RESEARCH

CHILEANJAR
CHILEAN JOURNAL of AGRICULTURAL RESEARCH

Bayesian analysis of the genetic structure of a Brazilian popcorn germplasm using data from simple sequence repeats (SSR)

Javier Saavedra[1], Tereza Aparecida Silva[2], Freddy Mora[3*], and Carlos Alberto Scapim[2]

Several studies have confirmed that popcorn (*Zea mays* L. var. *everta*) has a narrow genetic basis, which affects the quality of breeding programs. In this study, we present a genetic characterization of 420 individuals representing 28 popcorn populations from Brazilian germplasm banks. All individuals were genotyped using 11 microsatellite markers from the Maize Genetics and Genomics Database. A Bayesian clustering approach via Monte Carlo Markov chains was performed to examine the genetic differentiation (F_{ST} values) among different clusters. The results indicate the existence of three distinct and strongly differentiated genetic groups ($K = 3$). Moreover, the F_{ST} values (calculated among clusters) were significantly different according to Bayesian credible intervals of the posterior F_{ST} values. The estimates of posterior mean (and 95% credible interval) of the F_{ST} values were 0.086 (0.04-0.14), 0.49 (0.376-0.624) and 0.243 (0.173-0.324) for clusters 1, 2, and 3, respectively. Clusters 1 and 3 showed a high level of genetic diversity in terms of expected heterozygosity and number of alleles, indicating their potential for broadening the genetic basis of popcorn in future breeding programs. Additionally, the 11 microsatellites were informative and presented a suitable number of alleles for determining parameters related to genetic diversity and genetic structure. This information is important for increasing our knowledge regarding genetic relationships, for the identification of heterotic groups, and for developing strategies of gene introgression in popcorn.

Key words: Genetic differentiation, microsatellite markers, Bayesian clustering, Monte Carlo Markov chains.

INTRODUCTION

Popcorn (*Zea mays* L. var. *everta*) is an extreme form of flint corn (*Z. mays* L. var. *indurata*) that has a very hard corneous endosperm with only a small portion of soft starch and is characterized by its popping ability; the kernel pops upon heating as a result of the unique quality of the endosperm (Acquaah, 2006). Popcorn is a soft and tasty aliment that is appreciated in many countries (Rodovalho et al., 2008).

In the last several years, many efforts have been focused on studying genetic diversity and divergence among maize populations because the success of corn breeding programs depends mainly upon these parameters (Miranda et al., 2008; Leal et al., 2010). Although Miranda et al. (2008) stated that the popcorn germplasm has a restricted genetic base, Liu et al. (2003) demonstrated that popcorn is highly differentiated from other groups of corn by assaying 260 maize inbred lines, which are representative of the genetic diversity among essentially all public lines that are of importance to temperate breeding and include many important tropical and subtropical lines. However, according to Pereira et al. (2008), studies of genetic divergence in popcorn are still scarce.

In Brazil, there is a growing annual demand for popcorn, and an approximate consumption of 70 000 t yr^{-1} is estimated. In total, 71% of this production is imported, mainly from Argentina and USA (Scapim et al., 2006; Mora and Scapim, 2007), because there are a limited number of genetically improved cultivars (Vilarinho et al., 2003). Information regarding genetic diversity and population structure in elite breeding materials is of fundamental importance for breeding programs (Van Inghelandt et al., 2010) and could facilitate either germplasm classification or the development of genetic enhancement strategies to increase genetic diversity (Humphreys et al., 2005). In this sense, Miranda et al. (2008) indicated that an appropriate level of parental divergence is required to obtain higher yielding hybrids. Blanco et al. (2005) suggested that the genetic base of breeding populations of popcorn could be enhanced through the introduction of new germplasm. Furthermore, Melani and Carena (2005) indicated that the success of a popcorn breeding program depends on parental selection and the accurate identification of heterotic groups. An analysis of the amount and distribution of genetic variation within and among populations of a crop can increase our understanding of the historical process underlying genetic diversity and provide basic information for breeding programs (Hartings et al., 2008).

[1]Universidad de Chile, Facultad de Ciencias, Las Palmeras 3425, Ñuñoa, Santiago, Chile.
[2]Universidade Estadual de Maringá, Departamento de Agronomia, 87020-900, Maringá PR, Brasil.
[3]Universidad de Talca, Instituto de Biología Vegetal y Biotecnología, 2 Norte 685, Talca, Chile. *Corresponding author (fmora@utalca.cl).
Received: 19 November 2012.
Accepted: 13 May 2013.
doi:10.4067/S0718-58392013000200003.

DNA marker technology could greatly facilitate breeding through the determination of heterotic groups (Yuan et al., 2000). In this context, many authors have demonstrated practical applications for molecular marker techniques using restriction fragment length polymorphisms (RFLPs) (Ajmone-Marsan et al., 1998; Benchimol et al., 2000), amplified fragment length polymorphisms (AFLPs) (Oliveira et al., 2004; Legesse et al., 2007) and simple sequence repeats (SSRs) or microsatellites (Reif et al., 2003). Genetic analyses using either Wright's F-statistics or genetic distances remain the most common focus for the characterization of genetic differentiation (Latch et al., 2006; Bracco et al., 2009). This information has been used to assign individuals to genetically differentiated groups.

Over the last several years, many clustering methods have been proposed for the separation of a set of individuals into different subpopulations to study the correspondence between the inferred genetic clusters and known predefined populations (Rodríguez-Ramilo et al., 2009). Moreover, significant potential lies in the ability to identify the structure of genetic diversity within and among accessions, which can be relevant for the optimization of collections, the planning of seed regeneration, and the successful implementation of pre-breeding approaches (Hartings et al., 2008). According to Latch et al. (2006), most of the recent advances in clustering techniques have been made in a Bayesian statistical framework to allow simultaneous estimation of many interdependent parameters in complex models. The advantages of utilizing a Bayesian framework in the analysis of genetic experiments have been emphasized and discussed by several authors (Mora et al., 2009; Cané-Retamales et al., 2011; Li et al., 2011; Safner et al., 2011; Arriagada et al., 2012).

The present study aimed to determine the genetic structure of 28 popcorn populations maintained in several Brazilian breeding programs. A Bayesian modeling approach was used in the data analysis. We focused on three key issues: a) selecting and characterizing SSR markers in a Brazilian popcorn germplasm, b) determining genetically differentiated groups of popcorn, and c) determining the degree of genetic structure of the different clusters.

