

## Evaluation of genetic variability to form heterotic groups in popcorn

C. Vittorazzi<sup>1</sup>, A.T. Amaral Junior<sup>1</sup>, A.G. Guimarães<sup>1</sup>, F.H.L. Silva<sup>1</sup>,  
G.F. Pena<sup>1</sup>, R.F. Daher<sup>1</sup>, I.F.S. Gerhardt<sup>1</sup>, G.H.F. Oliveira<sup>1</sup>,  
P.H.A.D. Santos<sup>1</sup>, Y.P. Souza<sup>1</sup>, S.H. Kamphorst<sup>1</sup> and V.J. Lima<sup>1</sup>

<sup>1</sup>Laboratório de Melhoramento Genético Vegetal, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos dos Goytacazes, RJ, Brasil

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

Genet. Mol. Res. 17 (3): gmr18083  
Received July 18, 2018  
Accepted September 20, 2018  
Published August XX, 2018  
DOI <http://dx.doi.org/10.4238/gmr18083>

**ABSTRACT.** In view of the narrow genetic base of popcorn, probably due to its evolution by selection from flint maize types alone, knowledge about genetic divergence is imperative for the formation of heterotic groups. Thus, our objective was to identify heterotic groups of popcorn lines; we did so by exploiting the representative genotype collection of the Active Popcorn Germplasm Bank of the State University of Northern Rio de Janeiro. Thirty-eight popcorn genotypes from different origins were analyzed by two methodologies to identify divergent groups. In the first method, the genotypic data were processed to determine the number of groups, based on Bayesian clustering algorithms, and two clustering methods (UPGMA and Ward), based on three genetic distance algorithms, weighted index, unweighted index, and an index of genetic distance or dissimilarity, proposed by Smouse and Peakall. The second methodology identified groups based on simultaneous use of morphoagronomic and molecular information and extracting the genetic distance matrix by the Gower algorithm, and later applying UPGMA and Ward clustering methods. The consistency of the clustering methods was evaluated by cophenetic correlation coefficients. The significance of these coefficients was examined by the Mantel test. There was significant genetic variability among corn popcorn accesses at morphological and molecular levels. There also

was agreement between multivariate clustering techniques, mainly when using genotypic data provided by microsatellite markers. heterotic groups were identified; these were formed mainly according to the origin of each genotype and/or geographic origin. We found that there is sufficient heterosis to develop new cultivars.

**Key words:** *Zea mays*; Specialty maize; Microsatellite marker; Genetic divergence

## INTRODUCTION

Since popcorn has a narrow genetic base due to its probable evolution by selection from only flint-type common maize varieties, as described by Kantety et al. (1995), knowledge about heterotic groups is extremely important. Evaluations in diversity studies for the identification of heterotic groups have been carried out basically using phenotypic data based on morphological evaluations (Miranda et al., 2008).

With the advance of molecular techniques and the increasingly frequent use of highly informative genetic markers, identification of variability available directly at the DNA level has been widely used (Cruz et al., 2012). Despite constituting one of the pillars of plant breeding for diverse crops, the use of molecular markers for the identification of genetic variability and constitution of heterotic groups in popcorn is still not common, hampering the development of superior genotypes for most production areas.

In the U.S., Santa cruz-Varela et al. (2004) used 29 morphological markers, 18 isoenzymatic loci and 31 SSR loci to assess the relationships among representative popcorn germplasms from the Americas. They analyzed 56 populations from the U.S.A. and nine from Latin American countries, which were grouped in clusters, based on grouping and principal component analysis. By clustering the genotypes, three heterotic groups were proposed, consisting of: i) "Yellow Pearl Popcorn", which is the most important commercial group of the USA, supposedly derived from introductions of the Chilean race Curagua in England, in the 19<sup>th</sup> century; (ii) "North American Pointed Rice Popcorns", probably derived from a traditional complex of popcorn races from Latin America; and (iii) "North American" and "North American Early Popcorns", which are closely related to the "flint" maize types from northern U.S.

On the other hand, in common maize breeding programs, in which more research efforts have been invested, heterotic groups were efficiently identified by RFLP (Pinto et al., 2003), AFLP (Oliveira et al., 2004), and more recently by SSR procedures (Barata and Carena, 2006). Proposals of heterotic groups based on SSR microsatellite markers were made by Reif et al. (2003), who reported the formation of heterotic groups in agreement with the genealogy of subtropical "flint" and "dent" germplasm from the International Maize and Wheat Improvement Center (CIMMYT).

Currently, the State University of Northern Rio de Janeiro (UENF) Popcorn Germplasm Bank has dozens of popcorn accessions, including: i) populations of recurrent selection cycles of UENF, the State University of Maringá (UEM) and the Brazilian Agricultural Research Corporation – Maize and Sorghum (EMBRAPA); ii) national compounds and varieties from Brazil and abroad that originated the improved genotypes in the country, obtained in partnership exchange with EMBRAPA Maize and Sorgho, UEM, the

Federal University of Viçosa (UFV), the Agronomic Institute of Campinas (IAC) and CIMMYT; (iii) races from several Latin American countries, from CIMMYT; (iv) lines from popcorn breeding programs of the UFV, UEM and UENF; v) hybrids recommended by national and international companies, developed by the IAC and Pioneer, as well as lines developed from these commercial hybrids; and vi) single-cross hybrids between lines of the UENF breeding program. The main popcorn cultivars and hybrids recommended in the country are included in the UENF Germplasm Bank.

In this context, we attempted to identify heterotic popcorn maize groups, exploiting the genotypic representativeness of the UENF Germplasm Bank, with the objective of contributing to the development of popcorn hybrids in Brazil.

## MATERIAL AND METHODS

For the development of this study, 38 popcorn genotypes (Table 1) from different origins, maintained at the Active Popcorn Germplasm Bank of UENF, were analyzed.

Table 1. Description of popcorn genotypes of the Active Germplasm Bank of the State University of Northern Rio de Janeiro (UENF).

