Genetic Distance Based on Simple Sequence Repeats and Heterosis in Tropical Maize Populations

J. C. Reif, A. E. Melchinger,* X. C. Xia, M. L. Warburton, D. A. Hoisington, S. K. Vasal, G. Srinivasan, M. Bohn, and M. Frisch

ABSTRACT

Heterotic groups and patterns are of fundamental importance in hybrid breeding of maize (*Zea mays* L.). The major goal of this study was to investigate the relationship between heterosis and genetic distance determined with simple sequence repeat (SSR) markers. The objectives of our research were to (i) compare the genetic diversity within and between seven tropical maize populations, (ii) test alternative hypotheses on the relationship between panmictic midparent heterosis (PMPH) and genetic distances determined with SSR markers, and (iii) evaluate the use of SSR markers for grouping of germplasm and establishing heterotic patterns in hybrid breeding of tropical maize. Published data of a diallel of seven tropical maize populations evaluated for agronomic traits in seven environments were reanalyzed to calculate PMPH in population hybrids. In addition, 48 individuals from each population were sampled and assayed with 85 SSR markers covering the entire maize genome. A total of 532 alleles in the 7 × 48 genotypes assayed were detected. The analysis of molecular variance (AMOVA) revealed that 89.8% of the variation was found within populations and only 10.2% between populations. The correlation between PMPH and the squared modified Roger's distance (MRD) based on SSR markers was significantly positive ($P < 0.05$) only for grain yield ($r = 0.63$). With SSR analyses, it was possible to assign Population 29 (Pop29) to the established Heterotic Group A and propose new heterotic groups (Pop25, Pop43). We conclude that SSR markers provide a powerful tool for grouping of germplasm and are a valuable complementation to field trials for identifying groups with satisfactory heterotic response.

GENETIC DIVERSITY in maize plays a key role for future breeding progress. The development of molecular markers provides a tool for assessing the genetic diversity at the DNA level in plant species (Melchinger and Gumber, 1998). In particular, SSR markers show potential for large-scale DNA fingerprinting of maize genotypes due to the high level of polymorphism detected (Smith et al., 1997), their analyses by automated systems (Sharon et al., 1997), and their high accuracy and repeatability (Heckenberger et al., 2002).

Most evidence in maize suggests that the genetic basis of heterosis is partial to complete dominance (Hallauer et al., 1988; Stuber et al., 1992). Overdominance has long been discussed as the basis of heterosis (East, 1936; Crow, 1948). However, many data supporting overdominance presumably resulted from pseudooverdominance, arising from dominant alleles in repulsion phase linkage (Stuber et al., 1992; Crow, 1999). Epistasis, particularly between linked loci, may also be an explanation for heterosis in maize (Cockerham and Zeng, 1996). No data exclude the possibility of all three mechanisms contributing to heterosis, albeit in different proportions.

Lamkey and Edwards (1999) coined the term *panmictic midparent heterosis* to describe the deviation in performance between a population cross and the mean of its two parent populations in Hardy-Weinberg equilibrium. Quantitative genetic theory shows that in the absence of epistasis and two alleles per locus, PMPH is a function of the product of the dominance effect and the square of the difference in gene frequencies at the respective locus (Falconer and Mackay, 1996, p. 255), which corresponds to the square of the MRD (Melchinger, 1999). In fact, a linear increase in PMPH with increasing genetic distance (Hypothesis 1) was hypothesized in a diallel of U.S. maize populations (Moll et al., 1962).

In contrast, experimental data reported by Moll et al. (1965) in a study with tropical maize populations of diverse geographic origin suggest that PMPH increases with increasing genetic distance only up to an optimum level but thereafter decreases in extremely wide crosses (Hypothesis 2). The authors explained this by fertility distortion in wide crosses and epistatic interactions of genes. While Moll et al. (1962, 1965) inferred the genetic distance from the geographic origin of the populations, to our knowledge no attempts have been made to verify or falsify the above hypotheses with more reliable data based on molecular markers.

The choice of heterotic groups is fundamental in hybrid breeding of maize (Melchinger and Gumber, 1998). While heterotic patterns in temperate maize have been established more than 50 yr ago, a clearly defined heterotic pattern does not exist in the tropical maize of the CIMMYT germplasm. Therefore, before embarking on a hybrid breeding program, CIMMYT conducted several diallel studies for identifying populations showing not only good per se performance but also high heterosis in their crosses (Beck et al., 1990; Crossa et al., 1990; Vasal et al., 1992a,b,c). Genetic distances based on molecular markers have been suggested as a tool for grouping of similar germplasm as a first step in identifying promising heterotic patterns (Melchinger, 1999).

The major goal of this study was to investigate the relationship between heterosis and genetic distance de-

J.C. Reif, A.E. Melchinger, M. Bohn, and M. Frisch, Inst. of Plant Breeding, Seed Sci., and Population Genetics, Univ. of Hohenheim, 70593 Stuttgart, Germany; X.C. Xia, M.L. Warburton, D.A. Hoisington, S.K. Vasal, and G. Srinivasan, Int. Maize and Wheat Improvement Center (CIMMYT), Apdo. Postal 6-641 06600 Mexico D.F., Mexico. Received 7 June 2002. *Corresponding author (melchinger@uni-hohenheim.de).

Published in Crop Sci. 43:1275–1282 (2003).