MATERIALS AND METHODS

Population sampling and DNA extraction

A set of 420 individuals representing 28 populations of popcorn was assessed in this study (Table 1). Fifteen seeds from each population were germinated in pots at Núcleo de Pesquisa Aplicada à Agricultura (NUPAGRI), Paraná State, Brazil. The number of individuals per population was based on the studies of Hoxha et al. (2004) and Liu et al. (2009). Fifteen days after germination, genomic DNA was extracted according to the methodology proposed by Hoisington et al. (1994) with modifications. Fresh leaves (300 µg) were macerated in liquid N_2 and 800 µL of CTAB extraction buffer (1 M Tris-HCL [pH 7.5], 5 M NaCl, 0.5 M EDTA [pH 8.0], 1% CTAB and 140 mM β-mercaptoethanol). The DNA was quantified using a Quant-iT™ Assay Kit (Invitrogen, Carlsbad, California, USA) and was finally diluted to 10 ng μL^{-1}.

Table 1. Popcorn populations included in the present study.

Population	Origin	Population	Origin
ARZM 13 050	CIMMYT	BEIJA-FLOR	UFV
URUG 298	CIMMYT	UNB-2U C1*	UENF
ARZM 083	CIMMYT	UENFV-EXPLOSIVO C4*	UENF
PARA 172	CIMMYT	UNB-2U C3*	UENF
ARZM 07 049	CIMMYT	PA 091*	UEM
BOZM 260	CIMMYT	PR-023*	UEM
BOYA 462	CIMMYT	BRS ANGELA*	EMBRAPA
IAC-125**	IAC	SAM*	South America
IAC-112**	IAC	PARA 170	CIMMYT
JADE**	Pioneer Hi-bred	COLOMBIANA	UEM
ZELIA**	Pioneer Hi-bred	TATU 1	UEM
SE 013*	UEM	UNB-2U C0*	UENF
VIÇOSA*	UFV	UFVM2-BARÃO VIÇOSA*	UFV
RS-20*	IPAGRO/AGROESTE	TATU 2	UEM

CIMMYT: Centro Internacional para el Mejoramiento de Maíz y Trigo; EMBRAPA: Empresa Brasileira de Pesquisa Agropecuária; IAC: Instituto Agronômico de Campinas; UEM: Universidad Estadual de Maringá; UFV: Universidade Federal de Viçosa; UNFV: Universidade Estadual do Norte Fluminense.
*, **: Open-pollinated varieties and hybrids, respectively (Simon et al., 2004; Scapim et al., 2006; Silva et al., 2009; Carvalho et al., 2012).

SSR amplification

Eleven microsatellite primer pairs (from the Maize Genetics and Genomics Database: www.maizegdb.org/ssr.php) were selected, tested and amplified. Amplification was performed using 25 ng of template DNA with 2.0 mL of 10X reaction buffer, 2.0 mM $MgCl_2$, 0.1 mM of each dNTP, 1 unit of Taq DNA polymerase and 0.2 mM of each primer (forward and reverse) to obtain a final volume of 20 µL. The amplification was carried out based on touchdown PCR methodology (Don et al., 1991) using a TC 512 Temperature Cycling System (Techne, Staffordshire, UK).

Electrophoresis and DNA fragment visualization

The PCR products were separated on 4% agarose gels containing 50% common agarose and 50% metaphor agarose (Benchimol et al., 2005; García et al., 2007; Carvalho et al., 2012) using 0.5X TBE buffer (44.5 mM Tris base, 44.5 mM boric acid and 1 mM EDTA). The PCR products were stained with 2 µL of bromophenol blue. Each well was loaded with 20 µL of the PCR product. In the first well of each gel, 3 µL of 100 bp DNA Ladder (Invitrogen) was loaded. The electrophoresis was carried out for 4 h at 60 V. The gels were stained with a SYBR® Safe DNA Gel Stain (Invitrogen) solution for 45 min. Subsequently, the bands were visualized using a UV transilluminator and photographed using a Kodak EDAS 290 gel documentation system.

Genetic analysis

Cluster analyses were performed using Structure v.2.3.2 (Pritchard et al., 2000). An admixture model and correlated allele frequency model were used to analyze the dataset without prior population information. Structure uses multilocus genotype data to describe population genetic structure (Pritchard et al., 2000). At first, 20 runs of Structure were performed for each number of possible clusters (K = 1 to 28). The burn-in time was set to 50 000 samples, and the number of Monte Carlo Markov chains (MCMC) repeats after burn-in was set to 500 000 samples in each run (Latch et al., 2006; Li et al., 2009; Rodríguez-Ramilo et al., 2009). To avoid autocorrelation, a step of 100 was used. An ad hoc quantity based on the second order rate of change of the likelihood function with respect to K (ΔK) was used for estimating the number of clusters from Structure analysis. Evanno et al. (2005) suggested estimating ΔK as follows:

$$\Delta K = \frac{\{avg[L(K+1)] - 2avg[L(K)] + avg[L(K-1)]\}}{sd[L(K)]}$$

where *avg* is the arithmetic mean across replicates, and *sd* is the standard deviation of the replicated *L*(*K*). The value of K corresponds to the mean value of the distribution of ΔK (Rodríguez-Ramilo et al., 2009). Structure Harvester v.0.6.1 (Earl and vonHoldt, 2011) was used to process the Structure result files.

After determining the *K* value, a new analysis was performed with a burn-in period of 200 000 samples, 1 million MCMC samples and a step of 100. A Structure graphical bar plot of membership coefficients for populations and individuals was generated using the Distruct program (Rosenberg, 2004). The convergence of the Gibbs chains was determined using the test proposed by Heidelberger and Welch (1983), which was performed in the R program with the convergence diagnosis and output analysis (CODA) library (Contreras-Soto et al., 2011). Bayesian credible intervals (95%) were calculated using the Proc Univariate of SAS-STAT. The Kernel density estimation method was used for estimating mode values for each F_{ST}.

The number of alleles per locus (*Na*), observed and expected heterozygosity (*Ho* and *He*, respectively) and number of effective alleles (*Ne*) were calculated using Popgene v.1.32 (Yeh et al., 1999). The polymorphic information content (PIC) was estimated according to Smith et al. (1997):

$$PIC = 1 - \sum_{i=1}^{N} pi^2$$

where *pi* is the frequency of the i^{th} allele. This value is defined as the probability that two alleles taken at random from a population can be distinguished using the marker in question and is a measure of allele diversity at a locus (Hartings et al., 2008).