Genotypes	Type	Parent variety	Climatic adaptation	Institution of origin	PE
P1	Line	Hybrid Zélia	Temperate/Tropical	UEM	26.67
P2	Line	Compound CMS-42	Temperate/Tropical	UEM	19.33
P3	Line	Compound CMS-42	Temperate/Tropical	UEM	17.67
P4	Line	South American races	Temperate/Tropical	UEM	19.33
P5	Line	Hybrid Zaeli	Temperate/Tropical	UEM	26.33
P6	Line	Hybrid Zaeli	Temperate/Tropical	UEM	23.67
P7	Line	Hybrid Zaeli	Temperate/Tropical	UEM	24.67
P8	Line	Hybrid IAC112	Temperate/Tropical	UEM	22.67
P9	Line	Hybrid IAC112	Temperate/Tropical	UEM	27.33
P10	Line	Hybrid IAC112	Temperate/Tropical	UEM	25.33
L 51	Line S <sub>7</sub>	Beija-flor	Temperate/Tropical	UENF	16.00
L 52	Line S <sub>7</sub>	Beija-flor	Temperate/Tropical	UENF	15.67
L 53	Line S <sub>7</sub>	Beija-flor	Temperate/Tropical	UENF	29.33
L 54	Line S <sub>7</sub>	Beija-flor	Temperate/Tropical	UENF	19.00
L 55	Line S <sub>7</sub>	Beija-flor	Temperate/Tropical	UENF	24.67
L 59	Line S <sub>7</sub>	Beija-flor	Temperate/Tropical	UENF	19.00
L 61	Line S <sub>7</sub>	Beija-flor	Temperate/Tropical	UENF	5.00
L 63	Line S <sub>7</sub>	BRS-Angela	Tropical	UENF	21.67
L 65	Line S <sub>7</sub>	BRS-Angela	Tropical	UENF	19.00
L 66	Line S <sub>7</sub>	BRS-Angela	Tropical	UENF	22.33
L 69	Line S <sub>7</sub>	BRS-Angela	Tropical	UENF	20.33
L 70	Line S <sub>7</sub>	BRS-Angela	Tropical	UENF	21.00
L 71	Line S <sub>7</sub>	BRS-Angela	Tropical	UENF	15.67
L 74	Line S <sub>7</sub>	BRS-Angela	Tropical	UENF	19.33
L 75	Line S <sub>7</sub>	BRS-Angela	Tropical	UENF	18.33
L 76	Line S <sub>7</sub>	Viçosa	Temperate/Tropical	UENF	23.33
L 77	Line S <sub>7</sub>	Beija-flor	Temperate/Tropical	UENF	23.33
L 80	Line S <sub>7</sub>	Viçosa	Temperate/Tropical	UENF	6.33
L 88	Line S <sub>7</sub>	Viçosa	Temperate/Tropical	UENF	15.33
BOYA 462	OPP	SAR (Bolivia)	Temperate/Tropical	CIMMYT	8.00
URUG 298 Roxo	OPP	SAR (Uruguay)	Temperate/Tropical	CIMMYT	20.67
URUG 298 Amarelo	OPP	SAR (Uruguay)	Temperate/Tropical	CIMMYT	23.00
BOZM 260	OPP	SAR (Bolivia)	Temperate/Tropical	CIMMYT	9.33
CHZM 13 0136	OPP	SAR (Chile)	Temperate/Tropical	CIMMYT	12.67
ARZM 05 083	OPP	SAR (Argentina)	Temperate/Tropical	CIMMYT	5.00
ARZM 07 049	OPP	SAR (Argentina)	Temperate/Tropical	CIMMYT	11.67
ARZM 13 050	OPP	SAR (Argentina)	Temperate/Tropical	CIMMYT	15.33
PARA 172	OPP	SAR (Paraguay)	Temperate/Tropical	CIMMYT	9.33

IAC – Agronomic Institute of Campinas; UEM – State University of Maringá; CIMMYT – International Maize and Wheat Improvement Center; OPP – Open-pollinated population; SAR – South-American races – PE – Popping expansion (mg.mL<sup>-1</sup>);

The experimental design consisted of randomized blocks, with four replications within a growing season. Each plot consisted of one 5-m long row, at a spacing of 0.60 m between rows and 0.30 m between plants.

The fields were planted in October 2015 at the Experimental Station of UENF at the Antônio Sarlo State College of Agriculture, in the county of Campos dos Goytacazes, in the northern region of the state of Rio de Janeiro, for the phenotypic evaluation of the genotypes. Three seeds per planting hole were sown to ensure germination in all plots, at a depth of 0.05 m. At 21 days after emergence, thinning was performed, leaving the plant with greatest vigor per hole, resulting in a final population of approximately 55,555 plants.ha<sup>-1</sup>. Fertilization consisted of 800 kg.ha<sup>-1</sup> of mineral NPK fertilizer 04-14-08 at planting, a first sidedressing of 300 kg.ha<sup>-1</sup> of 20-00-20 (NPK) at 15 days after sowing, and a second sidedressing of 200 kg.ha<sup>-1</sup> of nitrogen, in the form of urea, at 30 days after emergence. All fertilizer quantities were determined as recommended, based on the chemical analysis of the soil of the experimental cultivation area.

Ten morphoagronomic characteristics of popcorn were analyzed: mean plant height (PH) - averaging data of five competitive plants per plot, measured from the soil level to the insertion point of the flag leaf, expressed in meters (m); mean ear height (EH) - averaging data of five competitive plants per plot, measured from the soil level to the insertion point of the first ear, in meters (m); mean stem diameter (StD) - averaging data of five competitive plants per plot, by measuring the stem diameter of the plant at 0.50 m above ground level, with a digital caliper, in centimeters (cm); mean tassel length (TL) - averaging data of five tassels per plot, by measuring the tassel length from the insertion point of the first branch to the apex of the main branch, in centimeters (cm); mean number of tassel branches (NRP) - obtained by the mean number of tassel branches of five plants per plot; days to flowering (FLOW) - defined by the number of days from sowing to the flowering peak in the plot, i.e., when simultaneous male and female flowering was observed in 50% of the plants of a plot; mean 100-seed weight (W100) - defined by the average weight of 100 seeds of two random replications per plot, weighed with a precision scale, in grams (g); mean ear length (EL) - obtained by averaging the length of five random ears per plot, measured with a ruler, in centimeters (cm); mean number of kernel rows per ear (NRE) - defined by the number of rows of kernels in five random ears per plot; and mean number of kernels per row on the ears (NKR) - obtained by counting the number of kernels per row in five random ears per plot.

Statistical analyses of the variables for the morphoagronomic traits were performed using software Genes (Cruz, 2016) in the analyses required for the study of identification of heterotic groups in the studied germplasm.

Analysis of variance for the randomized block design was performed, according to the following statistical model:

$$\gamma_{ij} = \mu + g_i + b_j + e_{ij} \quad (1)$$

where  $\gamma_{ij}$  is the value observed in the plot that received treatment  $i$  in block  $j$ ;  $\mu$  is the overall mean of the experiment;  $g_i$  is the effect of genotype  $i$ ,  $i = 1, 2, \dots, 38$ , where NID  $(0, \sigma_g^2)$ ;  $b_j$  is the effect of block  $j$ ,  $j = 1, 2, \dots, 4$ , where NID  $(0, \sigma_b^2)$ ; and  $e_{ij}$  is the experimental error associated with observation  $\gamma_{ij}$ , where NID  $(0, \sigma^2)$ .

For the molecular analyses, two seeds per genotype were sown in each cell (225 cm<sup>3</sup>) of a styrofoam tray containing commercial substrate, from which a sample consisting of young leaves of each genotype was collected 14 days after sowing.

