of two groups by the three grouping methods shows inconsistency of the procedures, particularly since the lines in the groups were not the same using the different methods. This is also evidenced by the classification of the lines into heterotic groups (Table 2), where the differential groupings by the hybrid index and hybrid mean are observed as well. Interestingly, the hybrid mean formed only groups 1 and 3, and only one line constituted group 2. Consequentially, it would be expected that the smallest proportion of lines would form group 2, since the highest means were observed for tester 2 in crosses with the evaluated lines (Table 3). In terms of combining ability, this demonstrates the more significant contribution of the respective tester to the performance of the hybrid combinations.

**Table 2.** Classification of 16 maize lines in heterotic groups using four classification methods with test cross data for three testers and by molecular data.

Line	Heterotic groups classified according to the method				
	Hybrid mean	Hybrid index	Mahalanobis distance	Euclidean distance	Molecular analysis
LIN 3 (1)	3	1 / 3	3.2	3.2	3
LIN 6 (2)	3	1 / 3	3.2	3.2	3
LIN 8 (2)	1	1 / 2 / 3	3.1	3.1	3
LIN 11 (3)	1	1 / 2 / 3	3.1	3.1	1
LIN 12 (3)	3	2 / 3	1	1	1
LIN 13 (4)	3	1 / 2 / 3	3.2	3.2	3
LIN 17 (5)	1	2	1	1	3
LIN 18 (6)	1	1 / 3	3.1	3.1	3
LIN 19 (6)	3	2 / 3	1	3.2	1
LIN 25 (8)	?	1 / 2	3.1	3.1	1
LIN 26 (8)	3	1 / 3	3.2	3.2	1
LIN 27 (8)	1	1	3.2	3.2	1
LIN 29 (8)	3	1 / 2 / 3	1	3.2	3
LIN 30 (8)	3	2 / 3	3.1	3.1	3
LIN 33 (11)	3	1 / 3	3.2	3.1	1
LIN 42 (11)	3	1 / 2 / 3	3.2	3.2	3

The grouping based on the hybrid mean (Table 2) detected one line, LIN 25, which could not be included in any group, due to its high combining ability with the three testers.

It had been expected that in the hybrid mean-based grouping, the highest proportion of the lines would be classified in group 3 - of the 16 lines considered, 10 were classified in group 3 (Table 2) - since tester 3 had the lowest mean performance (Table 3). It must be pointed out, however, that the clustering of most lines in group 3 was due to the lower combining ability of tester 3, and not only because of the lower heterosis of the hybrid combinations. Therefore, the data transformation to the same scale, as well as the establishment of an index related to the tester mean, may represent the heterosis of the hybrid combinations better than a direct evaluation of performance.

However, grouping by the hybrid index proved to be little discriminatory, since the lines could mostly be inserted in the same group by at least two distinct testers. It is worth pointing out that this method was based on the criterion that lines were included in the same group when the indices were below 1.05. The underlying principle for grouping determined that the heterosis in the lines had



to be lower than 5%, compared with the tester mean. In five of the sixteen lines considered, the index was lower than 1.05, in the mean of the four environments for the three testers, and in nine the index was below 1.05 for two testers. Only two lines were clustered in a separate group (Table 3).