The number of private alleles (*Pa*, defined as the alleles that were only detected in a particular population; Jiang et al., 2011) was determined using GenAlEx (Peakall and Smouse, 2006). The degree of genetic diversity within clusters was estimated using Arlequin v3.5.1.2 (Excoffier and Lischer, 2010). The average number of pairwise differences within populations was used in this analysis.

RESULTS AND DISCUSSION

DNA extraction, SSR amplification and heterozygosity for the total population

The average concentration of DNA extracted from 420 individuals was 174 ng μL^{-1}, which was sufficient for dilution and subsequent amplification of SSR regions. The characterization of 11 polymorphic microsatellite *loci* used to genotype the Brazilian germplasm is shown in Table 2. The 11 primer pairs evaluated allowed the detection of 54 alleles. The number of alleles per *locus* ranged from 2 to 8 (mean value of 4.81). The *bnlg1175 locus* was the most polymorphic in terms of the number of alleles. The average number of alleles per *locus* was lower than that reported by Morales et al. (2010) (average: 5.14), who used 21 SSR markers in heterotic maize populations, but was higher than that found by Bantte and Prasanna (2003) in their analysis of 23 tropical maize lines using 36 SSR markers (average: 3.25) and that found by Bracco et al. (2009) in their study of six maize landraces using 10 microsatellite markers (average: 2.94). The effective number of alleles ranged from 1.88 to 4.74 for the *loci umc1125* and *bnlg1175*, respectively. Only one private allele was found in the ARZM 07 049 population for the *locus mmc0271* with a frequency of 0.154.

The *Pa* was lower than that observed in previous studies. In fact, Vaz Patto et al. (2009) found 12 private alleles in 54 populations of maize, which were evaluated using 13 microsatellite markers, whereas Bracco et al. (2009) found 33 private alleles in six maize landraces using 10 microsatellite markers. According to Li et al. (2008), rare SSR *loci*, especially in plants with rare alleles at multiple *loci*, may indicate that these landraces have had limited genetic exchange with other landraces, and therefore may have rare alleles for various functional traits as well. Because these unique or rare alleles are likely caused by natural mutation and selection (Mousadik and Petit, 1996), they can be used not only in the specific categorization of germplasm collections but also in breeding and plant development as unique markers (Li et al., 2008).

According to Kalinowski (2004), the simplest measure of genetic diversity at a determined *locus* is the number of alleles (i.e., allelic richness). Moreover, the number of unique alleles in a population (private allelic richness) is a simple measure of genetic distinctiveness. In the present study, the explanation for the low number of private alleles could be attributed to the narrow genetic base of popcorn (Miranda et al., 2008), artificial selection processes or a common genetic base that is shared among

Table 2. Characterization of 11 polymorphic microsatellite *loci* used in this study. Columns correspond to microsatellite name (*locus*), forward (F) and reverse (R) primer sequences, repeat motif, bin location, number of alleles per *locus* (*Na*), number of effective alleles (*Ne*), observed heterozygosity (*Ho*) and polymorphic information content (*PIC*).

Locus	Primers	Repeat	Bin	*Na*	*Ne*	*Ho*	*PIC*
umc2401	F: TTTTCTTCTCCTTCCCTCACCTG	$(CTCTCT)_4$	8.05	3	2.39	0.044	0.583
	R: GCACCTGATGCAGTAGGGAGTC						
umc2108	F: CCTTTTAATTTGCAAAGAGAGGTGG	$(CTGCA)_4$	3.00	3	2.68	0.034	0.629
	R: CAAACTAAAGAAAGTCTCGGCAGC						
umc1125	F: CGTCCGACATCTTGCTTTTCTATC	$(CTCG)_5$	7.04	2	1.88	0.305	0.470
	R: TTTTACTTCTCAGCGGTAGATCGG						
mmc0271	F: CGTAATGCGTAGCAACATAG	$(GA)_{39}$	2.07	6	2.49	0.244	0.599
	R: CAACATCCTTTCCACCG						
umc2246	F: AGGCTCCAGCTCTAGGGGAGT	$(CCTCCT)_4$	2.00	5	2.79	0.108	0.643
	R: GTGAACTGTGTAGCGTGGAGTTGT						
bnlg1175	F: ACTTGCACGGTCTCGCTTAT	$(AG)_{38}$	2.04	8	4.74	0.215	0.790
	R: GCACTCCATCGCTATCTTCC						
umc1755	F: CTCCTCGGCATCCTACGAGATAC	$(CACGAG)_4$	2.07	5	2.92	0.018	0.659
	R: GTGCTCGTGCTCGTGGCTAT						
umc2166	F: TACGTCGTACATCGCCCACC	$(GCCTC)_4$	3.05	5	3.83	0.094	0.740
	R: GTCGTAGCCATATCAGTTGGAACG						
umc1118	F: ATCAGATTCCGAAGGGTCCATAAT	$(GAGCA)_4$	1.11	5	2.24	0.099	0.556
	R: GTAGTGAAATGAATCGTGAGAGCG						
umc1065	F: ACAAGGCCATCATGAAGAGCAGTA	$(ACA)_{17}$	2.06	6	4.71	0.308	0.789
	R: CACGGTCTGGCACACTAACCTTAT						
umc1642	F: CACTACAGCGCCTGTAACTGCC	$(GCTA)_6$	7.00	5	2.13	0.049	0.533
	R: CATGAGCTAAGCAAGAGGGGTATG						
Mean				4.81	2.98	0.138	0.636

the populations in the current study. The presence of private alleles indicates a unique genetic variability at certain loci, and this information is useful for identifying accessions with exclusive genetic variability, the selection of which can increase the allelic richness of gene banks (Borba et al., 2009). The measure *Ne* is equal to the actual number of alleles if and only if all alleles have the same frequency; if they do not, *Ne* is smaller (Reif, 2004). In the present study, the number of effective alleles (*Ne*) was lower than the number of alleles found at each *locus*, and the correlation between *Ne* and *He* values was significant.

PIC values (calculated for all analyzed *loci*) were higher than the estimates of observed heterozygosity, and these values ranged from 0.4699 to 0.7903 for the *loci umc1125* and *bnlg1175*, respectively, with an average of 0.6355. *Ho* values ranged from 0.0175 to 0.3083 for the *loci umc1755* and *umc1065*, respectively. This result could suggest the occurrence of inbreeding in the entire population (Labate et al., 2003).

The mean values of PIC are consistent with those estimated by Reif et al. (2003), Aguiar et al. (2008), and Morales et al. (2010) (0.60, 0.51, and 0.68, respectively). One SSR marker was reasonably informative (0.25 < PIC < 0.5), and 10 markers were highly informative (PIC < 0.5) according to Salas Fernandez et al. (2009).