The DNA was extracted according to a protocol proposed by Doyle and Doyle (1990), modified for the species. The step of selection of genetic microsatellite markers (EST-SSR) was initiated with a total of 193 primers. Among this primer set, 15 pairs were polymorphic and thus selected for resolution in a capillary electrophoresis fragment analyzer (Advanced Analytical). Coding of the genotype data matrix was carried out with the aid of Power Marker software (Liu and Muse, 2005). Genetic diversity was estimated based on the number of alleles (Na), the number of effective alleles (Ne) and the polymorphic information content (PIC).

Based on Bayesian clustering algorithms, according to the expected Hardy-Weinberg equilibrium and the absence of linkage disequilibrium between the analyzed loci within each population, the number of groups among the genotypes studied was determined with the software Structure (Pritchard et al., 2000).

To perform the analyses, the criteria "no admixture model" and correlated frequencies were used; after a burn-in period of 10,000, the MCMC (Markov Chain Monte Carlo) algorithm was run to obtain a sample of size of 50,000 repetitions, with a preset number of groups (*k*) ranging from 1 to 10. The maximum likelihood of adhesion between groups was determined using a threshold of 0.75. The proposal of Evanno et al. (2005) was used to infer the real value of *K* in the population with SSR-EST markers, since the likelihood distribution *L* (*k*) does not represent the true *K* value. For this inference, software STRUCTURE HARVESTER (Earl and Von, 2016) was used to process the STRUCTURE result files.

The genotypic data were analyzed using the software Genes (Cruz, 2016); two cluster analyses, UPGMA and Ward, were performed based on three algorithms of genetic distances, namely:

Weighted index:

$$S_{ii'} = \frac{1}{2} \sum_{j=1}^L p_j c_j \quad (2)$$

where  $p_j = \frac{a_j}{A}$  is the weight associated with locus *j*, determined by *a<sub>j</sub>*, which is the total number of alleles at locus *j*, and *A* the total number of alleles studied.

Unweighted index:

$$S_{ii'} = \frac{1}{2L} \sum_{j=1}^L c_j \quad (3)$$

where *L* is the total number of loci studied; and *c<sub>j</sub>* the number of common alleles between pairs of accessions *i* and *i'*.

Index *d*<sup>2</sup> of Smouse and Peakall (Smouse and Peakall, 1999):

$$d_{ii'}^2 = \frac{1}{2L} \sum_{j=1}^L \sum_{k=1}^{a_j} (y_{ijk} - y_{i'jk})^2 \quad (4)$$

where *y<sub>ijk</sub>* indicates the number of alleles *k* at locus *j* of plant *i*.

The morphoagronomic and molecular variables were analyzed simultaneously by the procedure proposed by Franco et al. (1998), named Ward-Modified Location Model (MLM), using software SAS (SAS Institute, 2003), based on the Gower algorithm (1971), expressed by:

$$S_{ijk} = \frac{\sum_{k=1}^p W_{ijk} \cdot S_{ijk}}{\sum_{k=1}^p W_{ijk}} \quad (5)$$

where  $k$  is the number of variables ( $k = 1, 2, \dots, p$ );  $i$  and  $j$ : two individuals that represent the accession;  $W_{ijk}$  is the weight given to comparison  $ijk$ , assigning value 1 to valid comparisons and value 0 to invalid comparisons (when the value of the variable is absent in one or both individuals); and  $S_{ijk}$ : contribution of variable  $k$  to the similarity between individuals  $i$  and  $j$ , with values between 0 and 1. For a qualitative (nominal) variable, if the value of variable  $k$  is the same for both individuals,  $i$  and  $j$ , then  $S_{ijk} = 1$ , otherwise,  $S_{ijk} = 0$ . For a quantitative (continuous) variable,  $S_{ijk} = \frac{1 - |x_{ik} - x_{jk}|}{R_k}$  where  $x_{ik}$  and  $x_{jk}$  are the values of variable  $k$  for individuals  $i$  and  $j$ , respectively, and  $R_k$  is the interval (maximum minus minimum value) of variable  $k$  in the sample. Division by  $R_k$  eliminates the differences between scales of the variables, producing a value between 0 and 1 and equal weights.

After estimating the genetic distance matrices, clustering was performed by hierarchical methods of mean linkage between groups or UPGMA and Ward's minimum variance (Ward, 1963), using software R (R Core Team, 2013). The cut-off point (PC) for determining the number of groups per cluster was defined according to the expression proposed by Mojena (1977):

$$PC = \hat{\mu}_\alpha + k \hat{\sigma}_\alpha \quad (6)$$

where:  $PC$  is the cutoff point of the dendrogram determined for the identification of the number of groups;  $\hat{\mu}_\alpha$  is the mean of the cluster fusion points;  $\hat{\sigma}_\alpha$  is the standard deviation of the fusion points between two dendrogram groups; and  $k$  is an arbitrarily established constant, where  $k = 1.25$ , used as proposed by Milligan and Cooper (1985).

The consistency of the clustering methods was assessed by the cophenetic correlation coefficients (CCC), where the significance of CCC was examined by the Mantel test (Mantel, 1967).

## RESULTS AND DISCUSSION

Analysis of variance results showed that only three among the 10 morphoagronomic characteristics - mean plant height (PH), mean ear height (EH) and mean stem diameter (StD) had no significant differences between the 38 popcorn germplasms (Table 2).

Table 2. Summary of variance analysis with the respective mean squares and degrees of freedom (DF), estimates of experimental variation coefficients (CV%), and means for the evaluated traits, in a randomized block design. Campos dos Goytacazes, RJ.

SV	DF	Mean Square <sup>1/</sup>									
		PH	EH	StD	TL	NRP	FLOW	W100	EL	NRE	NKR
Blocks	3	0.92	0.81	172.35	148.67	76.02	179.71	47.57	29.62	20.1	69.61
Genotypes	37	0.032 <sup>ns</sup>	0.04 <sup>ns</sup>	2.29 <sup>ns</sup>	81.94 <sup>**</sup>	151.19 <sup>**</sup>	96.86 <sup>**</sup>	18.40 <sup>**</sup>	16.15 <sup>**</sup>	9.64 <sup>**</sup>	146.97 <sup>**</sup>
Error	111	0.053	0.044	26.38	102.05	65.21	169.67	46.97	59.51	25.75	223.38
Mean		1.56	0.93	16.14	34.96	18.74	58.57	7.75	11.02	12.54	22.27
CV (%)		14.8	22.58	10.06	9.14	13.63	7.03	27.96	22.13	12.8	21.22

<sup>1/</sup>PH = Mean plant height (m); EH = Mean ear height (m); StD = Mean stem diameter (cm); TL = Mean tassel length (cm); NRP = Mean number of tassel branches; FLOW = Number of days to flowering; W100 = Mean 100-seed weight (g); EL = ear length (cm); NRE = Number of kernel rows; NKR = Number of kernels per row; <sup>ns</sup> = Not significant by the F test at 1 and 5% probability; <sup>\*\*</sup> = Significant by the F test at 1% probability

Contrasting results were found by Pandit et al. (2016), who observed the most variation in plant height and ear traits, based on the estimation of divergence among 20 maize genotypes. However, they also found differences among the genotypes in the trait days to flowering.