Abbreviations: ANOVA, analysis of variance; AMOVA, analysis of molecular variance; CIMMYT, International Maize and Wheat Improvement Center; GCA, general combining ability; MRD, modified Roger's distance; PC, principal coordinate; PCoA, principal coordinate analysis; PMPH, panmictic midparent heterosis; Pop, population; SCA, specific combining ability; SSR, simple sequence repeat.

termined with SSR markers. The objectives of our research were to (i) compare the genetic diversity within and between seven tropical maize populations, (ii) test alternative hypotheses on the relationship between PMPH and genetic distances determined with SSR markers, and (iii) evaluate the use of SSR markers for grouping of germplasm and establishing heterotic patterns for hybrid breeding of tropical maize.

MATERIALS AND METHODS

Field Trials

The field experiments were previously described in detail by Vasal et al. (1992a). Briefly, their investigation involved six tropical late white maize populations and one gene pool developed by CIMMYT (Table 1). The seven maize populations were crossed in a 7 × 7 diallel mating design at Poza Rica, Mexico, in the 1985 winter season. All possible 21 crosses were made in both directions using bulked pollen of each parent population. Seeds from each cross and its reciprocal were bulked to represent a particular cross. Seed increase of each parent population was done simultaneously by random mating to ensure Hardy-Weinberg equilibrium.

The parents and their crosses were evaluated in field trials for grain yield, days to silking, and plant height at seven locations (Tlaltizapán, Poza Rica, Silao, Tlacomulco, and Obregón in Mexico; Palmira in Colombia; and Nakornsawan in Thailand) during 1985-1986. The experimental design was a randomized complete block design with three replications at each location. The experimental unit consisted of two 5-m rows spaced 75 cm and a plant density of ≈53 333 plants ha^{-1}. All rows were hand-harvested and grain yield was calculated from dry ear weight at harvest assuming 80% shelling and adjusted to 155 g kg^{-1} grain moisture.

Simple Sequence Repeat Analyses

From each of the seven populations, 48 randomly chosen individuals were analyzed separately. The seeds used for extracting DNA were from the same selection cycle as the populations tested in the field trials; however, the populations were multiplied repeatedly by CIMMYT's maize genebank since 1985.

DNA was extracted employing the CTAB procedure (Clarke et al., 1989). The 85 SSR markers were chosen from the MaizeDB database (*http://nucleus. agron.missouri.edu/cgi-bin/ssr_bin.pl*) based on repeat unit and bin location to provide uniform coverage of the entire maize genome. The SSRs were multiplexed for maximum efficiency. Fragments were separated using acrylamide gels run on an ABI 377 automatic DNA sequencer. Fragment sizes were calculated with GeneScan 3.1 (Perkin Elmer/Applied Biosystems) using the Local Southern sizing method; allele identity was assigned using Genotyper 2.1 (Perkin Elmer/Applied Biosystems) and the two inbred lines CML51 and CML292 as control. Data have been stored in the MaizeDB database (*http://nucleus. agron.missouri.edu/cgi-bin/ssr_bin.pl*).

Statistical Analyses

Analyses of variance (ANOVA) were computed for the three plant traits. A mixed linear model was used with the assumption that effects of entries were fixed and all other effects were considered random. Following Analysis III of Gardner and Eberhart (1966), the sums of squares and degrees of freedom (27 *df*) for entries were orthogonally partioned into the contrast between parents *vs.* crosses (1 *df*), the variation among populations (6 *df*), and the variation among crosses (20 *df*) with a further subdivision into general combining ability (GCA) and specific combining ability (SCA) effects. A corresponding subdivision was made on the entry × environment interaction sums of squares. Entry mean squares were tested by *F* tests for significance by using the corresponding entry × environment mean squares. Entry × environment mean squares were tested for significance by using the pooled error mean square. The PMPH of each cross was calculated as the difference between the F_1 mean and the respective midparent mean across all environments.

The gene diversity (D) based on SSR data was calculated for each population according to Weir (1996, p. 151):

$$D = 1 - \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{a_i} p_{ij}^2, \qquad [1]$$

where p_{ij} is the frequency of the jth allele at the ith marker, a_i is the number of alleles at the ith marker, and m refers to the number of markers. In addition to D, we used the AMOVA to divide the genetic variation into components attributable to the variance between and within populations (Michalakis and Excoffier, 1996).

We calculated the MRD between two populations or individuals (Wright, 1978, p. 91; Goodman and Stuber, 1983) as:

Table 1. Description of the seven CIMMYT tropical late maize populations used in this study.

Population or pool	Name; selection cycle; Heterotic Group	Germplasm description
Pool24	Tropical Late White Dent; C_{21}; A	Mainly based on Tuxpeño germplasm but includes also some materials from Central America, the Carribean, and Zaire. White dent grain type. Tolerant to ear and stalk rots. Selected for resistance to fall armyworm.
Pop21	Tuxpeño-1; C_5; A	Composed of seven Tuxpeño races plus some familes from Pool 24. White dent grain type. Excellent standability and relatively short plant type. Fairly tolerant to most foliar diseases.
Pop22	Mezcla Tropical Blanc; C_6; A	Broad genetic base, including Tuxpeño, ETO Blanco, Antigua, and Central American germplasm. White dent-semident grain type. Improved for downy mildew resistance in Thailand and the Philippines.
Pop25	Blanco-Cristalino-3; C_0; B	Derived from tropical late white flint Pool 23. Composed of white flint selections from crosses among materials from Mexico, Colombia, the Caribbean, Central America, India, Thailand, and the Philippines. White flint grain type. Improved for husk cover and resistance to ear and stalk rot as well as root and stalk lodging.
Pop29	Tuxpeño Caribe; C_5; unassigned	Broad genetic base including Tuxpeño, Cuban flints, and ETO. White dent grain type. Improved for reduced plant height, stalk and root lodging, and husk cover.
Pop32	ETO Blanco; C_5; B	Developed in Colombia with germplasm from South America, Cuba, Mexico, and the U.S. cornbelt. White flint grain type. Improved for shorter plant type at CIMMYT.
Pop43	La Posta; C_5; unassigned	Tuxpeño synthetic composed of 16 S_1 lines. White grain type. Improved for resistance to streak virus in Nigeria.