**Table 3.** Indices on each test cross in relation to the mean of the tester used.

Line	CX15-037 (T1)					XE23-041B (T2)					CP99-005B (T3)				
	P1	P2	C1	C2	Mean	P1	P2	C1	C2	Mean	P1	P2	C1	C2	Mean
LIN 01 (1)	-	-	-	-	-	1.22	1.46	1.20	0.84	1.12	-	-	-	-	-
LIN 02 (1)	-	-	-	-	-	1.17	1.40	1.08	0.83	1.06	-	-	-	-	-
LIN 03 (1)	0.80	1.35	1.06	0.87	0.98	1.04	0.99	1.12	1.02	1.05	1.20	0.94	1.10	0.78	0.99
LIN 04 (1)	1.07	1.30	1.08	0.84	1.02	1.14	1.28	1.02	1.02	1.08	-	-	-	-	-
LIN 05 (2)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LIN 06 (2)	0.98	1.40	0.95	0.97	1.01	1.16	0.96	0.98	1.15	1.07	1.01	0.56	0.98	0.76	0.85
LIN 07 (2)	0.98	1.10	0.87	1.26	1.05	-	-	-	-	-	-	-	-	-	-
LIN 08 (2)	0.73	0.27	0.78	0.93	0.77	1.12	0.65	1.05	1.09	1.02	0.95	0.75	0.94	1.15	0.99
LIN 09 (3)	0.88	0.68	1.11	1.01	0.99	-	-	-	-	-	1.15	1.25	1.17	1.13	1.16
LIN 10 (3)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LIN 11 (3)	0.69	0.52	0.96	0.73	0.78	0.74	1.31	1.12	1.00	1.03	1.04	0.91	1.02	0.88	0.96
LIN 12 (3)	1.04	0.72	1.12	1.19	1.09	0.63	0.68	0.88	1.17	0.90	0.99	1.64	0.85	0.99	1.03
LIN 13 (4)	1.11	1.32	0.96	1.01	1.04	0.65	0.88	1.17	1.15	1.02	0.92	0.71	1.11	1.08	1.01
LIN 14 (4)	-	-	-	-	-	1.18	1.53	1.14	0.81	1.10	-	-	-	-	-
LIN 15 (4)	1.13	0.72	1.10	1.10	1.06	0.88	0.82	0.84	0.95	0.88	-	-	-	-	-
LIN 16 (4)	1.34	1.54	1.13	0.99	1.16	-	-	-	-	-	-	-	-	-	-
LIN 17 (5)	0.88	1.60	1.18	1.01	1.11	0.95	1.05	0.85	0.91	0.91	1.04	1.66	1.12	1.22	1.21
LIN 18 (6)	0.56	0.57	0.94	0.65	0.74	1.15	1.06	1.07	0.97	1.05	0.92	1.14	0.91	1.03	0.98
LIN 19 (6)	1.05	1.32	1.21	1.14	1.17	0.81	0.87	0.81	0.77	0.81	1.02	0.74	0.65	0.74	0.76
LIN 20 (6)	1.11	1.05	1.31	1.06	1.16	-	-	-	-	-	1.03	1.15	0.86	1.09	1.01
LIN 21 (7)	0.87	1.15	0.91	0.93	0.94	-	-	-	-	-	0.60	1.63	0.87	1.58	1.16
LIN 22 (8)	1.12	1.19	0.84	1.21	1.06	-	-	-	-	-	1.13	0.43	1.12	0.77	0.91
LIN 23 (8)	-	-	-	-	-	1.01	0.46	1.06	1.17	1.00	0.89	0.49	1.07	0.81	0.87
LIN 24 (8)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LIN 25 (8)	0.95	0.31	0.68	1.01	0.80	1.16	1.20	1.04	1.02	1.08	1.16	0.95	1.20	1.12	1.13
LIN 26 (8)	0.89	0.95	0.96	0.70	0.85	1.26	0.91	1.13	1.00	1.08	0.93	0.89	1.09	0.93	0.98
LIN 27 (8)	1.29	0.91	0.86	0.93	0.96	1.09	1.15	1.02	1.00	1.05	1.19	0.98	0.99	1.12	1.07
LIN 28 (8)	1.01	1.05	0.93	1.09	1.01	-	-	-	-	-	0.58	0.88	0.90	0.88	0.83
LIN 29 (8)	1.01	0.79	0.89	0.99	0.93	1.03	0.69	0.75	0.80	0.81	1.11	1.07	1.07	0.86	1.01
LIN 30 (8)	1.05	1.14	0.91	1.17	1.05	0.85	0.44	0.93	0.81	0.81	0.89	1.33	0.94	0.92	0.98
LIN 31 (9)	-	-	-	-	-	1.08	1.31	1.14	1.00	1.10	-	-	-	-	-
LIN 32 (10)	0.77	0.54	0.81	0.91	0.81	0.98	0.71	0.88	0.99	0.91	-	-	-	-	-
LIN 33 (11)	0.77	0.60	1.05	0.97	0.92	0.99	1.20	0.91	1.16	1.05	1.02	0.55	0.97	1.22	1.01
LIN 34 (12)	0.83	1.64	1.00	1.21	1.11	-	-	-	-	-	-	-	-	-	-
LIN 35 (12)	1.03	1.99	0.98	1.13	1.15	0.88	0.94	0.72	1.12	0.91	-	-	-	-	-
LIN 36 (12)	1.26	1.22	1.26	1.14	1.21	1.31	1.14	0.95	0.92	1.04	-	-	-	-	-
LIN 37 (13)	0.94	1.12	1.18	0.91	1.04	-	-	-	-	-	-	-	-	-	-
LIN 38 (13)	-	-	-	-	-	0.70	1.03	1.01	1.21	1.02	-	-	-	-	-
LIN 39 (13)	1.03	1.13	0.86	1.05	0.99	-	-	-	-	-	1.26	1.52	1.05	1.06	1.16
LIN 40 (14)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LIN 41 (14)	0.97	0.90	0.97	0.98	0.97	1.06	0.80	1.07	1.02	1.01	-	-	-	-	-
LIN 42 (14)	1.25	0.83	0.92	0.76	0.91	1.11	0.95	0.98	1.08	1.04	0.97	0.84	1.03	0.91	0.95
LIN 43 (14)	0.85	1.27	1.08	1.15	1.09	0.64	1.50	0.97	1.14	1.03	-	-	-	-	-
LIN 44 (14)	1.02	1.25	1.03	1.10	1.08	0.87	1.24	1.08	1.26	1.12	-	-	-	-	-
Mean	4.263	2.743	9.356	9.008	6.343	5.955	4.288	10.116	9.878	7.559	4.488	3.141	7.955	7.916	5.875