In general, the values of the coefficients of experimental variation for the traits were satisfactory, with low values for mean tassel length and days to flowering, as well as mean number of tassel branches and mean number of kernel rows per year, aside from high estimates for mean 100-seed weight, ear length, and number of kernels per rows, according to the classification scale of coefficients of variation proposed by Pimentel Gomes (1990).

The significant differences among the germplasms detected in seven variables showed the possibility of morphology-based diversity studies of these genotypes. Consequently, the means of the characteristics were used, eliminating those without significant differences.

Table 3 shows the 15 SSR-EST primers, with a mean annealing temperature for each locus varying between 59 and 65°C. Per locus, 2-5 alleles were identified, and 42 alleles for the 15 loci, by Fragment Analyzer (Advanced Analytical), resulting in an overall mean of 2.88 alleles per locus.

Table 3. SSR-ESTs primers and the respective mean annealing temperatures (AT°C), number of alleles (Na), number of effective alleles (Ne) and polymorphic information content (PIC).

<i>Locus</i>	AT°C	Na	Ne	PIC
umc1118	62	2	1.85	0.35
ole1	62	3	2.20	0.45
umc1252	63	3	1.52	0.30
umc1139	63	2	1.82	0.34
umc1982	63	3	1.88	0.43
phi96100	59	5	3.39	0.67
umc1073	64	2	1.98	0.37
umc2319	64	3	1.38	0.25
glb1	62	3	1.99	0.42
umc1241	63	2	1.73	0.33
umc2112	63	3	2.98	0.59
umc1130	63	2	1.82	0.36
bip2	64	3	2.09	0.45
umc1515	62	3	2.00	0.40
y1	65	3	2.94	0.59
<b>General mean</b>		2.80	2.10	0.42
<b>Total</b>		42		

AT: annealing temperature, Na: observed number of alleles; Ne: effective number of alleles, PIC: polymorphism information content.

The quantification of the genetic polymorphism of each locus of the studied population and the indication of the level of information and potential of each locus to detect differences between the genotypes, based on their genetic relations, are defined by the polymorphic information content (PIC) (Botstein et al., 1980; Rajendran et al., 2014). According to the classification recommended by Botstein et al. (1980), a locus is considered highly informative when the PIC value is  $> 0.5$ , moderately informative when  $0.5 > \text{PIC} > 0.25$ , and uninformative when  $< 0.25$ .

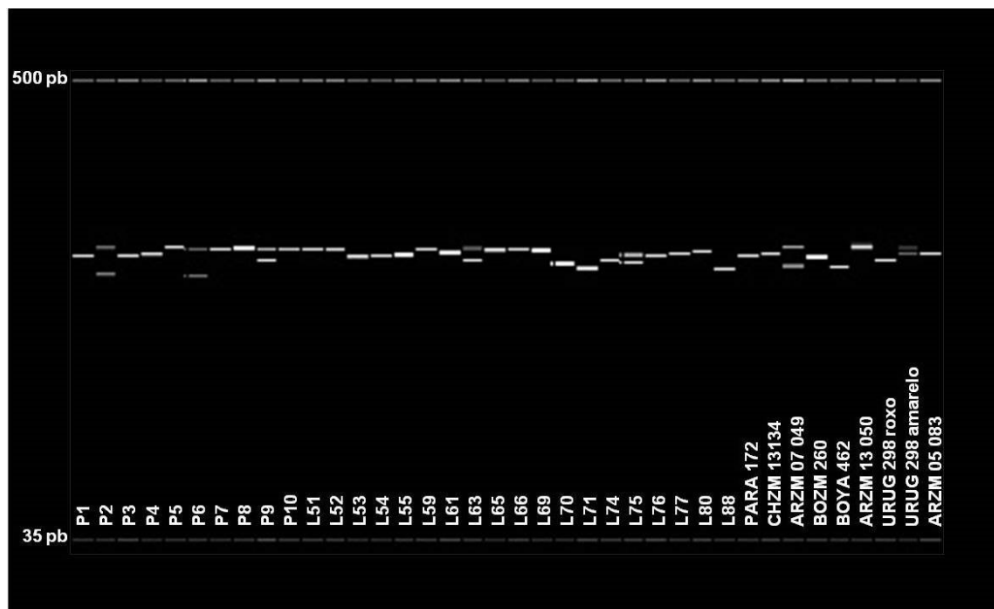
Three of the 15 microsatellite loci were highly polymorphic, 12 loci were considered moderately polymorphic, and all microsatellite loci were informative. The overall mean PIC was 0.42, indicating a moderate polymorphism level. The highest and lowest PIC values (0.67 and 0.25, respectively) were observed for the *phi96100* and *umc2319* loci. Similar results were observed by Pena et al. (2016), who reported a mean



PIC value of 0.40, in an analysis of  $S_3$  popcorn progenies. The results of this study are considered satisfactory, since Zhang et al. (2014) reported that several studies involving this type of codominant marker detected low to medium levels of polymorphism. The cause of the low PIC value was ascribed to the selection that possibly affected the variation in conserved regions of the SSR-ESTs markers (Scott et al., 2000).

The highest PIC was reported by Silva et al. (2015), in an SSR-EST-marker analysis of popcorn genotypes (PIC = 0.53). However, the presence of a locus with practically null polymorphic information (0.06) was observed, reinforcing the quality of the results obtained in this study, since the locus with the lowest PIC value was much higher than this value (umc2319; PIC = 0.25). According to Romero-Severson et al. (2001), the PIC weighs the number of alleles and the relative frequency of these alleles; thus, lower PIC values indicate a high frequency of one or two alleles.

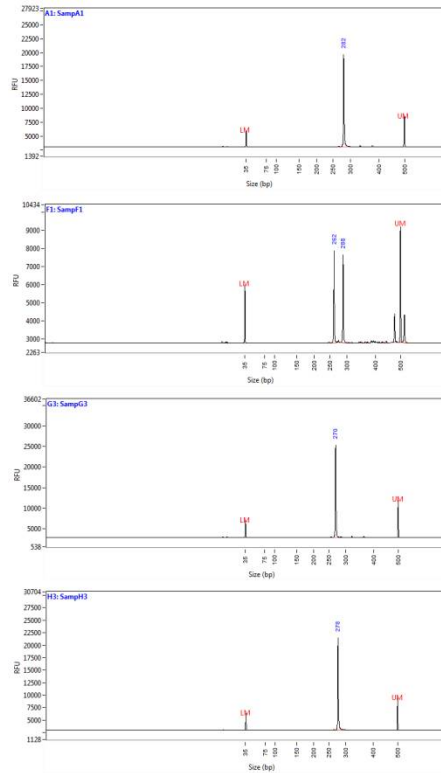
Figure 1 shows the virtual gel analysis with Fragment Analyzer of locus *phi96100* for the 38 analyzed popcorn accessions, in which, based on the resulting band patterns, five alleles were identified for this locus.