Bayesian clustering method and population assignment

The method proposed by Evanno et al. (2005), which is based on the second order rate of change of the likelihood function with respect to K (ΔK), showed a clear peak at K = 3 (Figure 1). The estimated curve was essentially flat after that peak, and significant changes were not observed up to K = 28 (data not shown). Heidelberger and Welch's test confirmed the convergence of the Gibbs chains for the F_{ST} values, taking into account K = 3 (Figure 2) and the indicated non-autocorrelation.

In cluster 1 (*Clus1*) (Figure 3), the populations TATU 1 and TATU 2 were obtained from breeding programs conducted by the Universidade Estadual de Maringá (PR, Brazil). Additionally, the populations PR-023 and PA091 (open-pollinated populations derived from American hybrids) (Silva et al., 2009) were included within *Clus1*. These results are in agreement with the findings of

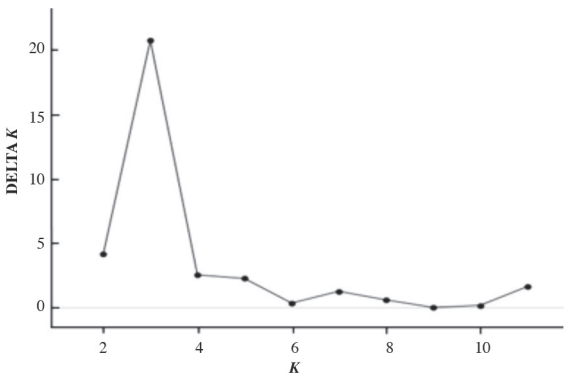

Figure 1. Determination of genetically differentiated groups (*K* = 3) using the method of Evanno et al. (2005). The Bayesian clustering analysis was performed using an admixture model and a correlated allele frequency model.

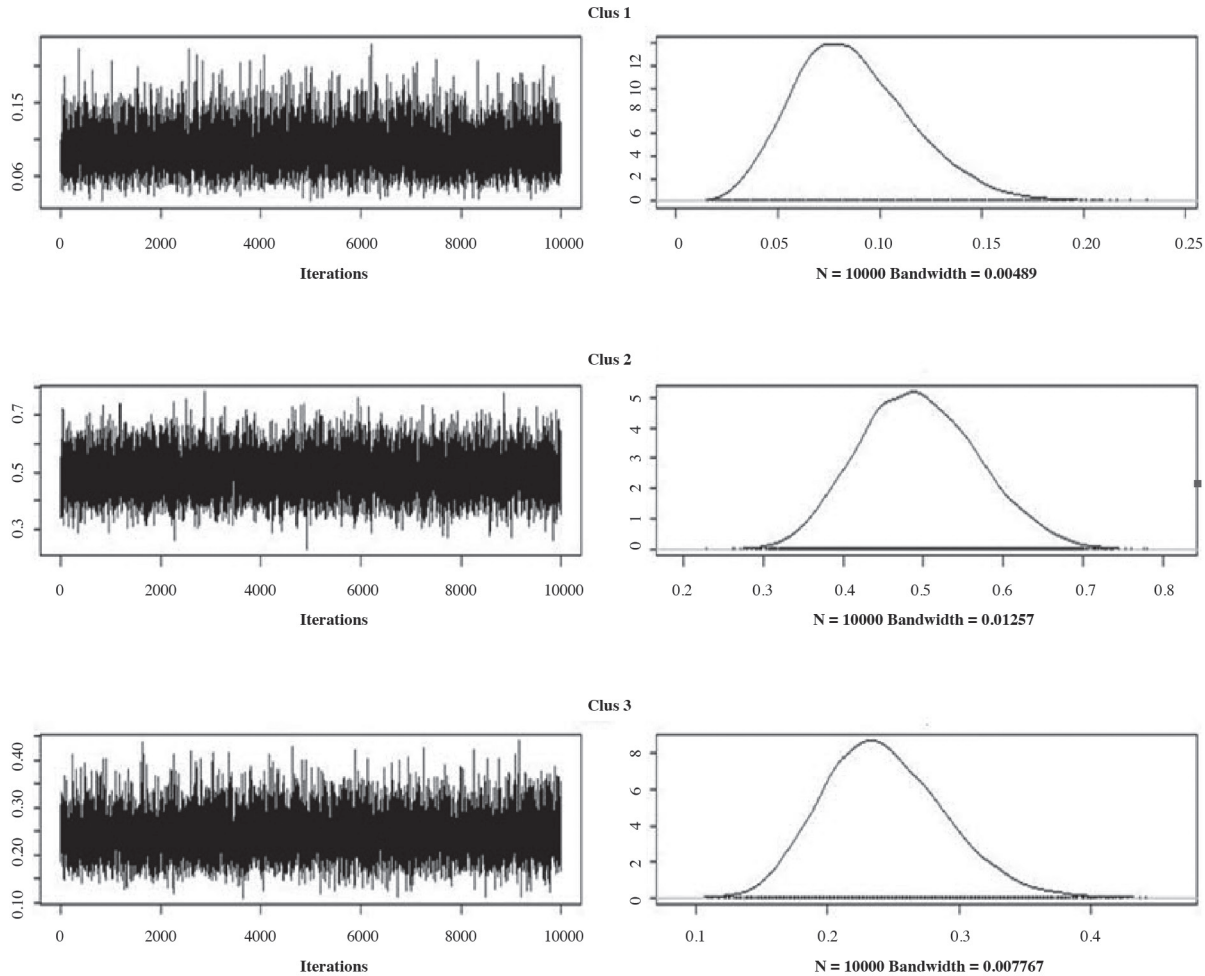

Figure 2. Marginal posterior distributions of F_{ST} values of each genetic group determined through a Bayesian clustering approach.

previous studies (Oliveira, 2010), i.e., either individuals in such populations may share a common ancestry, or an adaptation to local environmental conditions (Western Paraná) may be taking place, thus affecting the genetic similarity. These results do not agree with the findings of Silva et al. (2009), who found that populations PR-023 and PA091 were in two different groups; however, the authors used bulked DNA samples and the clustering method of Tocher, and it is not possible to determine the proportion of individuals with a particular allelic profile using this genotyping method.