Table 2. Means (above diagonal) and panmictic midparent heterosis (below diagnoal) for grain yield, days to silking, and plant height of seven CIMMYT tropical late white maize populations and their crosses averaged across data from seven environments during 1985 and 1986.

	Populations						
	Pool24	**Pop21**	**Pop22**	**Pop25**	**Pop29**	**Pop32**	**Pop43**
	Grain yield, Mg ha^{-1}						
per se	**6.36**	**6.66**	**7.12**	**6.31**	**6.51**	**5.96**	**7.05**
Pool24		**7.22**	**6.90**	**6.80**	**6.78**	**6.56**	**6.98**
Pop21	**0.71**		**7.34**	**7.40**	**6.98**	**7.15**	**7.83**
Pop22	**0.16**	**0.45**		**6.92**	**7.21**	**7.55**	**7.55**
Pop25	**0.47**	**0.92**	**0.21**		**6.78**	**6.68**	**7.07**
Pop29	**0.37**	**0.40**	**0.40**	**0.37**		**7.34**	**7.06**
Pop32	**0.64**	**0.84**	**1.01**	**0.55**	**1.11**		**7.40**
Pop43	**0.10**	**0.98**	**0.47**	**0.39**	**0.28**	**0.90**	***0.49***†
	Plant height, cm						
per se	**217.0**	**217.9**	**212.9**	**205.9**	**204.1**	**217.1**	**234.9**
Pool24		**224.9**	**217.2**	**216.6**	**213.9**	**218.5**	**227.1**
Pop21	**7.5**		**219.6**	**216.7**	**210.8**	**223.8**	**226.8**
Pop22	**2.3**	**4.2**		**208.3**	**212.1**	**216.3**	**216.7**
Pop25	**5.2**	**4.8**	**−1.1**		**215.1**	**213.0**	**220.8**
Pop29	**6.1**	**−0.2**	**3.6**	**10.1**		**211.2**	**216.9**
Pop32	**−0.5**	**6.3**	**1.3**	**1.5**	**0.6**		**229.6**
Pop43	**−9.4**	**0.4**	**−7.2**	**0.4**	**−2.6**	**3.6**	***8.8***†
	Days to silking, d						
per se	**68.0**	**69.8**	**67.4**	**66.4**	**68.3**	**69.1**	**70.4**
Pool24		**69.4**	**67.3**	**67.9**	**68.0**	**67.8**	**69.8**
Pop21	**0.5**		**68.2**	**67.7**	**69.4**	**68.8**	**68.9**
Pop22	**−0.4**	**−0.4**		**66.3**	**66.4**	**67.1**	**67.1**
Pop25	**0.7**	**−0.4**	**−0.6**		**66.5**	**66.9**	**67.9**
Pop29	**−0.3**	**0.4**	**−1.5**	**−0.9**		**68.0**	**69.2**
Pop32	**−0.7**	**−0.7**	**−1.2**	**−0.9**	**−0.7**		**68.9**
Pop43	**−1.3**	**−1.2**	**−1.8**	**−0.5**	**−0.2**	**−0.9**	***1.4***†

† **LSD0.05 of the means.**

$$\mathrm{MRD} = \sqrt{\frac{1}{2m}\sum_{i=1}^{m}\sum_{j=1}^{a_i}(p_{ij} - q_{ij})^2}, \qquad [2]$$

where p_{ij} and q_{ij} are allele frequencies of the jth allele at the ith marker in the two entries under consideration and a_i and m as defined above. Standard errors of MRD estimates were obtained by using a bootstrap procedure with resampling over markers and individuals.

Associations among the populations were revealed with principal coordinate analysis (PCoA) (Gower, 1966) based on MRD estimates. Multiple regression analysis was used to study the relationship between PMPH and squared modified Roger's distance (MRD^2). The PCoA was performed with the statistical software R (Ihaka and Gentleman, 1996) and multiple regression analysis with the statistical software SAS (SAS Institute, 1988).

RESULTS

Agronomic Trials

The combined ANOVA showed highly significant ($P < 0.01$) differences among the 28 entries (7 populations, 21 crosses) for all three traits, but no significant genotype × environment interactions (Table 2 of Vasal et al., 1992a). The comparison of parents vs. crosses, which provides a measure for average PMPH, was significant ($P < 0.01$) only for grain yield and amounted to 0.56 Mg ha^{-1}. Grain yield differed significantly ($P < 0.01$) among the seven parent populations as well as among the 21 crosses and ranged from 5.96 Mg ha^{-1} (Pop32) to 7.12 Mg ha^{-1} (Pop22) for the parent populations and from 6.56 Mg ha^{-1} (Pop32 × Pool 24) to 7.83 Mg ha^{-1} (Pop21 × Pop43) for the crosses (Table 2). The variation among the crosses was mainly due to significant ($P < 0.01$) GCA effects, whereas SCA effects were not significant for any trait.

Maximum PMPH for grain yield was observed in cross Pop29 × Pop32 with 1.11 Mg ha^{-1}, although it was not the top yielding cross. Minimum PMPH was observed in cross Pop43 × Pool24 with 0.10 Mg ha^{-1}.