**Figure 1.** Virtual gel analysis with Fragment Analyzer, of locus *phi96100* for the 38 popcorn genotypes.

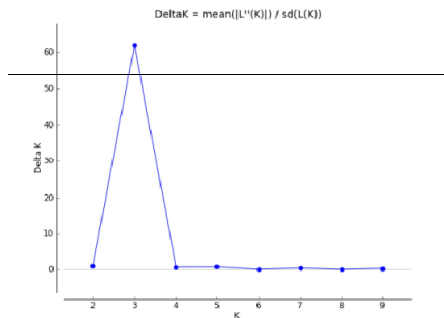
The diagram of locus *phi96100* (Figure 2) shows four randomly selected genotypes, three of which were homozygous and one heterozygous, as an illustration of the five possible alleles for this locus, evidenced by the presence of peaks according to the scale of base pairs.





**Figure 2.** Diagram of locus *phi96100* in homozygous and heterozygous genotypes demonstrating its five allelic forms (allele 1: 262bp, allele 2: 270bp, allele 3: 278bp, allele 4: 282bp, and allele 5: 288bp).

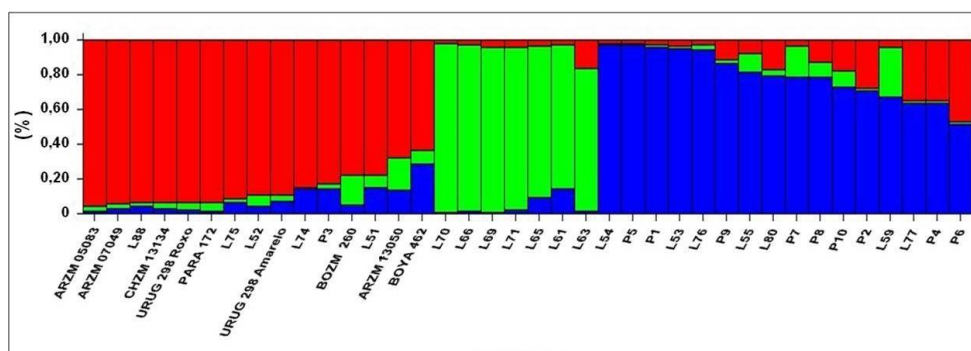
As an alternative for genotype classification in different divergence groups, the software Structure (Pritchard et al., 2000) was used for the analysis of molecular data. The 42 alleles were used to infer the genetic groups for the 38 genotypes, where the Bayesian approach demonstrated a tendency to form three groups, when the  $\Delta K$  value was highest, with  $K = 3$ , identified by the peak shown in Figure 3, confirmed by the method of Evanno et al. (2005).



**Figure 3.** Delta K values ( $\Delta K$ ) for the respective group numbers ( $K$ ).

The  $\Delta K$  curve reaches practically zero at the other points, except for the peak observed when  $K=3$  (Figure 3). This determination of the group number varies according to the study population. In this regard, Silva et al. (2015) determined a  $K$  value of 2, in a microsatellite analysis of 31 popcorn accessions. Four groups were presented by Santos-Schneider (2013), in an evaluation of 18 popcorn lines with the same marker type.

The distribution of the genotypes in the divergent groups defined by the Bayesian model was determined by the probability of adhesion at a probability level of  $> 0.75$ . However, the probability of adhesion of some genotypes was  $< 0.75$ , so these were included in the groups of higher probability of superior adhesion (Figure 4). It was observed that all genotypes allocated in group II (green) had high adhesion to this group (namely: L70, L66, L69, L71, L65, L61, and L63). Only for some genotypes of the groups I (red), i.e., ARZM 13 050 and BOYA 462 and III (blue) - L77, P4 and P6 – the probability values were not  $> 0.75$ , but all of these had values  $> 0.60$ , considered satisfactory for allocation in these groups with higher adhesion probability.



**Figure 4.** Bayesian-based clustering of 38 popcorn genotypes with the identified groups ( $K = 3$ ): Group I - red; Group II - green; Group III - blue.

The open-pollinated populations of the South American countries other than Brazil (Argentina, Bolivia, Chile, Paraguay and Uruguay) were all allocated in the same group, together with some lines of different origins, indicating the strong similarity between the South American races. It can be assumed that the alleles of these populations refer to the same gene pool or, at least, to nearby gene pools. This confirms the reliability of molecular markers for the identification of divergent popcorn groups, constituting a tool of interest for future research along this line.

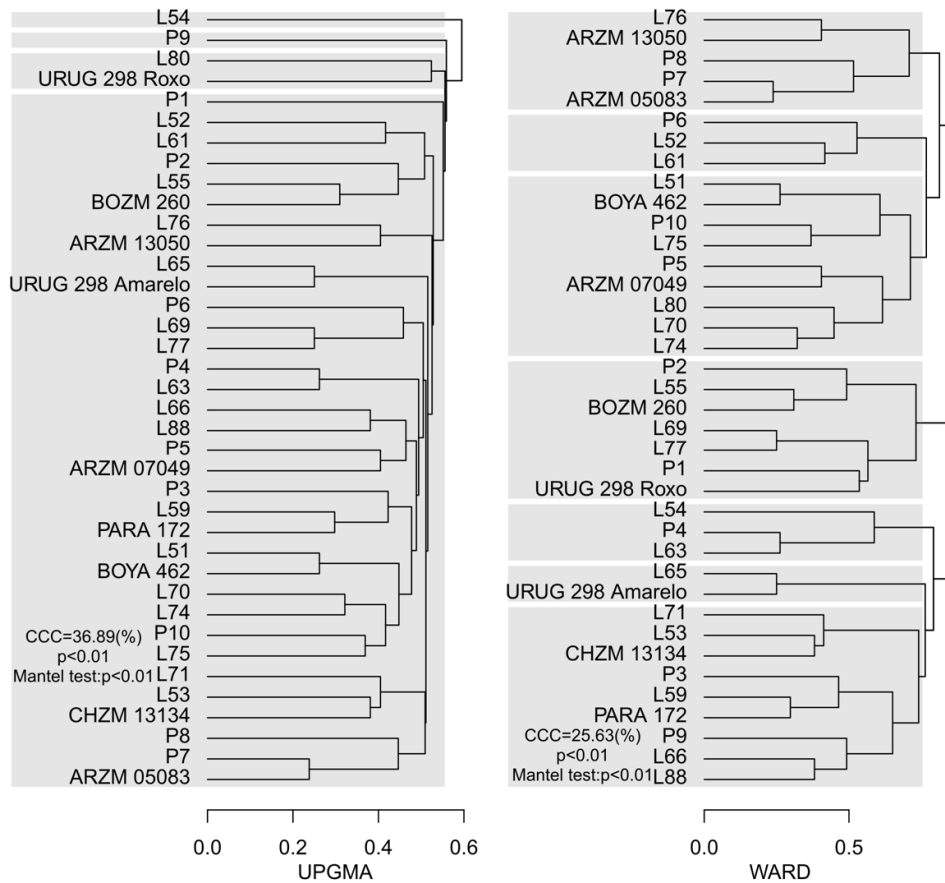
All 33 genotypes in group II were derived from the popcorn variety BRS Angela, 28 with predominantly white and five with yellow endosperm, maintained at the Active Germplasm Bank of Embrapa Maize and Sorghum, denominated material of tropical origin (Pacheco et al., 2004). Group III, the largest one, was composed of 16 lines, all derived from national germplasms, indicating a greater genetic proximity between these genotypes.

