PERGAMINO, ARGENTINA
Instituto Nacional de Tecnología Agropecuaria
CORDOBA, ARGENTINA
Universidad Nacional de Cordoba

Clustering methods for determining heterotic patterns using molecular markers

--Ornella, LA; Morales Yokobori, ML; Decker, V; Nestares, G; Eyherabide, G; Balzarini, M

In hybrid maize breeding programs, efficiency of procedures to identify inbreds used to develop outstanding single crosses strongly affects the success of the program (Hallauer and Miranda, 1988). The best hybrid combinations can be identified using information from diallel (which are prohibitive with large numbers of inbreds) or topcrosses to testers (Terron et al., Agron. Meso. 8:26-34, 1997). When a large number of germplasm exists but no established heterotic groups are available, genetically similar germplasms can be identified with molecular markers. On the basis of this information, field trials can be planned more efficiently (Reif et al., Crop Sci. 43:1275-1282, 2003).

Several studies have been published in the last few years using molecular markers to study genetic divergence with variable results (Dias et al., 2004 Genet. Mol. Res. 3:356-368). According to Reif et al. (Crop Sci. 41:1-7, 2005), the choice of a coefficient for studying divergence depends on the marker system properties involved and on the study objectives, among other conditions. According to these authors, several studies ignore these conditions, especially those related to the coefficient properties, which are connected to the study objective, which are very important for decision making considering the proper coefficient to be used. These studies usually employ the same similarity coefficients for dominant markers, such as RAPDs, and codominant and multiallele markers, such as simple sequence repeats (SSR), even though some of these coefficients are specific for dichotomic variables. Most similarity coefficients are based on comparisons between the occurrence of common and different bands (indicated by ones and zeros in common in a data matrix), while genetic dissimilarity coefficients, such as Roger's modified distance and Nei's distance, make use of information on allele frequency obtained by molecular markers, especially microsatellites (Balestre et al., Genet. Mol. Res. 7:695-705, 2008; Reif et al., 2005).

Pritchard et al. (Genetics 155:945-959, 2000) introduced the software, Structure, which has been used with relative success in maize (Camus-Kulandaivelu et al., Crop Sci 47:887-890, 2007). Given a value for the number of populations (*K*), Structure uses a Bayesian framework to assign lines from the entire sample to clusters in such a way that Hardy-Weinberg disequilibrium and linkage disequilibrium (LD) are maximally explained (Pritchard, et al., 2000). The purpose of this study was to evaluate the reliability of clustering methods based on molecular marker information to replace and/or complement topcross trials in assigning lines to heterotic groups of temperate germplasm.

For the analysis, we used the results of the molecular characterization of 21 microsatellite loci evenly distributed in the genome of 26 inbred lines. All lines except one (B73) were developed by INTA (Instituto Nacional of Tecnologia Agropecuaria) from different sources (mainly landraces) and belong to the Argentine Orange Flint heterotic group. Results were partially published in Morales Yokobori et al. (MNL 79:36-37, 2005). The entire set of 26 lines was previously grouped into four heterotic groups by topcross (Table 1) (Eyherabide et al., Plant Breeding: The Arnel R. Hallauer International Symposium, Blackwell Publishing, pp. 352–379, 2006).

Table 1. Clustering of lines established by topcross (Eyhérabide et al., 2006; Nestares et al., 1999).

Heterotic Group	Inbreds
I	B73, lp17, lp32, lp521, lp122
II	lp123, lp153, lp22, lp44, lp662, lp70, P1338
III	lp13, lp146, lp147, lp19, lp199, ZN6
IV	lp38, lp62, lp103, lp109, lp110, lp138, lp152, lp140

Cluster analysis was performed using the Unweighted Pair Group Method using Arithmetic averages (UPGMA) and on the basis of Modified Roger's distance (MRD). According to Melchinger (The Genetics and Exploitation of Heterosis in Crops, pp. 99–118, 1999*)*, heterosis is a function of the dominance effect of the QTL and of MRD between parents. Reif et al. (2005) states that MRD is especially suitable in studies based on (i) the prediction of heterosis with genetic dissimilarities or (ii) the establishment of heterotic groups.