Simple Sequence Marker Data

The 85 SSR primers generated a total of 532 alleles in the 336 genotypes (7 populations × 48 individuals) analyzed. The number of alleles per marker across all seven populations was on average 6.3 and ranged from 2 to 16 (Table 3). Gene diversity D within the seven populations ranged from 0.503 to 0.580 with a mean of 0.539 (Table 3). Values of MRD between pairs of populations averaged 0.258 and ranged from 0.203

Table 3. Gene diversity D within populations, average number ($\bar{a}$) and standard deviation σ_a of alleles per population.

	Population							
Statistic	**Pool24**	**Pop21**	**Pop22**	**Pop25**	**Pop29**	**Pop32**	**Pop43**	**Total**
D	**0.559**	**0.548**	**0.535**	**0.527**	**0.580**	**0.518**	**0.503**	**0.593**
$\bar{a}$	**4.247**	**4.259**	**4.226**	**4.000**	**4.294**	**3.541**	**3.553**	**6.259**
σ_a	**2.029**	**1.814**	**1.679**	**1.766**	**1.792**	**1.593**	**1.687**	**2.583**

Table 4. Modified Roger's distances between populations (above diagonal) and their standard error (below diagonal).

	Population						
Population	**Pool24**	**Pop21**	**Pop22**	**Pop25**	**Pop29**	**Pop32**	**Pop43**
Pool24		**0.219**	**0.203**	**0.224**	**0.216**	**0.270**	**0.248**
Pop21	**0.016**		**0.222**	**0.272**	**0.236**	**0.305**	**0.286**
Pop22	**0.014**	**0.016**		**0.250**	**0.233**	**0.284**	**0.268**
Pop25	**0.021**	**0.024**	**0.021**		**0.259**	**0.263**	**0.278**
Pop29	**0.017**	**0.015**	**0.016**	**0.023**		**0.285**	**0.274**
Pop32	**0.017**	**0.019**	**0.020**	**0.023**	**0.023**		**0.318**
Pop43	**0.019**	**0.018**	**0.018**	**0.026**	**0.019**	**0.021**	

(Pool24 × Pop22) to 0.318 (Pop32 × Pop43) with significant ($P < 0.01$) differences between MRD estimates (Table 4). The AMOVA revealed that 89.8% of the molecular genetic variance was found within populations and 10.2% between populations (Table 5).

In the PCoA based on MRD estimates for the populations, the first three principal coordinates (PC) explained 27.3, 22.1, and 15.8% of the total variation, respectively (Fig. 1). Pop21, Pop22, Pop29, and Pool24 were clearly separated from Pop32 and Pop25 with respect to the first principal coordinate (PC1). Pop43 and Pop25 were separated from the other populations with respect to PC2 and PC3. Principal coordinate analysis based on individual plants also resulted in a clear separation between a cluster consisting of Pop21, Pop22, Pop29, and Pool24 and a cluster comprising Pop25, Pop32, and Pop43 (Fig. 2).

Relationship between Panmictic Midparent Heterosis and Marker Data

The MRD^2 was plotted against PMPH of grain yield, plant height, and days to silking (Fig. 3) and analyzed with multiple regression. The MRD^2 was significantly correlated with PMPH for grain yield ($r = 0.63$; $P < 0.01$) and negatively for days to silking ($r = -0.44$; $P < 0.05$) and plant height ($r = -0.13$). Neither the quadratic nor the cubic regression model gave a significantly better fit to the data than the linear regression (data not shown).

DISCUSSION

CIMMYT's maize germplasm bank contains about 8000 accessions of tropical maize for use in breeding. Breeding efforts at CIMMYT in the early 1960s and 1970s were focused on population improvement via recurrent selection and, therefore, emphasized formation of genetically broad-based populations and pools disregarding heterotic patterns and combining ability (Vasal et al., 1999). Their mixed genetic constitution makes the task of assigning them to genetically diverse and complementary heterotic groups difficult. To achieve this goal, germplasm originally developed by intermating genetically diverse races were grouped according to ecology, grain color, and maturity. The groups were tested in diallel designs, each involving six to 10 populations or pools. On the basis of their performance data, the populations were categorized (Vasal et al., 1999). Pop21, Pop22, and Pool24 were assigned to Heterotic Group A, while Pop25 and Pop32 were allotted to Heterotic Group B. Pop29 and Pop43 have not yet been assigned to these or other heterotic groups.

Table 5. Analysis of molecular variance of the seven tropical maize populations analyzed with 85 SSR markers.

Source of variation	df	SS	Variance components	% variation
Among populations	**6**	**1 443.8**	**2.3**	**10.2**
Within populations	**665**	**13 430.6**	**20.2**	**89.8**
Total	**671**	**14 874.4**	**22.5**	**100.0**

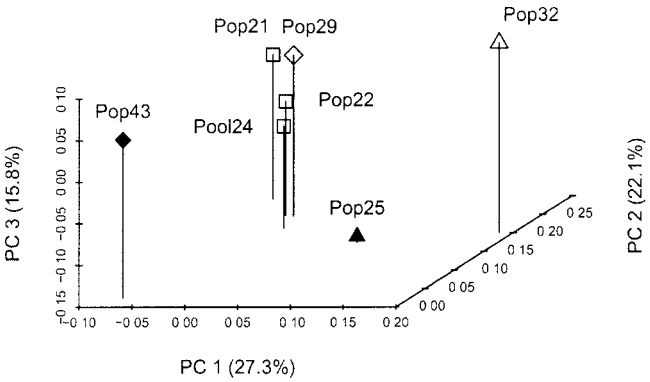

Fig. 1. Principal coordinate analysis of the seven tropical maize populations based on modified Roger's distance. PC1, PC2, and PC3 are the first, second, and third principal coordinate, respectively. Heterotic Group A (Pop21, Pop22, and Pool24), Heterotic Group B (Pop25, Pop32), and populations not yet assigned to heterotic groups (Pop29, Pop43) are shown.