The UNB-2U populations are derived from different recurrent selection cycles (UNB-2U C0, UNB-2U C1, UNB-2U C3 and UENFV-EXPLOSIVO C4). Of these populations, only UNB-2U C0 and UNB-2U C3 were included in *Clus1*. Oliveira (2010) found significant genetic diversity among UNB-2U populations based on the Jaccard index and ISSR markers, which allowed the identification

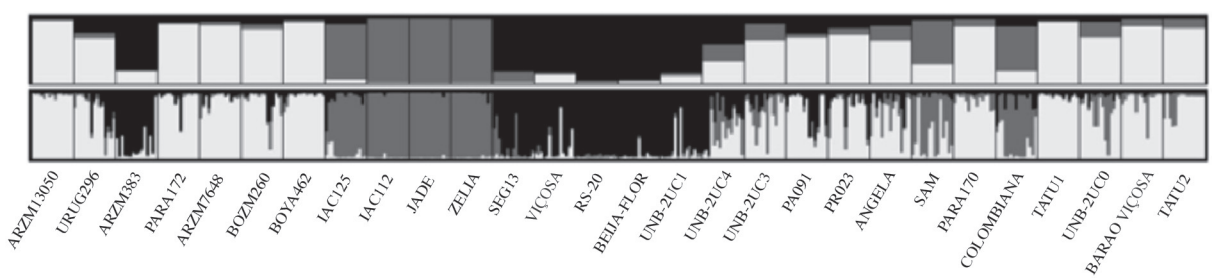

Figure 3. Bar plot showing the probability of membership for the 28 popcorn populations (above) and the 420 individuals (below) assessed based on microsatellite data. Each population and each individual is represented by a vertical line, and each genetic cluster is represented by a distinct color. The graph was constructed using DISTRUCT software (Rosenberg, 2004).

of two genetically distinct groups. Furthermore, Vilela et al. (2008) assessed the impact of recurrent selection on genetic variability of the UNB-2U population after three cycles of selection using RAPD markers (UNB-2U C0, UNB-2U C1, and UNB-2U C2), concluding that the use of different recurrent selection strategies did not lead to a genetic narrowing of the population under selection. In the current clustering study, it is therefore not surprising that the UNB-2U populations were assigned to different groups. In fact, the only population that was not assigned to any clusters ($p < 0.60$) was UENFV-EXPLOSIVO C4. This genetic differentiation is consistent with the findings of Simon et al. (2004), who stated that the effects of inbreeding depression generate dramatic changes in allele frequencies in popcorn compared with other types of corn. All populations from the CIMMYT breeding programs (with the sole exception of ARZM 083) were included in *Clus1*, among which there are several South American populations (i.e., BOYA462, BOZM260 and URUG 298 from Bolivia, Colombia and Uruguay, respectively, and PARA170 and PARA172 from Brazil) (Amaral-Júnior et al., 2011).

The JADE and ZELIA populations, which are triple hybrids from breeding programs developed by Pioneer Hi-Bred, were included in *Clus2*. Additionally, both the simple hybrid IAC-112 and the triple hybrid IAC-125 (derived from IAC-112) were also assigned to this cluster. In a previous study of genetic clustering using ISSR markers (Mendes et al., 2010), JADE, ZELIA and IAC-125 were included in the same cluster. Additionally, SAM (originating from Argentina) (Sawazaki, 1995) and COLOMBIANA (from breeding programs at the State University of Maringá) were incorporated into this cluster.

The VIÇOSA and BEIJA-FLOR populations were grouped into *Clus3*; both populations are pioneers in the genetic improvement programs conducted by the Federal University of Viçosa (Sawazaki, 1995). Mendes et al. (2010) indicated that the VIÇOSA population has not been subjected to an artificial selection process, which could mean that its genetic variability has not been affected yet. Additionally, ARZM 083, UNB-2U C1, SE013 and RS-20 were included in *Clus3*. ARZM 083 is derived from CIMMYT, and it is the only population that was not grouped into *Clus1*. This result is in agreement with the genetic analysis of Carvalho et al. (2012), who found that the ARZM 083 genotype had the lowest observed and expected mean heterozygosity values; furthermore, this genotype was the most genetically distant among the 8 popcorn genotypes that were analyzed from CIMMYT.

According to Oliveira (2010), the SE013 population and the RS-20 variety are most likely derived from North American hybrids.

Genetic parameters of the clusters

The levels of genetic differentiation for each cluster are shown in Table 3. Bayesian credible intervals (P = 95%) of the F_{ST} values (estimated using marginal posterior distribution) confirmed the existence of significant differences between each cluster. Cluster 2 (*Clus2*) showed the highest F_{ST} value, followed by cluster 3 (*Clus3*) and cluster 1 (*Clus1*).

Table 3. Bayesian point estimates (mode, median, and mean) and credible intervals (CI, 95% probability) of F_{ST} estimates.

				(CI 95%)	
Cluster	Mode*	Median	Mean	Lower	Upper
Clus1	0.075	0.084	0.087	0.044	0.140
Clus2	0.487	0.491	0.494	0.376	0.624
Clus3	0.234	0.240	0.243	0.173	0.324

*Mode values were calculated by the kernel density estimation method.

The cut-off probability for assignment to a cluster was assumed to be 0.6 according to the information provided by the Q-matrix (Pritchard et al., 2000), which accounts for the probability of membership of an individual and/or population in a particular cluster. Thus, individuals with probabilities lower than 0.6 were integrated into a fourth group (referred to as the Mixed Group) (Table 4).

Table 5 summarizes the other genetic parameters that were determined within each cluster. Three private alleles were found in *Clus1* for the *locus mmc0271*, and one private allele was found in the Mixed Group for the *locus umc1755*. *He* values ranged from 0.3735 to 0.6333, and the genetic diversity (based on the average number of pairwise differences within a population) ranged from 2.927 to 4.642. The high level of F_{ST} (0.49) found in *Clus2* was due to its low allelic richness; *Clus2* also had a value of 1.83 for *Ne* and the lowest values for *He*, *Ho* and genetic diversity. These results could explain the high susceptibility of ZELIA to fungal pathogens such as *Puccinia sorghi*, *P. polysora*, *Helminthosporium turcicum*, *H. maydis*, and *Physopella zeae*, among others (EMBRAPA, 2011). In contrast, *Clus1* showed the highest genetic variability in terms of *He* and allelic richness.

Table 4. Grouping of popcorn populations based on Q-matrix information.