T1, T2 and T3 = testers 1, 2 and 3. P1 and P2 = Palotina 1 and 2. C1 and C2 = Cascavel 1 and 2. Same numbers in parentheses indicate lines derived from the same germplasm. Mean = in kg/ha from all hybrids obtained with the testers. - = lost data or crossing not carried out with set grouping of testers.

This distortion in classification based on the mean of the hybrid index can be explained by the fact that when the environment mean is considered, the heterosis of each tester decreases to below the heterosis level in the separate analysis of the environments. This drawback of the index can be minimized by evaluating the data considering not only the mean, but the index in each environment. This can be inferred from the example that the index of LIN 08 was equal to or higher than 1.05 in three of the four environments for tester 2 (Table 3), whereas the average index was lower.

Regardless of the fact that the procedure clearly needs to be further developed, the grouping based on the mean of all hybrids with the tester represented an attempt of establishing a method of unbiased heterotic group formation that lends itself to automation, with the further advantage that the group formation is mutually exclusive.

By way of explanation, it should be added that the constitution of heterotic groups based on the direct evaluation of each hybrid is influenced by the combining ability of the tester, i.e., those test crosses with testers that had high combining ability have a higher mean, and the lines tend not to be classified in this group.

Based on the characterization of genetic diversity among maize inbred lines developed at CIMMYT for hybrid production by using 32 RFLP markers for the purpose of defining potential heterotic groups formed with molecular data and identifying the most representative testers for each potential groups, Warburton et al. (2005) concluded that lines did not cluster based on heterotic response (as determined based on hybrid performance with testers), but lines related by pedigree usually did cluster together. Also, Legesse et al. (2007) on the evaluation of 26 inbred lines crossed with six testers (population and line) in five locations in Ethiopia, concluded that the genetic distances derived from the inbred lines versus all testers and from the population testers' sub-group were not positively correlated with hybrid performance and midparent heterosis for most traits. On the other hand, cluster analysis using AFLP markers separated the tester parents from the corresponding inbred lines in agreement with their pedigree records.