Genetic Diversity among and within the Populations

In this study, we found on average across the seven populations 6.3 alleles per marker. Lu and Bernardo

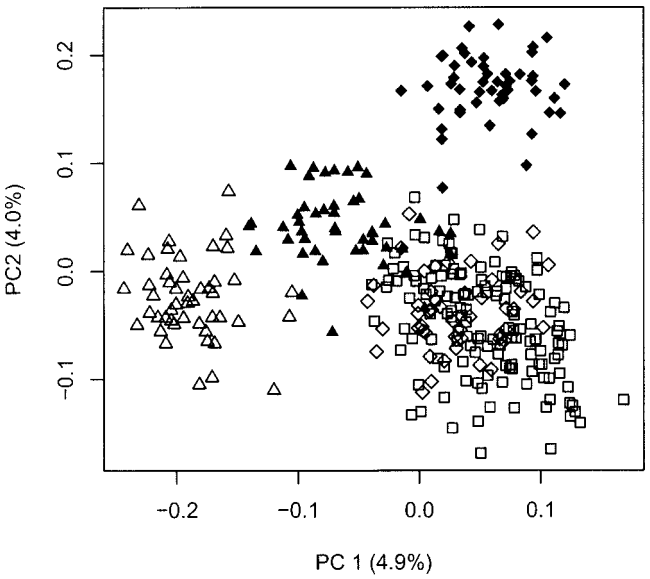

Fig. 2. Principal coordinate analysis of individuals from seven tropical maize populations based on modified Roger's distance. PC1 and PC2 are the first and second principal coordinate, respectively. Heterotic Group A (Pop21, Pop22, and Pool24, open squares), Heterotic Group B (Pop25, filled triangles; Pop32, open triangles), and populations not yet assigned to heterotic groups (Pop29, open diamonds; Pop43, filled diamonds) are shown.

(2001) detected for 40 U.S. inbred lines an average of 4.9 alleles using 83 SSR markers. Senior et al. (1998) reported an average of five alleles in their study with 94 elite maize inbreds, representative of the diversity in the U.S. maize germplasm, and 70 SSR markers. Hence, the total number of alleles per marker was higher in our study than previously reported in the literature. This and the high average number of alleles per population (Table 3) in our study suggests a broad genetic base of the seven populations.

Pop29 had the highest gene diversity *D* followed by Pool24 and Pop21 (Table 3). This is consistent with pedigree information (Table 1) because the populations have been established using a wide range of germplasm. The lowest *D* value observed for Pop43 is also in accordance with its pedigree, because it was generated from 16 S_1 lines including only Tuxpeño germplasm. Ranking of the populations based on *D* was almost identical with their ranking based on the average number of alleles per marker (rank correlation $r_s = 0.93$; $P < 0.01$). Altogether, the high percentage (89.8%) of the molecular variance revealed by the AMOVA (Table 5) within populations is in harmony with the broad genetic base of the materials used for their synthesis (Table 1). Since related germplasm such as various sources from Tuxpeño or ETO entered different populations, it was also not surprising to find only a minor variance between populations (Table 5). A more detailed analysis of the population subdivision with test statistics of the AMOVA was not possible, because this would require knowledge of the gametic phase for linked loci (Michalakis and Excoffier, 1996), which cannot be determined from SSR analyses of heterozygous individuals.

Correlation between MRD^2 and Panmictic Midparent Heterosis

We investigated the correlation between PMPH and MRD^2 because quantitative genetic theory suggests a linear relationship between both measures under certain assumptions (Falconer and Mackay, 1996, p. 255). This is in harmony with related studies on midparent heterosis in crosses of inbred lines (see Melchinger et al., 1991; Boppenmaier et al., 1993), where the commonly employed Roger's distance (1972) is equal to MRD^2 (Melchinger, 1993). A low correlation between PMPH and MRD^2 can be attributable to several causes: (i) a poor association between heterozygosity estimated from marker data and heterozygosity at quantitative trait loci affecting the trait examined, (ii) a poor association between heterozygosity and heterosis at quantitative trait loci in the crosses examined (Charcosset et al., 1991), (iii) existence of multiple alleles (Cress, 1966), and (iv) epistasis (Moll et al., 1965).

The low correlations between MRD^2 and PMPH for plant height and days to silking were mostly due to small PMPH estimates for these traits (Table 2). By comparison, the corresponding correlation for grain yield was surprisingly high ($r = 0.63$; $P < 0.01$). This is consistent with the relative large contribution of SCA effects to the total sums of squares, which accounted for 33% of the genetic variation among crosses for this trait (Vasal et al., 1992a). In accordance with quantitative genetic theory (Melchinger, 1999) the correlation of MRD^2 was lower with hybrid performance ($r = 0.41$; $P < 0.05$) than with PMPH for grain yield ($r = 0.63$; $P < 0.01$). On the basis of a literature survey with single crosses produced from inbreds, Melchinger (1999) pointed out that only intragroup crosses show a correlation between parental genetic distance and midparent heterosis, but not intergroup crosses. However, a closer examination of the graph between MRD^2 and PMPH (Fig. 3) did not provide any clue in this direction.