According to Santos-Schneider (2013), the Bayesian approach applied in this study to determine divergent groups can be efficient as an orientation for crosses of popcorn lines, to develop hybrids by exploiting heterosis.

In an analysis of different popcorn populations, Bracco et al. (2009) pointed out that the Bayesian-based technique for group determination can be efficient in the establishment of heterotic groups and consequently the optimization of strategies for the exploitation of genetic resources, aiming at the breeding of superior popcorn genotypes.

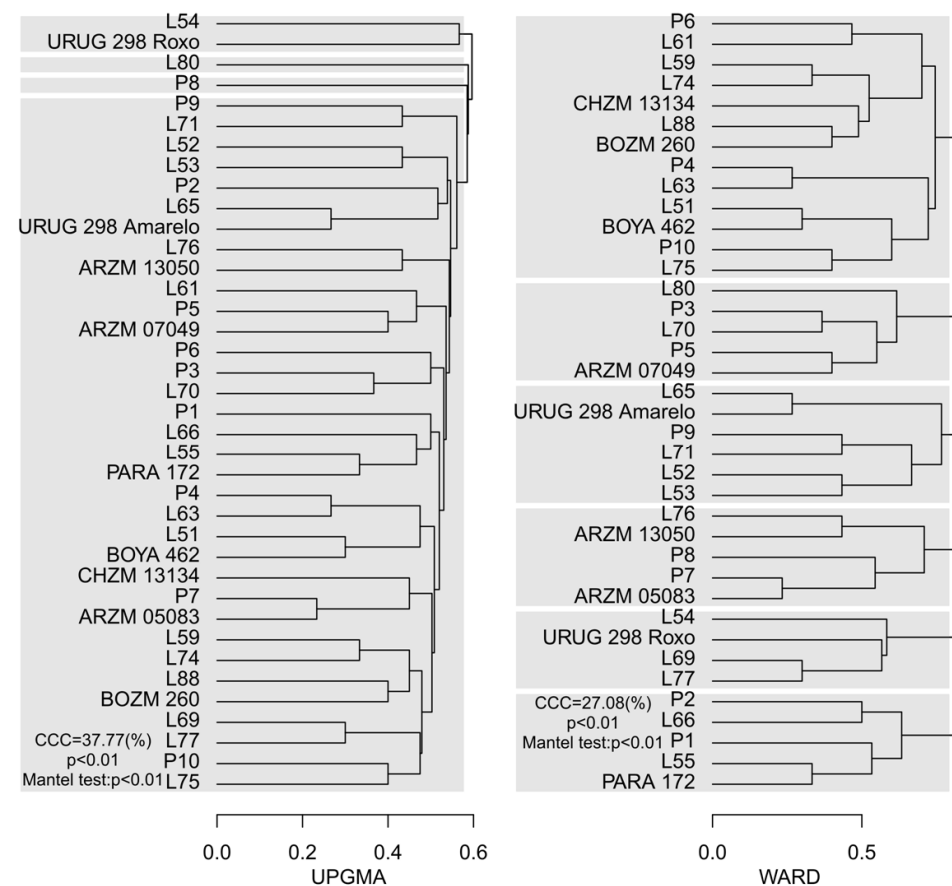
Following up the above analysis, two cluster analyses (UPGMA and Ward) were performed based on the Weighted Index and Unweighted Index, as well as the Smouse and Peakall  $d^2$  index.

Clustering by the UPGMA hierarchical method, based on genetic dissimilarity matrices using the weighted as well as the unweighted index, only separated four genotypes from a large group composed of 34 genotypes (Figure 5 and 6). Also, the CCC obtained for these two clusters was relatively low (36.89% for the weighted and 37.77% for the unweighted index), revealing inconsistency of this clustering for this set of genotypes.



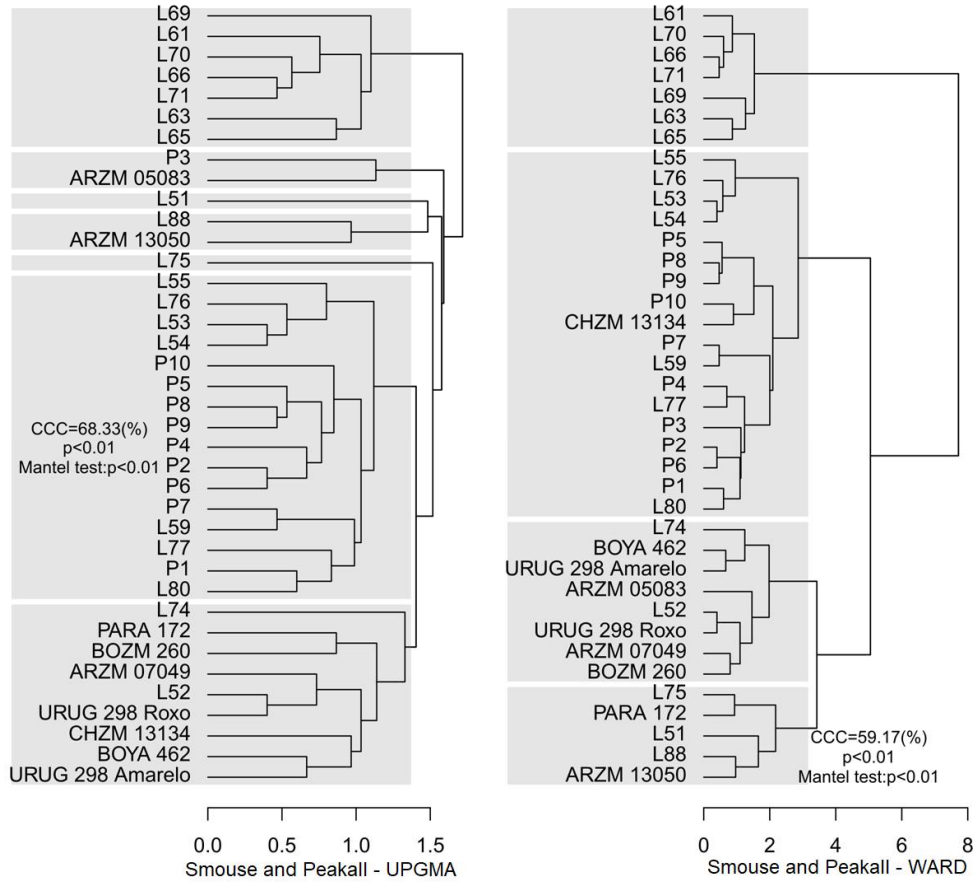
**Figure 5.** Dendrograms obtained by the hierarchical methods UPGMA and Ward, based on the Weighted Index, for 38 popcorn genotypes, with a cutoff point determined as proposed by Mojena (1977), and the cophenetic correlation coefficient (CCC).