Both distance and clustering were performed using InfoStat/P, v1.1 (Grupo InfoStat, FCA, Córdoba Argentina). Four groups were determined by visual inspection of dendrograms (Table 2). Lines were also subdivided into 4 genetic clusters using **Structure** (Pritchard et al., 2000). We set the parameter K = 4, the number of heterotic populations previously established by topcross. Burn-in time and replication number were both set to 500,000. Results can be seen in Table 3.

Table 2. Clustering of lines based on UPGMA (Unweighted Pair Group Method with Arithmetic average) and Roger's Modified Distance.

Cluster	Inbreds
1	lp38, lp44
2	lp152, p1338, ZN6, lp199, lp521,lp117
3	lp138, lp22, lp32,lp62,lp110,lp19
4	lp103, lp122, lp123, lp109, lp13, lp662, lp153, lp70, B73, lp140, lp146, lp147

Table 3. Clustering of lines according to Structure software (Pritchard et al., 2000).

Cluster	Inbreds
A	lp103, lp122 , lp123, lp22,lp32 , lp38, lp44
B	B73, lp110, lp138, lp140, lp19, lp62, lp662
C	lp117, lp152, lp199, lp521,p1338, ZN6
D	lp109, lp13, lp146, lp147, lp153, lp70

A script in R language (http://www.r-project.org/) was made in order to determine the best level of agreement between clustering based on molecular data (this work) and clustering based on topcross (Eyhérabide et al., 2006). This allows identification of the best match between molecular and topcross groups. Concordance was measured by Cohen's Kappa coefficient (psy package of R Project). Cohen's kappa measures the agreement between two raters who each classify *N* items into *C* mutually exclusive categories. K< 0 indicates no agreement whereas 1 indicates a perfect match. Kappa values ranged from 0.16 to 0.24 (Table 4), which indicates a fair agreement.

To the present, distance-based methods are most frequently applied (Reif et al., 2005); however, we found that STRUCTURE grouping shows better agreement with topcross data than distance-based methods (Table 4). This could be attributed to: a) the

Table 4. Rate of agreement between topcross grouping and clusters, based on molecular data.

	UPGMA-Modified Rogers Distance	Structure
Topcross	0.1612903	0.3346457

low cophenetic coefficient (0.65) which indicates the goodness of fit of the cluster to the distance matrix, and b) the better performance of Structure per se (Pritchard et al., 2000). All clustering methods, like most reported in the bibliography, combined molecular information under the assumption that loci contribute in similar fashion to heterosis: lines clustered together display similar heterotic perfomance independently of the cross under evaluation. However, it has been reported in testcross trials that QTLs (quantitative trait loci) responsible for grain yield detected with only one tester were not necessarily detected for the rest of the testers (Austin et al., Crop Sci. 40:30-39, 2000; Mihaljevic et al., Crop Sci. 45:114-122, 2005). If so, not only markers associated with loci that positively affect heterosis must be selected for clustering, but more refined clustering algorithms must be designed in order to account for tester effects.

***Aknowlegments*:** We would like to thank Dr. E. Tapia (FCElyA-UNR-Argentina) for his technical advice.

PIACENZA, ITALY
Università Cattolica S. Cuore (UCSC)

Fine mapping and gene expression analysis of *de18*, a defective endosperm mutant of maize affecting auxin metabolism

--Lanubile, A; Pasini, L; Marocco, A

The *defective endosperm* (*de*) mutants of maize are a class of mutations affecting the final size and shape of the endosperm. On the basis of their pattern of accumulation of dry matter and total proteins in the seed, the mutants are divided into three major classes showing, with respect to the wild type B37: 1) a reduced accumulation rate of dry matter throughout the whole grain-filling period; 2) an initial normal rate followed by an early slowing down, and 3) an initial lag in the accumulation rate coupled with an early termination.

The mutant *de18* drastically reduces the growth rate of the grain throughout the period of development time and, at maturity, the final seed weight is less than one half that of the wild type counterpart. Preliminary histological results indicate a normal number of cells in the endosperm of *de18* seed while the cell dimensions are altered. It is also known that the auxin IAA levels in *de18* endosperm are several times lower with respect to the wild type. In fact, during the five developmental stages covering the most important part of the grain-filling period, namely 12, 15, 20, 30, and 40 DAP, the content of total indole-acetic acid (IAA) in *de18* is at least 15 times lower than in B37. As a consequence, when naftalen-acetic acid (NAA), a synthetic auxin which mimics IAA in its biochemical functions and which is not degraded to such a high extent as IAA, is added to developing seeds, a larger increase in seed weight is observed in the mutant (Torti et al., Theor. Appl. Genet. 72:602-605, 1986).