While Hypothesis 1 postulates a linear relation between MRD^2 and PMPH, under Hypothesis 2 a quadratic or cubic regression is expected to fit the data better than linear regression. However, in our study neither a quadratic nor a cubic regression model gave a significantly better fit to the data than linear regression. This is in accordance with the graphs shown in Fig. 3. Consequently, our results confirm Hypothesis 1 for the tropical maize germplasm investigated here.

A decrease in PMPH of genetically very distant populations is generally attributed to the lack of coadaption between both allelic and nonallelic combinations of genes from the two parental haploid genomes, resulting in reduced or negative dominance and negative epistatic effects, respectively (Falconer and Mackay, 1996, p. 255). A major reason for the absence of an optimum in the relationship between genetic distance and PMPH in our study could be that all populations (Table 1) were more or less well adapted to the test environments. In addition, we did not include extremely wide crosses, as was the case in the experiment of Moll et al. (1965).

For hybrid breeding, Melchinger and Gumber (1998) recommended the following criteria for the choice of heterotic patterns: (i) high mean performance and large genetic variance in the hybrid population; (ii) high per se performance and good adaption of the parent populations to the target region(s); (iii) low inbreeding depression, if hybrids are produced from inbreds. Under Hypothesis 1 (PMPH increases with increasing genetic distance), genetic distance could be used as a further criterion for the identification of heterotic patterns. Considering all four criteria, the following promising heterotic patterns can be suggested: (i) Heterotic Group A with Heterotic Group B; (ii) Pop43 with Heterotic Group A or B; (iii) Pop29 with Heterotic Group B or Pop43.

Grouping of Germplasm

We chose the MRD as genetic distance measure because of its mathematical and genetic properties. In particular, it is an Euclidean distance, which is an often-overlooked prerequisite for most multivariate analysis methods (Jacquard, 1974, p. 465). Furthermore, in the absence of epistasis and two alleles per locus, PMPH is a linear function of the product of the dominance effect and the square of the MRD (Melchinger, 1999).

Principal coordinate analysis based on MRD revealed very clearly a major split between the populations from