Analyzing the results obtained by Ward's clustering, based on distance matrices using the weighted as well as the unweighted indices, a greater distribution of genotypes was observed; seven groups were formed by the weighted, and six by the unweighted index (Figures 5 and 6). The low values of the cophenetic correlation coefficients of these clustering methods (25.63 and 27.08%) indicate a limited reliability of these distances for genotype clustering, making it impossible to use these algorithms for conclusive clustering of heterotic groups.



**Figure 6.** Dendrograms obtained by UPGMA and Ward's method of hierarchical cluster analysis based on the Unweighted Index, for 38 popcorn genotypes using a cutoff point proposed by Mojena (1977) and the cophenetic correlation coefficient (CCC).

The results obtained by hierarchical clustering based on molecular markers, considering the dissimilarity measure proposed by Smouse and Peakall (1999), were more consistent than those obtained with the weighted and unweighted indices, providing satisfactory cophenetic correlation coefficients (68.33 and 59.17% for UPGMA and Ward, respectively, Figure 7).



**Figure 7.** Dendrograms obtained by UPGMA and Ward's hierarchical clustering based on the Smouse and Peakall Index, for 38 popcorn genotypes using a cutoff point proposed by Mojena (1977) and the cophenetic correlation coefficient (CCC).

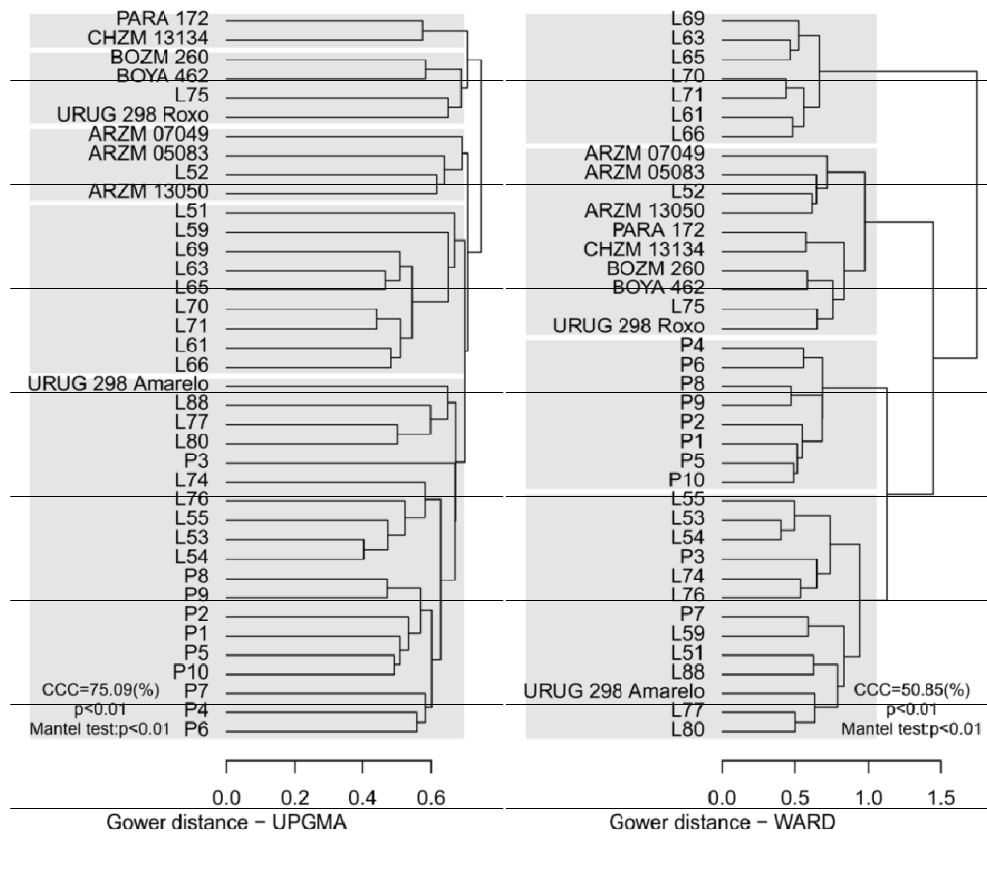
Based on the clustering using the distance measure of Smouse and Peakall (1999), seven groups were formed by the UPGMA method and four by Ward's method. In both clustering methods, a group consisting of seven lines was formed, confirming the genetic similarity between them and dissimilarity with the other genotypes. Interestingly, these lines were the same as the constituents of group II in the Bayesian-based analyses (Figure 3).

The characterization of groups formed by white-endosperm germplasm was presented by Amaral Júnior et al. (2011), who analyzed the genetic variability of popcorn in comparison to other maize types, based on ISSR markers. They identified the formation of a group composed exclusively of popcorn genotypes with white endosperm. This may contribute to a possible identification of heterotic popcorn groups based on kernel color.

The clustering techniques used, based on dissimilarity measures proposed by Smouse and Peakall (1999), showed a strong similarity between popcorn populations, since these were generally related in distinct groups of the descendant lines of national germplasm. This may influence the future development of superior genotypes by directed

crosses between parents derived from these divergent groups, circumventing the difficulty highlighted by Miranda et al. (2008), of limited heterosis and genetic variability for popping expansion in the Brazilian popcorn populations.

Morphoagronomic and molecular characteristics were analyzed simultaneously for the application of UPGMA and Ward's hierarchical clustering, using the genetic distance matrix provided by the Gower algorithm for this purpose (Figure 8).



**Figure 8.** Dendrograms obtained by UPGMA and Ward's hierarchical cluster analysis based on genetic distances of the Gower algorithm considering morphoagronomic and molecular characteristics, for 38 popcorn genotypes, using a cutoff point proposed by Mojena (1977) and the cophenetic correlation coefficient (CCC).

With a hybrid matrix of quantitative and qualitative variables, the 38 genotypes were separated into five groups by the UPGMA method and divided into four groups by Ward's clustering. High similarity with the analyzed groupings based on the dissimilarity measures proposed by Smouse and Peakall was observed (Figure 7). The formation of one group with the popcorn lines L61, L63, L65, L66, L69, L70, and L71, derived from the variety BRS-Angela (Pacheco, 2004) is noteworthy, revealing a strong genetic proximity

between these lines, which have an obvious morphological characteristic, the white kernel color.

The open-pollinated popcorn populations that we analyzed were arranged in three groups with similarity to each other by the UPGMA method (Figure 8). By Ward's method, the populations were all grouped into a single group, confirming the similarity between them. An exception in both clusters was the germplasm URUG 298 Amarelo, which was allocated in groups together with Brazilian lines.