The *de18* mutant was backcrossed five times with the inbred B37. Segregation data, obtained from F2 ears, were used to evaluate the 3:1 segregating ratio by the chi-squared test. The F2 population was selfed and 16 homozygous wild type and 24 mutant F3 families were selected and used for mapping. The Bulked Segregant Analysis (Michelmore et al., Proc. Natl. Acad. Sci. USA 88:9828-9832, 1991) was applied with 191 *Eco*RI/*Mse*I primer combinations in order to find AFLP (Vos et al., Nucl. Acids Res. 23:4407-4414, 1995) markers linked to the mutation. Only AFLP bands segregating in coupling with the mutant and polymorphic between B37 and *de18* were considered. The polymorphic AFLPs were also tested in B73 and Mo17 in order to permit their use as bridges for their integration into the IBM2 reference map, using the software MapMaker 3.0.

The *de18* locus mapped to chromosome 10 bin 10.03 and was fully linked to the AFLP marker E3443_23, as shown by the absence of recombinants in the F_3 population (Figure 1). Other AFLP markers, E4335_14, E4539_7, E4335_19, E3537_4, E3942_17 (all with three recombinants) and E4445_1 (one recom-

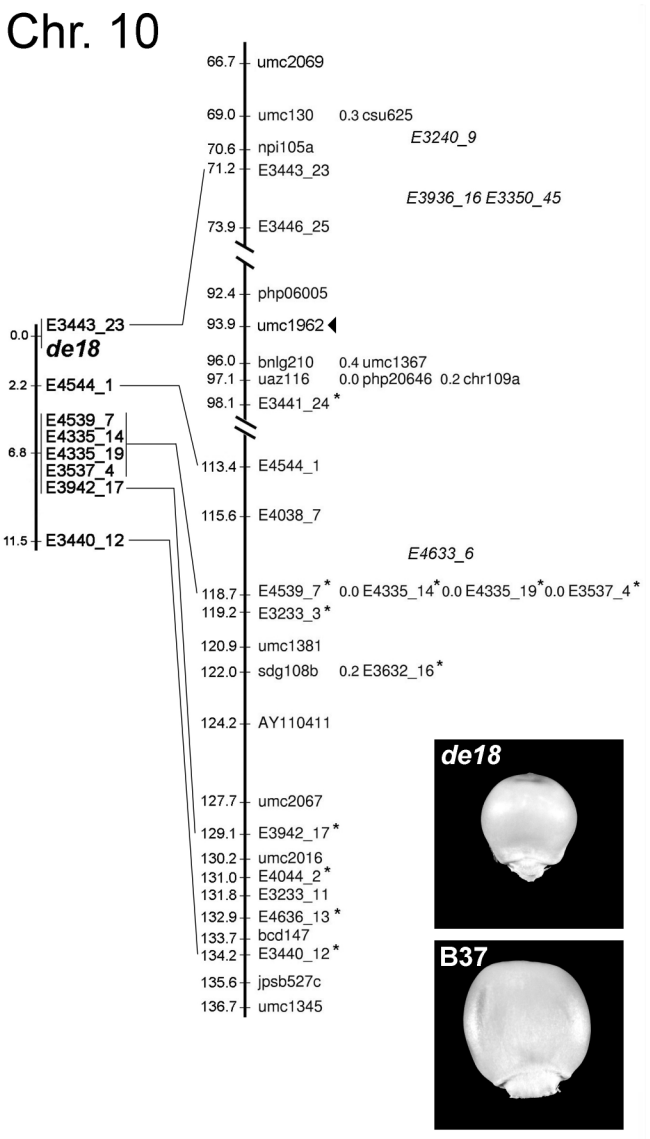

Figure 1. Map position of the *de18* locus on chromosome 10 bin 10.03. The arrow indicates the *umc1962* SSR marker in complete linkage with the mutant. AFLP markers are labelled E; the asterisk indicates the AFLP marker mapped using the NEAR command from MapMaker3.0; italics indicate AFLPs placed with the TRY command from MapMaker3.0, assigned to a map interval and reported on the right of the backbone. Map distances are reported in cM.