In Brazil, Balestre et al. (2008) with the purpose to correlate the genetic distance of single-cross hybrids with yield, heterosis and specific combining ability in the double-cross hybrid synthesis, concluded that there was a medium correlation between genetic distance and heterosis ( $r = 0.40$ ) and genetic distance and specific combining ability ( $r = 0.38$ ). Besides, the authors demonstrated that the SSR markers were efficient in placing hybrids in different heterotic groups and were also useful in eliminating the most negative heterosis and specific combining ability.

The concordance proportions of the classification of the lines in heterotic groups (Table 4), when one method allocated a line to more than one group, was considered coincident with another method that would classify the same line in any one of these groups.

**Table 4.** Proportion of concordances in the classification of 16 maize lines in heterotic groups by five methods based on test cross data with three testers or on molecular data.

Strategy	Hybrid index	Mahalanobis distance	Euclidean distance	Molecular analysis
Hybrid mean	14/15	8/15	10/15	8/15
Hybrid index		11/16	12/16	13/16
Mahalanobis distance			14/16	9/16
Euclidean distance				9/16

In the cases where a method classifies the line in two groups, the classification is considered coincident when one group coincides.

Owing to this criterion, the concordance with all other methods was highest by the hybrid index classification, since most lines were assigned to more than one group and the probability of concordance was consequently higher.

The classification of lines in heterotic groups using Mahalanobis' generalized distance and average Euclidean distance was also quite coincident. This high concordance had also been expected, since the two methods used the same data set. The high concordance between these estimates of genetic distance measures was demonstrated above. Only the classification was not completely coincident, due to the residual correlations between the indices obtained with the respective testers, which are related to the variability of each tester in the different environments.

The proportion of concordance of the groupings obtained by Mahalanobis' generalized and the average Euclidean distance with the grouping obtained based on the SSR markers on the one hand, and the grouping concordance of these three methods with the approach based on the hybrid mean on the other hand was much the same (Table 4), with a percentage of over 50%.

Admittedly, the concordance in the formation of heterotic groups with SSR markers was not absolute for grouping based on the hybrid mean; notwithstanding, the use of molecular protocols is a valuable tool. By conventional procedures, the number of lines to be characterized in a maize hybrid breeding program is generally rather high. It is not unusual that programs comprise thousands of lines, as recommended for tropical germplasms by Warburton et al. (2002) and in the compared analysis between different molecular markers by Pejic et al. (1998). Furthermore, test crossing with a large number of lines is extremely labor and cost-intensive, besides requiring an adequate infra-structure and too much time. Moreover, the results of test crosses are influenced by genotype by environment interaction, and may vary according to the year and location of evaluation of the tester-derived hybrids.

It is worth adding that the formation of heterotic groups using information from test crosses is tester-dependent. The discrimination of heterotic groups based on estimates of genetic diversity obtained with molecular marker data, on the other hand, is not influenced by the testers and is not susceptible to genotype by environment interaction, besides requiring less time, since no cross has to be performed.

## ACKNOWLEDGMENTS

The authors thank Conselho Nacional de Pesquisa (CNPq) and Fundação Araucária for financial support.

## REFERENCES

- Ajmone-Marsan P, Castiglioni P, Fusari F, Kuiper M, et al. (1998). Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. *Theor. Appl. Genet.* 96: 219-227.
- Anderson JA, Churchill GA, Autrique JE, Tanksley SD, et al. (1993). Optimizing parental selection for genetic linkage maps. *Genome* 36: 181-186.
- Austin DF, Lee M, Veldboom LR and Hallauer AR (2000). Genetic mapping in maize with hybrid progeny across testers and generations: grain yield and grain moisture. *Crop Sci.* 40: 30-39.
- Balestre M, Machado JC, Lima JL, Souza JC, et al. (2008). Genetic distance estimates among single cross hybrids and correlation with specific combining ability and yield in corn double cross hybrids. *Genet. Mol. Res.* 7: 65-73.
- Barata C and Carena MJ (2006). Classification of North Dakota maize inbred lines into heterotic groups based on molecular and testcross data. *Euphytica* 151: 339-349.