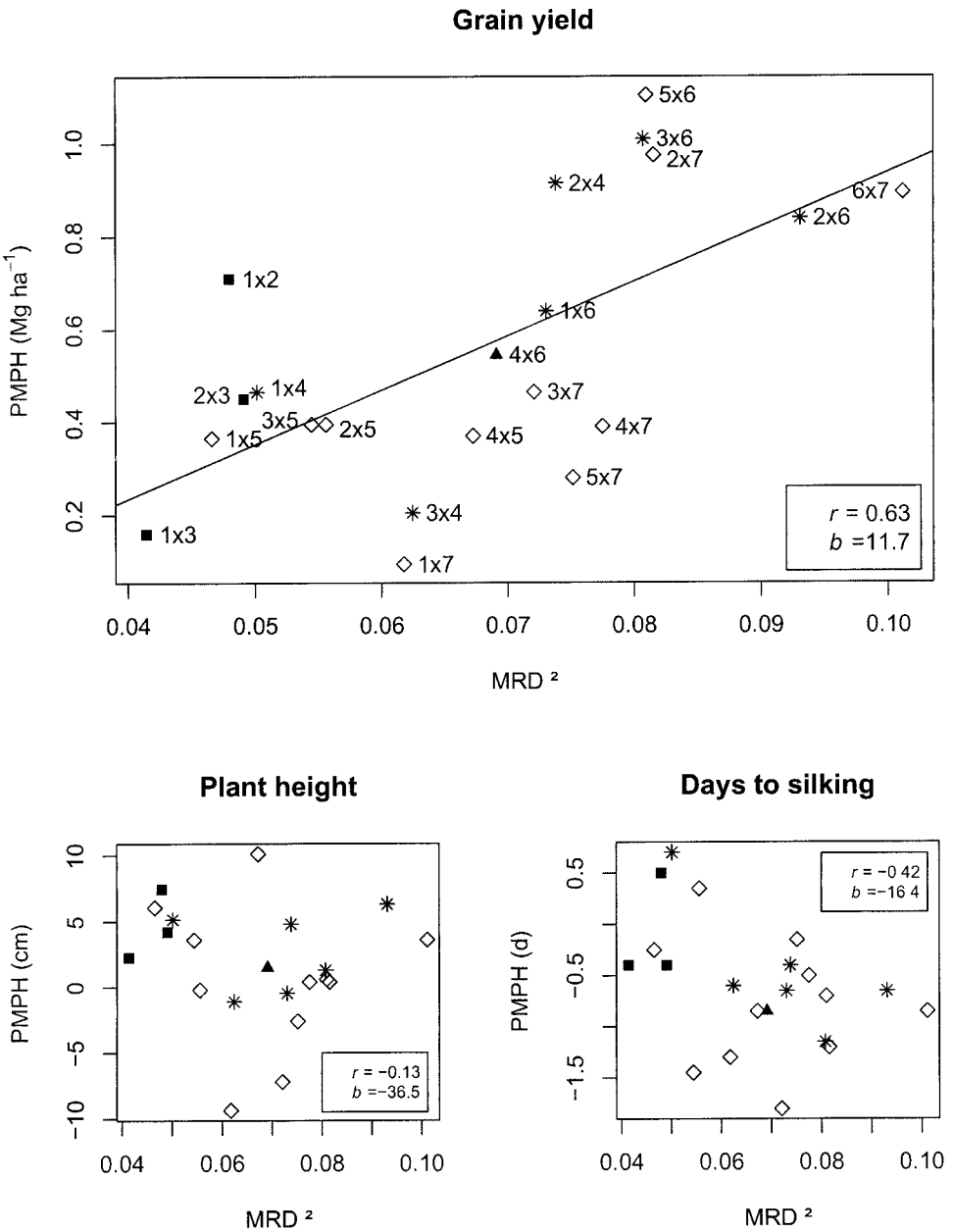

Fig. 3. Relationship between squared Roger's distance (MRD²) and panmictic midparent heterosis (PMPH) for grain yield, plant height, and days to silking. Intrapool crosses within Heterotic Group A (filled squares) and Group B (filled triangles), interpool crosses between A and B (*), and miscellaneous (open diamonds) are shown. 1 = Pool24, 2 = Pop21, 3 = Pop22, 4 = Pop25, 5 = Pop29, 6 = Pop32, 7 = Pop43; *r* is the correlation coefficient and *b* the slope coefficient.

Heterotic Group A and Pop32 (Fig. 1). Pop25 is separated from the other populations by PC3 and had an average MRD at the population level to Heterotic Group A of 0.24 and to Pop32 of 0.26. The assignment of Pop25 to Heterotic Group B together with Pop32 originally based on testcross data was not supported by our molecular data. This could be interpreted as an indicator that Pop25 should have been established as a separate Heterotic Group C. The values of PMPH (Table 2) support this hypothesis in that Pop25 had a low average PMPH with Heterotic Group A. In addition, PCoA accurately portrayed the relationship of Pop43 to Heterotic Group A and B. It is closer to Heterotic Group A ($\overline{MRD}$ = 0.26) than to Heterotic Group B ($\overline{MRD}$ = 0.29), but the distance from Pop43 to Heterotic Group A was higher than the average distance between Heterotic Groups A and B. This together with the diallel analysis suggests classification of Pop43 as a separate Heterotic Group D. According to the PCoAs (Fig. 1, 2), Pop29 could be assigned to Heterotic Group A, because it had a smaller average MRD to Heterotic Group A (0.22) than to B (0.26). The diallel analysis supports this suggestion.

In conclusion, classification of the seven populations based on SSR data mostly confirmed the results from the diallel data except the assignment of Pop25 to Heterotic

Group B. Furthermore, it was possible to assign Pop29 to the established Heterotic Groups A and to propose new heterotic groups (Pop25, Pop43). When a large number of germplasm exists but no established heterotic groups are available, genetically similar germplasm can be identified with molecular markers. On basis of this information, field trials can be planned more efficiently. Thus, by using molecular data to focus the search for heterotic groups on a smaller number of promising heterotic patterns and evaluating these intensively, breeders should arrive at a more economic and solid approach for making this important decision at the beginning of a hybrid breeding program.

CONCLUSIONS

The results of this study suggest that molecular marker-based analyses, and in particular SSR technology, offers a reliable and effective means of assessing genetic diversity within and between maize populations. The AMOVA revealed a high within population variance, as expected from the origin and genetic background of these populations. For the establishment of heterotic groups to be used in hybrid breeding, a higher variance between populations would have been advantageous because this should result in higher PMPH and, consequently, a higher performance of crosses between them.

Simple sequence repeat markers provide a valuable tool for grouping of germplasm and are a good complementation to field trials for identifying groups of genetically similar germplasm. Consequently, field trials for identification of promising heterotic patterns can be planned more efficiently based on prior information obtained from SSR analyses.

ACKNOWLEDGMENTS

The molecular marker analysis of this research were supported by funds from the German "Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung" Project No. 98.7860.4-001-01. J.C. Reif was supported by the "Vater und Sohn Eiselen Stiftung", Ulm, Germany, for carrying out parts of this research at CIMMYT in Mexico. The authors thank three anonymous reviewers for their valuable suggestions.

REFERENCES

Beck, D.L., S.K. Vasal, and J. Crossa. 1990. Heterosis and combining ability of CIMMYT's tropical early and intermediate maturity maize (*Zea mays* L.) germplasm. Maydica 35:279–285.

Boppenmaier, J., A.E. Melchinger, G. Seitz, H.H. Geiger, and R.G. Herrmann. 1993. Genetic diversity for RFLPs in European maize inbreds. III. Performance of crosses within versus between heterotic groups for grain traits. Plant Breed. 111:217–226.

Charcosset, A., M. Lefort-Buson, and A. Gallais. 1991. Relationship between heterosis and heterozygosity at marker loci: A theoretical computation. Theor. Appl. Genet. 81:571–575.

Clarke, B.C., L.B. Moran, and R. Appels. 1989. DNA analyses in wheat breeding. Genome 32:334–339.

Cockerham, C.C., and Z.B. Zeng. 1996. Design III with marker loci. Genetics 143:1437–1456.

Cress, C.E. 1966. Heterosis of the hybrid related to gene frequency differences between two populations. Genetics 53:269–274.

Crossa, J., S.K. Vasal, and D.L. Beck. 1990. Combining ability estimates of CIMMYT tropical late yellow maize germplasm. Maydica 35:273–278.

Crow, J.F. 1948. Alternative hypotheses of hybrid vigor. Genetics 33:477–487.

Crow, J.F. 1999. Dominance and overdominance. p. 49–58. *In* J.G. Coors and S. Pandey (ed.) The genetics and exploitation of heterosis in crops. ASA, CSSA, and SSSA, Madison, WI.

East, E.M. 1936. Heterosis. Genetics 21:375–397.

Falconer, D.S., and T.F. Mackay. 1996. Introduction to quantitative genetics. 4th ed. Longman Group, London.