Cluster	Population
Clus1	ARZM 13 050, URUG 298, PARA 172, ARZM 07 049, BOZM 260, BOYA 462, UNB-2U C3, PA 091, PR-023, BRS ANGELA, PARA 170, TATU 1, UNB-2U C0, UFVM2-BARAO VIÇOSA, TATU 2
Clus2	IAC-125, IAC-112, JADE, ZELIA, SAM, COLOMBIANA
Clus3	ARZM 083, SE 013, VIÇOSA, RS-20, BEIJA-FLOR, UNB-2U C1
Mixed Group	UENFV-EXPLOSIVO C4

Table 5. Number of alleles per *locus* (*Na*), number of effective alleles (*Ne*), number of private alleles (*PA*), observed heterozygosity (*Ho*), expected heterozygosity (*He*) and internal genetic diversity for the three clusters and a Mixed Group ($p < 0.6$).

Cluster	*Na*	*Ne*	*Pa*	*Ho*	*He*	Genetic diversity
Clus1	4.63	3.01	3	0.151	0.633	4.64
Clus2	3.63	1.83	-	0.113	0.373	2.92
Clus3	4.18	2.54	-	0.114	0.567	3.98
Mixed Group	3.81	2.54	1	0.175	0.576	4.37

Within this cluster, germplasms from CIMMYT are known to be tolerant to major diseases that affect popcorn and are early flowering varieties (Carvalho et al., 2012); therefore, this cluster may be an important source of genes that can enhance popcorn genetic improvement programs. Furthermore, the high level of heterozygosity exhibited by this cluster could allow for more adaptive plasticity to different environmental conditions; therefore, this cluster should be considered for future studies of adaptation and agronomic performance.

Nei's genetic identity and genetic distance values are shown in Table 6. The lowest genetic distance was found between *Clus1* and the Mixed Group. In fact, 50% of the individuals that comprise the Mixed Group are from different populations grouped into *Clus1*, according to the Q-matrix values for each assessed individual, whereas 19.04% and 11.92% of the individuals are included in Clus2 and Clus3, respectively. The remaining proportion corresponds to individuals of the UENFV-EXPLOSIVO C4 population. Additionally, the presence of private alleles found only in *Clus1* and the Mixed Group, the fact that *Clus1* and the Mixed Group have the two highest values of genetic diversity within populations and the low genetic differentiation value of *Clus1* (0.087) provide clear evidence of the high similarity between *Clus1* and the Mixed Group. These results, combined with the high degree of genetic differentiation of *Clus2* and *Clus3* (Table 3), could allow for the identification of heterotic groups, thus maximizing important agronomic traits in this crop (Yuan et al., 2000; Reif et al., 2003; Miranda et al., 2008).

In the long term, and to avoid exhausting the variability at the within-population level, it would be advisable to monitor the levels of genetic diversity available and to introgress valuable alleles from other populations to prevent the loss of complementary gene interactions due to inbreeding (Hartings et al., 2008). In the current study, the use of SSR markers from common maize allowed the identification of differentiated groups in popcorn populations. This result was consistent with the findings of many previous studies (Silva et al., 2009; Leal et al., 2010; Mendes et al., 2010; Ribeiro et al., 2010; Carvalho et al., 2012), which indicated that the population origin and the relationship among populations could be the main causes of divergence. The natural mutation rate of SSRs is typically too low to be a major factor responsible for population differentiation at the observed levels (Jiang et al., 2011).

Table 6. Nei's genetic identity (above the diagonal) and genetic distance (below the diagonal) among the three clusters (*Clus1*, *Clus2* and *Clus3*) and the Mixed Group performed using a Bayesian clustering process with microsatellite markers.

Cluster	*Clus1*	*Clus2*	Clus3	*Mixed Group*
Clus1	-	0.741	0.704	0.907
Clus2	0.298	-	0.675	0.879
Clus3	0.350	0.392	-	0.834
Mixed Group	0.096	0.128	0.181	-

CONCLUSIONS

The main advantages of the Bayesian clustering model were the use of individual genetic information for the estimation of Q-matrix and the subsequent assignment of populations to each cluster. The analysis of Gibbs chains and the estimation of the Bayesian credible interval confirmed the existence of significant differences between the F_{ST} values of each cluster. This information may be critical for germplasm management, introgression of genetic material, determination of heterotic groups and association mapping studies in popcorn. As an example of the latter, the use of information related to the genetic structure of populations (Q-matrix information) in association mapping studies has been incorporated into statistical models as a fixed effect, improving the control of both type I and type II error rates. Because no association mapping studies have been performed using popcorn populations to examine agronomic traits, we highlight the information generated regarding the genetic structure of popcorn populations for possible use in these types of studies. Additionally, we emphasize the Bayesian paradigm as an important issue to keep in mind in future studies related to genetic differentiation of popcorn germplasms.

ACKNOWLEDGEMENTS

Financial support for this work was provided by the National Council for Scientific and Technological Development (CNPq).

LITERATURE CITED

Acquaah, G. 2006. Principles of plant genetics and breeding. 569 p. Blackwell, Oxford, UK.

Aguiar, C.G., I. Schuster, A.T. Amaral-Júnior, C.A. Scapim, and E.S.N. Vieira. 2008. Heterotic groups in tropical maize germplasm by test crosses and simple sequence repeat markers. Genetics and Molecular Research 7:1233-1244.

Ajmone-Marsan, P., P. Castiglioni, F. Fusari, M. Kuiper, and M. Motto. 1998. Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. Theoretical and Applied Genetics 96:219-227.

Amaral-Júnior, A.T., E.C. Oliveira, L.S. Azeredo, L.S. Candido, T.R. Conceição, C. Vittorazzi, et al. 2011. Assessment of genetic diversity among maize accessions using inter simple sequence repeats (ISSR) markers. African Journal of Biotechnology 10:15462-15469.

Arriagada, O., F. Mora, J.C. Dellarossa, M.F.S. Ferreira, G.D.L. Cervigni, and I. Schuster. 2012. Bayesian mapping of quantitative trait loci (QTL) controlling soybean cyst nematode resistant. Euphytica doi:10.1007/s10681-012-0696-y.

Bantte, K., and B.M. Prasanna. 2003. Simple sequence repeat polymorphism in Quality Protein Maize (QPM) lines. Euphytica 129:337-344.

Benchimol, L.L., C.L. Jr. Souza, A.A.F. Garcia, P.M.S. Kono, C.A. Mangolin, A.M.M. Barbosa, et al. 2000. Genetic diversity in tropical maize inbred lines: heterotic group assignment and hybrid performance determined by RFLP markers. Plant Breeding 119:491-496.

Benchimol, L.L., C.L. Jr. Souza, and A.P. Souza. 2005. Microsatellite-assisted backcross selection in maize. Genetics and Molecular Biology 28:789-797.