- Benchimol LL, Souza CL Jr, Garcia AAF, Kono PMS, et al. (2000). Genetic diversity in tropical maize inbred lines: heterotic group assignment and hybrid performance determined by RFLP markers. *Plant Breed.* 119: 491-496.
- Birchler JA, Auger DL and Riddle NC (2003). In search of the molecular basis of heterosis. *Plant Cell* 15: 2236-2239.
- Cruz CD (2006). Programa Genes: Biometria. Editora Universidade Federal de Viçosa, Viçosa.
- Cruz CD and Carneiro PCS (2003). Modelos Biométricos Aplicados ao Melhoramento Genético. Editora Federal de Viçosa, Viçosa.
- Doyle JJ and Doyle JL (1990). Isolation of plant DNA from fresh tissue. *Focus* 12: 13-15.
- Legesse BW, Myburg AA, Pixley KV, Twumasi-Afriyie S, et al. (2007). Relationship between hybrid performance and AFLP based genetic distance in highland maize inbred lines. *Euphytica* 162: 313-323.
- Mohammadi SA and Prasanna BM (2003). Analysis of genetic diversity in crop plants - salient statistical tools and considerations. *Crop Sci.* 43: 1235-1248.
- Oliveira KM, Laborda PR, Garcia AAF, Paterniani MEAG, et al. (2004). Evaluating genetic relationships between tropical maize inbred lines by means of AFLP profiling. *Hereditas* 140: 24-33.
- Pejic I, Ajmone-Marsan P, Morgante M, Kozumplick V, et al. (1998). Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs. *Theor. Appl. Genet.* 97: 1248-1255.
- Pinto RMC, Souza CL Jr, Carlini-Garcia LA, Garcia AAF, et al. (2003). Comparison between molecular markers and diallel crosses in the assignment of maize lines to heterotic groups. *Maydica* 48: 63-73.
- Reif JC, Melchinger AE, Xia XC, Warburton ML, et al. (2003). Use of SSRs for establishing heterotic groups in subtropical maize. *Theor. Appl. Genet.* 107: 947-957.
- Ricci GC, Silva N, Pagliarini MS and Scapim CA (2007). Microsporogenesis in inbred line of popcorn (*Zea mays* L.). *Genet. Mol. Res.* 6: 1013-1018.
- Schuster I and Cruz CD (2003). Similaridade genética obtida por um coeficiente de coincidência simples para dados codominantes e multi-alelicos. In: 49º Congresso Brasileiro de Genética, Sociedade Brasileira de Genética, Águas de Lindóia.
- Sênior ML, Murphy JP, Goodman MM and Stuber CW (1998). Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. *Crop Sci.* 38: 1088-1098.
- Statsoft Inc. (1999). Statistica for Windows (Computer Program Manual). Version 5.5, Tulsa. Available at [<http://www.statsoft.com>].
- Tollenaar M, Ahmadzadeh A and Lee EA (2004). Physiological basis of heterosis for grain yield in maize. *Crop Sci.* 44: 2086-2094.
- Troyer AF (1999). Background of U.S. hybrid corn. *Crop Sci.* 39: 621-626.
- Warburton ML, Xianchun X, Crossa J, Franco J, et al. (2002). Genetic characterization of CIMMYT inbred maize lines and open pollinated populations using large scale fingerprinting methods. *Crop Sci.* 42: 1832-1840.
- Warburton ML, Ribaut JM, Franco J, Crossa J, et al. (2005). Genetic characterization of 218 elite CIMMYT maize inbred lines using RFLP markers. *Euphytica* 142: 97-106.
- Ward JH (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58: 236-244.