The distribution of these populations in similar groups corroborates the data of Silva et al. (2015), analyzing the genetic diversity of a corn-popcorn germplasm bank, in which the use of microsatellite markers allowed the allocation of these same populations in a single group.

Two Brazilian lines (L52 and L75) were joined with populations from other countries by both clustering methods, isolated from the national lines, which may influence the direction of possible directed crosses, with the lines clustered in different groups for the breeding of superior hybrids.

Based on the molecular groupings by the distance of Smouse and Peakall and the clustering resulting from combined morphoagronomic and molecular information, the formation of three divergent groups in popcorn can be assumed, highlighting the formation of a group composed of the lines that present as a particularity the color of the kernel. In the other groups, the characteristics corresponding to the morphology of kernel and ears are highlighted in the contribution to the genotypic differentiation.

## CONCLUSIONS

There was genetic variability between corn popcorn accesses at morphological and molecular levels. There was agreement between multivariate clustering techniques, mainly when using genotypic data provided by microsatellite markers. Heterotic groups were identified; they were formed based on the varieties of origin of each genotype and/or geographic origin. Therefore, heterosis can be exploited to develop new cultivars.

## CONFLICTS OF INTEREST

The authors declare no conflict of interest

## ACKNOWLEDGMENTS

The authors are thankful to the Foundation for Research Support of the State of Rio de Janeiro (FAPERJ), the National Council for Scientific and Technological Development (CNPq), for funding the project, and the Coordination of Improvement of Higher Education Personnel (CAPES) for a doctoral scholarship for the first author.

## REFERENCES

- Amaral Júnior AT, de Oliveira EC, Gonçalves LSA, Scapim CA, et al. (2011). Assessment of genetic diversity among maize accessions using inter simple sequence repeats (ISSR) markers. *Afr. J. Biotechnol.* 10: 15462-15469.
- Barata C and Carena M (2006). Classification of North Dakota maize inbred lines into heterotic groups base on molecular and testcross data. *Euphytica.* 151: 339-349.



- Botstein D, White RL, Skolnick M, Davis RW, (1980). Construction of a genetic linkage map in man restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32: 314-331.
- Bracco M, Lia VV, Gottlieb AM, Hernández JC, et al. (2009). Genetic diversity in maize landraces from indigenous settlements of Northeastern Argentina. *Genetica.* 135: 39-49.
- Cruz CD, Regazzi AJ and Carneiro PCS (2012). Modelos biométricos aplicados ao melhoramento genético. 4th edn. UFV, Viçosa.
- Cruz CD, (2016). Genes Software – extended and integrated with the R, Matlab and Selegen. *Acta Sci. Agron.* 38: 547-552.
- Doyle JJ, Doyle JL, (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin.* 19:11-15.
- Earl DA, Von HBM, (2016). Structure Harvest: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* <http://link.springer.com/article/10.1007%2fs12686-011-09548-7>.
- Evanno G, Regnaut S, Goudet J, (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611-2620.
- Franco J, Crossa J, Villaseñor J, Taba S, et al. (1998). Classifying genetic resources by categorical and continuous variables. *Crop Sci.* 38: 1688-1696.
- Gower JC, (1971). A general coefficient of similarity and some of its properties. *Biometrics.* 27: 857-874.
- Kantety RV, Zeng X, Bennetzen J, Zehr BE, (1995). Assessment of genetic diversity in dent and popcorn (*Zea mays* L.) inbred lines using inter-simple sequence repeat (ISSR) amplification. *Mol. Breed.* 1: 365-373.
- Liu K and Muse SV (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics.* 21: 2128–2129.
- Mantel N, (1967). The detection of disease clustering and a generalized regression approach. *Cancer research.* 27: 209-220.
- Milligan GW and Cooper MC (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika.* 50: 159-179.
- Miranda GV, Souza LV, Galvão JCC, Guimarães LJM, et al. (2008). Genetic variability and heterotic groups of Brazilian popcorn populations. *Euphytica.* 162: 431-440.
- Mojena R (1977). Hierarchical grouping methods and stopping rules: an evaluation. *Comput. J.* 20: 359-363.
- 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.
- Pacheco CAP, Gama EEG, Parentoni, SN, Santos M, et al. (2004). Variedade de milho pipoca BRS Angela: novo ciclo de seleção. *Embrapa Milho e Sorgo.* Comunicadotécnico.
- Pandit M, Chakraborty M, Haider ZA, Pande A, et al. (2016). Genetic diversity assay of maize (*Zea mays* L.) inbreds based on morphometric traits and SSR markers. *Afr. J. Agric. Res.* 11: 2118-2128.
- Pena GF, Amaral Junior AT, Ribeiro RM, et al. (2016). Inference of genetic diversity in popcorn S<sub>3</sub> progenies. *Genet. Mol. Res.* 15: gmr.15028456.
- Pimentel Gomes F (1990). Curso de estatística experimental. 12th edn. Nobel, São Paulo.
- Pinto RMC, Souza Júnior CL, 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.
- Pritchard JK, Stephens M, Donnell YP, (2000). Inference of population structure using multilocus genotype data. *Genetics.* 155: 945-959.
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available at [<https://www.Rproject.org/>].
- Rajendran A, Muthiah A, Joel J, Shanmugasundaram, P, et al. (2014). Heterotic grouping and patterning of quality protein maize inbreds based on genetic and molecular marker studies. *Turk. J. Biol.* 38: 10-20.
- 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.
- Romero-Severson J, Smith JSC, Ziegler J, Hauser J, Joe L, et al. (2001). Pedigree analysis and haplotype sharing within diverse groups of *Zea mays* L. inbreds. *Theor. Appl. Genet.* 103: 567–574.
- Santacruz-Varela A, Widrelechner MP, Ziegler KE, Alvarado RJ, et al. (2004). Phylogenetic relationships among North American popcorns and their evolutionary links to Mexican and South American popcorns. *Crop Sci.* 44: 456-1467.
- Santos-Schneider JF (2013). Proposta de formação de grupos heteróticos e linhagens de milho-pipoca (*Zea mays* L.) utilizando marcadores microsatélites. Doctoral thesis, Universidade Estadual de Maringá, UEM, Maringá.
- SAS Institute Inc (2003). Statistical Analysis System user's guide. Version 9.1. edn. Cary, NC, USA.
- Scott KD, Egger P, Seaton G, et al. (2000). Analysis of SSRs derived from grape ESTs. *Theor. Appl. Genet.* 100: 723-726.
- Silva TA, Cantagalli LB, Saavedra J, Lopes AD, et al. (2015). Population structure and genetic diversity of Brazilian popcorn germplasm inferred by microsatellite markers. *Electron. J. Biotechnol.* 18: 181-187.
- Smouse PE and Peakall R (1999). Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity.* 82: 561-573.

- Ward JH (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58: 236-244.
- Zhang M, Mao W, Zhang G, Wu F (2014). Development and characterization of polymorphic EST-SSR and genomic SSR markers for Tibetan annual wild barley. *PLoS ONE* 9: e94881.