Blanco, M.H., C.A.C. Gardner, W. Salhuana, and N. Shen (eds.) 2005. Proceedings of the 41st Annual Illinois Corn Breeders School, Champaign, Illinois. March 2005. The Illinois Maize Breeding and Genetics Laboratory, University of Illinois, Urbana, Illinois, USA.

Borba, T.C., C. Mendes, É. Guimarães, T.O. Brunes, J.R. Fonseca, R.V. Brondani and C. Brondani. 2009. Genetic variability of Brazilian rice landraces determined by SSR markers. Pesquisa Agropecuária Brasileira 44:706-712.

Bracco, M., V.V. Lia, A.M. Gottlieb, J. Cámara Hernández, and L. Poggio. 2009. Genetic diversity in maize landraces from indigenous settlements of Northeastern Argentina. Genetica 135:39-49.

Cané-Retamales, C., F. Mora, F. Vargas-Reeve, S. Perret, and R. Contreras-Soto. 2011. Bayesian threshold analysis of breeding values, genetic correlation and heritability of flowering intensity in *Eucalyptus cladocalyx* under arid conditions. Euphytica 178:177-183.

Carvalho, M.S., C.A. Mangolin, C.A. Scapim, T.A. Silva, and M.F.P. Silva. 2012. A collection of popcorn as a reservoir of genes for the generation of lineages. Molecular Biotechnology doi:10.1007/s12033-012-9527-6.

Contreras-Soto, R., F. Mora, S. Perret, F. Vargas-Reeve, y C. Cané-Retamales. 2011. Predicción bayesiana del comportamiento poblacional de *Eucalyptus cladocalyx* para características binarias de componentes de florecimiento y supervivencia en zonas áridas de Chile. Interciencia 36:644-649.

Don, R.H., P.T. Cox, B.J. Wainwright, K. Baker, and J.S. Mattick. 1991. Touchdown PCR to circumvent spurious priming during gene amplification. Nucleic Acids Research 19:4008.

Earl, D.A., and B.M. vonHoldt. 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4:359-361.

EMBRAPA. 2011. Tabela 2. Comportamento das cultivares de milho disponíveis no mercado brasileiro na safra 2009/10 em relação às principais doenças. Empresa Brasileira de Pesquisa Agropecuária. Available at http://www.cnpms.embrapa.br/milho/cultivares/TABELA2.html (accessed November 2012).

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software Structure: a simulation study. Molecular Ecology 14:2611-2620.

Excoffier, L., and H.E.L. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10:564-567.

García, A.F., J.L. Albertini, M.I. Zucchi, and A.P. Souza. 2007. Microsatellite molecular markers in the cultivar identification of Brazilian soybean for human consumption. Crop Breeding and Applied Biotechnology 7:155-164.

Hartings, H., N. Berardo, G.F. Mazzinelli, P. Valoti, A. Verderio, and M. Motto. 2008. Assessment of genetic diversity and relationships among maize (*Zea mays* L.) Italian landraces by morphological traits and AFLP profiling. Theoretical and Applied Genetics 117:831-842.

Heidelberger, P., and P.D. Welch. 1983. Simulation run length control in the presence of an initial transient. Operations Research 31:1109-1114.

Hoisington, D., M. Khairallah, and D.F. González de León. 1994. Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory. 3rd ed. CIMMYT, Mexico D.F., Mexico.

Hoxha, S., M.R. Shariflou, and P. Sharp. 2004. Evaluation of genetic diversity in Albanian maize using SSR markers. Maydica 49:97-103.

Humphreys, M.W., R.S. Yadav, A.J. Cairns, L.B. Turner, J. Humphreys, and L. Skøt. 2005. A changing climate for grassland research. New Phytologist 169:9-26.

Jiang, Z., H. Xia, B. Basso, and B. Lu. 2011. Introgression from cultivated rice influences genetic differentiation of weedy rice populations at a local spatial scale. Theoretical and Applied Genetics 124:309-322.

Kalinowski, S.T. 2004. Counting alleles with rarefaction: Private alleles and hierarchical sampling designs. Conservation Genetics 5:539-543.

Labate, J.A., K.R. Lamkey, S.E. Mitchell, S. Kresovich, H. Sullivan, and J.S.C. Smith. 2003. Molecular and historical aspects of corn belt dent diversity. Crop Science 43:80-91.

Latch, E.K., G. Dharmarajan, C.G. Jeffrey, E. Glaubitz, and O.E. Jr. Rhodes. 2006. Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. Conservation Genetics 7:295-302.

Leal, A.A., C.A. Mangolin, A.T. Amaral-Júnior, L.S.A. Gonçalves, C.A. Scapim, A.S. Mott, et al. 2010. Efficiency of RAPD versus SSR markers in determination of genetic diversity among popcorn lines. Genetics and Molecular Research 9:9-18.

Legesse, B.W., A.A. Myburg, K.V. Pixley, S. Twumasi-Afriyie, and A. Botha. 2007. Relationship between hybrid performance and AFLP based genetic distance in highland maize inbred lines. Euphytica 162:313-323.

Li, Y., R. Guan, Z. Liu, Y. Ma, L. Wang, L. Li, et al. 2008. Genetic structure and diversity of cultivated soybean (*Glycine max* (L.) Merr.) landraces in China. Theoretical and Applied Genetics 117:857-871.

Li, H., W. Han, Y. Zhu, J. Shu, X. Zhang, and K. Chen. 2009. Analysis of genetic structure and relationship among nine indigenous Chinese chicken populations by the Structure program. Journal of Genetics 88:197-203.

Li, M., J. Kantanen, A. Michelson, and U. Saarma. 2011. Genetic components of grey cattle in Estonia as revealed by microsatellite analysis using two Bayesian clustering methods. BMC Research Notes 4:37.

Liu, K., M. Goodman, S. Muse, S.J. Smith, E. Buckler, and J. Doebley. 2003. Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics 165:2117-2128.

Liu, Z., R. Guo, J. Zhao, Y. Cai, F. Wang, et al. 2009. Genetic diversity of two important groups of maize landraces with same name in china revealed by M13 Tailed-Primer SSRs. Agricultural Sciences in China 8:15-23.

Melani, M.D., and M.J. Carena. 2005. Alternative maize heterotic patterns for the northern corn belt. Crop Science 45:2186-2194.

Mendes, T.O.P., A.T. Amaral-Júnior, L.S.A. Gonçalves, C.A. Scapim, L.A. Peternelli, and V.Q.R. Silva. 2010. Pi statistics underlying the evaluation of stability, adaptability and relation between the genetic structure and homeostasis in popcorn. Acta Scientiarum Agronomy 32:269-277.