Gardner, C.O., and S.A. Eberhart. 1966. Analysis and interpretation of the variety cross diallel and related populations. Biometrics 22: 439–452.

Goodman, M.M., and C.W. Stuber. 1983. Races of maize: VI. Isozyme variation among races of maize in Bolivia. Maydica 28:169–187.

Gower, J.C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–388.

Hallauer, A.R., W.A. Russell, and K.R. Lamkey. 1988. Corn breeding. p. 463–564. *In* G.F. Sprague and J.W. Dudley (ed.) Corn and corn improvement. 3rd ed. Agron. Monogr. 18. ASA, CSSA, and SSSA, Madison, WI.

Heckenberger, M., A.E. Melchinger, J.S. Ziegle, L.K. Joe, J.D. Hauser, M. Hutton, and M. Bohn. 2002. Variation of DNA fingerprints among accessions within maize inbred lines with regard to the identification of essentially derived varieties. I. Genetic and technical sources of variation in SSR data. Mol. Breed. 10:181–191.

Ihaka, R., and R. Gentleman. 1996. A language for data analysis and graphics. J. Comput. Graph. Stat. 5:299–314.

Jacquard, A. 1974. The genetic structure of populations. p. 465–466. *In* Biomathematics. Vol. 5. Springer Verlag, Berlin.

Lamkey, K.R., and J.W. Edwards. 1999. Quantitative genetics of heterosis. p. 31–48. *In* J.G. Coors and S. Pandey (ed.) The genetics and exploitation of heterosis in crops. ASA, CSSA, and SSSA, Madison, WI.

Lu, H., and R. Bernardo. 2001. Molecular marker diversity among current and historical maize inbreds. Theor. Appl. Genet. 103: 613–617.

Melchinger, A.E. 1993. Use of RFLP markers for analysis of genetic relationship among breeding materials and prediction of hybrid performance. p. 621–628. *In* D.R. Buxton et al. (ed.) International crop science I. CSSA, Madison, WI.

Melchinger, A.E. 1999. Genetic diversity and heterosis. p. 99–118. *In* J.G. Coors and S. Pandey (ed.) The genetics and exploitation of heterosis in crops. ASA, CSSA, and SSSA, Madison, WI.

Melchinger, A.E., and R.K. Gumber. 1998. Overview of heterosis and heterotic groups in agronomic crops. p. 29–44. *In* K.R. Lamkey and J.E. Staub (ed.) Concepts and breeding of heterosis in crop plants. CSSA Spec. Publ. 25. CSSA, Madison, WI.

Melchinger, A.E., M.M. Messmer, M. Lee, W.L. Woodman, and K.R. Lamkey. 1991. Diversity and relationships among U.S. maize inbreds revealed by restriction fragment length polymorphisms. Crop Sci. 31:669–678.

Michalakis, Y., and L. Excoffier. 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics 142:1061–1064.

Moll, R.H., J.H. Longquist, J.V. Fortuna, and E.C. Johnson. 1965. The relation of heterosis and genetic divergence in maize. Genetics 52:139–144.

Moll, R.H., W.S. Salhuana, and H.F. Robinson. 1962. Heterosis and genetic diversity in variety crosses of maize. Crop Sci. 2:197–198.

SAS Institute. 1988. SAS language guide for personal computers. Release 6.03 ed. SAS Inst., Cary, NC.

Senior, M.L., J.P. Murphy, M.M. Goodman, and C.W. Stuber. 1998. Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. Crop Sci. 38:1088–1098.

Sharon, E.M., S. Kresovich, C.A. Jester, C.J. Hernandez, and A.K. Szewc-McFadden. 1997. Application of multiplex PCR and fluorescence-based, semi-automated allele sizing technology for genotyping plant genetic resources. Crop Sci. 37:617–624.

Smith, J.S.C., E.C.L. Chin, H. Shu, O.S. Smith, S.J. Wall, M.L. Senior, S.E. Mitchell, S. Kresovich, and J. Ziegle. 1997. An evaluation of utility of SSR loci as molecular markers in maize (*Zea mays* L.):

comparisons with data from RFLPs and pedigree. Theor. Appl. Genet. 95:163–173.

Stuber, C.W., S.E. Lincoln, D.W. Wolff, T. Helentjaris, and E.S. Lander. 1992. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823–839.

Vasal, S.K., H. Cordova, S. Pandey, and G. Srinivasan. 1999. Tropical maize and heterosis. p. 363–373. *In* J.G. Coors and S. Pandey (ed.) The genetics and exploitation of heterosis in crops. ASA, CSSA, and SSSA, Madison, WI.

Vasal, S.K., G. Srinivasan, D.L. Beck, J. Crossa, S. Pandey, and C. de Leon. 1992a. Heterosis and combining ability of CIMMYT's tropical late white maize germplasm. Maydica 37:217–223.

Vasal, S.K., G. Srinivasan, J. Crossa, and D.L. Beck. 1992b. Heterosis and combining ability of CIMMYT's subtropical and temperate early maturity maize germplasm. Crop Sci. 32:884–890.

Vasal, S.K., G. Srinivasan, S. Pandey, H.S. Cordova, G.C. Han, and F. Gonzalez. 1992c. Heterotic patterns of ninety two white tropical CIMMYT maize lines. Maydica 37:259–270.

Weir, B.S. 1996. Genetic data analysis II. 2nd ed. Sinauer Associates, Sunderland, MA.

Wright, S. 1978. Evolution and genetics of populations. Vol. IV. Univ. of Chicago Press, Chicago, IL.