Miranda, G.V., L.V. Souza, J.C. Cardoso, L.J. Moreira, A.V. Melo, and I.C. Santos. 2008. Genetic variability and heterotic groups of Brazilian popcorn populations. Euphytica 162:431-440.

Mora, F., R. Gleadow, S. Perret, and C.A. Scapim. 2009. Genetic variation for early flowering, survival and growth in sugar gum (*Eucalyptus cladocalyx* F. Muell) in southern Atacama Desert. Euphytica 169:335-344.

Mora, F., and C.A. Scapim. 2007. Predicting breeding values of population effect of maize evaluated in Brazil and Paraguay. Agricultura Técnica 67:139-146.

Morales, M., V. Decker, and L. Ornella. 2010. Analysis of genetic diversity in Argentinean heterotic maize populations using molecular markers. Ciencia e Investigación Agraria 37:151-160.

Mousadik, A., and R.J. Petit. 1996. High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. Theoretical and Applied Genetics 92:832-835.

Oliveira, E. 2010. Marcadores ISSR na formação de grupos heteróticos e na inferencia evolutiva de milho pipoca. Tesis M.Sc. Universidade Estadual do Norte Fluminense, Campo dos Goytacazes, Rio de Janeiro, Brasil.

Oliveira, K.M., P.R. Laborda, A.A.F. Garcia, M.E. Paterniani, and A.P. Souza. 2004. Evaluating genetic relationships between tropical maize inbred lines by means of AFLP profiling. Hereditas 140:24-33.

Peakall, R., and P.E. Smouse. 2006. GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.

Pereira, L.K., C.A. Scapim, C.A. Mangolin, A. Machado, M.F. Silva, A.P. Cleso, et al. 2008. Heterozigosity following half-sib recurrent selection in popcorn using isoenzyme markers. Electronic Journal of Biotechnology 1:107-115.

Pritchard, J.K., M. Stephens, and P. Donelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Reif, J. 2004. Assessing the genetic diversity in crops with molecular markers: Theory and experimental results with CIMMYT wheat and maize elite germplasm and genetic resources. Tesis PhD. Sc. Institut für Pflanzenzüchtung, Saatgutforschung und Populations genetik, Universität Hohenheim, Stuttgart-Hohenheim, Germany.

Reif, J.C., A.E. Melchinger, X.C. Xia, M.L. Warburton, D.A. Hoisington, S.K. Vasal, et al. 2003. Use of SSRs for establishing heterotic groups in subtropical maize. Theoretical and Applied Genetics 107:947-957.

Ribeiro, A.P., R.J. Barth, A.T. Amaral-Júnior, C.A. Mangolin, M.F. Silva, and C.A. Scapim. 2010. Genetic diversity of breeding popcorn lines determined by SSR markers. Electronic Journal of Biotechnology 1:13.

Rodovalho, M., F. Mora, E. Mendes, and C.A. Scapim. 2008. Survival heritability in 169 families of white grain popcorn: A Bayesian approach. Ciencia e Investigación Agraria 35:303-309.

Rodríguez-Ramilo, S.T., M.A. Toro, and J. Fernández. 2009. Assessing population genetic structure via the maximisation of genetic distance. Genetics Selection Evolution 41:49.

Rosenberg, N.A. 2004. DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes 4:137-138.

Safner, T., M.P. Miller, B.H. McRae, M.J. Fortin, and S. Manel. 2011. Comparison of Bayesian clustering and edge detection methods for inferring boundaries in landscape genetics. International Journal of Molecular Science 12:865-889.

Salas Fernandez, M.G., I. Kapran, S. Souley, M. Abdou, I.H. Maiga, C.B. Acharya, et al. 2009. Collection and characterization of yellow endosperm sorghums from West Africa for biofortification. Genetic Resources and Crop Evolution 56:991-1000.

Sawazaki, E. 1995. Melhoramento do milho-pipoca. 21 p. Instituto Agronômico de Campinas, Campinas, São Paulo, Brasil.

Scapim, C.A., R.J.B. Pinto, A.T. Amaral-Júnior, F. Mora, and T.S. Dandolini. 2006. Combining ability of white grain popcorn populations. Crop Breeding and Applied Biotechnology 6:136-143.

Silva, T.A., R.J.B. Pinto, C.A. Scapim, C.A. Mangolin, M.F. Silva, and M.S. Carvalho. 2009. Genetic diversity in popcorn genotypes using microsatellites in bulk genomic DNA. Crop Breeding and Applied Biotechnology 9:31-36.

Simon, G.A., C.A. Scapim, C.A. Pacheco, R.J.B. Pinto, A.L. Braccini, e A. Tonet. 2004. Depressão por endogamia em populaçoes de milho-pipoca. Bragantia 63:55-62.

Smith, J.S.C., E.C.L. Chin, H. Shu, O.S. Smith, S.J. Wall, et al. 1997. An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPs and pedigree. Theoretical and Applied Genetics 95:163-173.

Van Inghelandt, D., A.E. Melchinger, C. Lebreton, and B. Stich. 2010. Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theoretical and Applied Genetics 120:1289-1299.

Vaz Patto, M.C., M.L. Alves, N.F. Almeida, C. Santos, P. Mendes, Z. Satovic, et al. 2009. Is the bread making technological ability of Portuguese traditional maize landraces associated with their genetic diversity? Maydica 54:297-311.

Vilarinho, A., J.M. Sorinano, J.F. Santos, and T.M.C. Marinho. 2003. Efficiency in selecting S1 and S2 popcorn progenies for development of inbred lines. Bragantia 62:9-17.

Vilela, F.O., A.T. Amaral-Júnior, M.G. Pereira, C.A. Scapim, A.P. Viana and S. Freitas Júnior. 2008. Effect of recurrent selection on the genetic variability of the UNB-2U popcorn population using RAPD markers. Acta Scientiarum Agronomy 30:25-30.

Yeh, F.C., R. Yang, and T. Boyle. 1999. POPGENE. Version 1.31. Microsoft Window-based Freeware for Population Genetic Analysis. University of Alberta, Edmonton, Alberta, Canada.

Yuan, L.X., S.H. Zhang, J.H. Fu, X.H. Li, Z.B. Peng, Z. Tian, et al. 2000. Study on genetic diversity and heterotic groups in maize (*Zea mays* L.). Scientia Agricultura Sinica 33:9-14.