Theor Appl Genet (2003) 107:947–957
DOI 10.1007/s00122-003-1333-x

J. C. Reif · A. E. Melchinger · X. C. Xia · M. L. Warburton · D. A. Hoisington · S. K. Vasal · D. Beck · M. Bohn · M. Frisch

Use of SSRs for establishing heterotic groups in subtropical maize

Received: 18 November 2002 / Accepted: 19 April 2003 / Published online: 27 June 2003

Abstract Heterotic groups and patterns are of fundamental importance in hybrid breeding. The objectives of our research were to: (1) investigate the relationship of simple sequence repeats (SSR) based genetic distances between populations and panmictic midparent heterosis (PMPH) in a broad range of CIMMYT maize germplasm, (2) evaluate the usefulness of SSR markers for defining heterotic groups and patterns in subtropical germplasm, and (3) examine applications of SSR markers for broadening heterotic groups by systematic introgression of other germplasm. Published data of two diallels and one factorial evaluated for grain yield were re-analyzed to calculate the PMPH in population hybrids. Additionally, 20 pools and populations widely used in CIMMYT's breeding program were assayed with 83 SSR markers covering the entire maize genome. Correlations of squared modified Roger's distance (MRD^2) and PMPH were mostly positive and significant, but adaption problems caused deviations in some cases. For intermediate- and early-maturity subtropical germplasm, two heterotic groups could be suggested consisting of a flint and dent composite. We concluded that the relationships between the populations obtained by SSR analyses are in excellent agreement with pedigree information. SSR markers are a valuable complementation to field trials for identifying heterotic groups and can be used to introgress exotic germplasm systematically.

Keywords Heterotic groups · SSRs · Heterosis · Mega-environment · Genetic distance

Communicated by F. Salamini

J. C. Reif · A. E. Melchinger (✉) · M. Frisch
Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany
e-mail: melchinger@uni-hohenheim.de
Fax: +49-711-4592343

M. L. Warburton · D. A. Hoisington · S. K. Vasal · D. Beck
CIMMYT, International Wheat and Maize Improvement Center. Apdo. Postal 6-641, 06600 Mexico DF, Mexico

X. C. Xia
National Maize Improvement Center of China, China Agricultural University, 100094, Beijing, China

M. Bohn
Crop Science Department, Universiy of Illinois, 1102 South Goodwin Avenue, Urbana, IL 61801, USA

Introduction

Recognition of heterotic patterns among genetically divergent groups of germplasm is fundamental in hybrid breeding for maximum exploitation of heterosis (Hallauer et al. 1988). Lamkey and Edwards (1999) coined the term panmictic midparent heterosis (PMPH) for the difference between a hybrid population and the mean of its two parent populations in Hardy-Weinberg equilibrium. Under the assumptions of two alleles per locus and no epistasis, PMPH is a function of the dominance effect at each locus and the square of the difference in allele frequency between the populations (Falconer and Mackay 1996); the latter corresponds to the square of the modified Roger's distance (MRD^2).

Using the geographic origin as a crude indicator for the genetic distance, Moll et al. (1962) in their study with U.S. maize observed a linear increase in PMPH with increasing genetic distances. In contrast, experimental data reported by Moll et al. (1965) in a study with tropical and U.S. maize populations suggested an increase of PMPH with increasing genetic distance only up to an optimum level, but a decrease in extremely wide crosses. The authors explained this decline by fertility distortion in wide crosses, adaptation problems and epistatic interactions of genes. The relationship between mid-parent heterosis of single-cross hybrids and the genetic distance of their parental inbreds, determined with molecular markers, were investigated both in theory (Charcosset and Essioux 1994) and numerous experiments with maize and other crops (Brummer 1999). Melchinger (1999) pointed out that only intragroup crosses show a correlation between parental genetic distance and midparent hetero-

sis, but for intergroup hybrids, heterosis is at best only loosely correlated with the parental genetic distance.

If heterosis of hybrids increases monotonically with increasing genetic distance of the parents, genetic distances based on molecular markers should be a useful tool for establishing promising heterotic groups and patterns (Melchinger and Gumber 1998). Introgression of exotic germplasm is often suggested for increasing the genetic differences between opposite heterotic populations with an expected increase in heterotic response (Beck et al. 1991; Vasal et al. 1992a, b; Ron Parra and Hallauer 1997).

Over the past 35 years, breeders at the International Maize and Wheat Improvement Center (CIMMYT) have developed numerous germplasm pools, populations, and open-pollinated varieties (OPV) based on mixtures of germplasm originating from various backgrounds (CIMMYT 1998). A series of combining ability studies was conducted to determine heterotic relationships among CIMMYT populations and pools. Several of the populations demonstrated good general combining ability, and various promising heterotic patterns were identified (Crossa et al. 1990; Beck et al. 1991; Vasal et al. 1992a, b). However, no conclusions were drawn about clearly defined heterotic groups. With the establishment of a hybrid breeding program, the question of suitable heterotic groups becomes relevant for subtropical maize germplasm (Vasal et al. 1999).

The objectives of our research were to: (1) investigate the relationship of simple sequence repeat (SSR) based genetic distances between populations and PMPH in a broad range of CIMMYT maize germplasm, (2) evaluate the usefulness of SSR markers for defining heterotic groups and patterns in subtropical germplasm, and (3) examine applications of SSR markers for broadening heterotic groups by systematic introgression of other germplasm.

Materials and methods

For reducing the large collection of germplasm from CIMMYT's gene bank, to a size which can be handled efficiently for breeding purposes, more than 100 populations were established using germplasm from different sources. Additionally, 30 broad-based back-up pools were formed to reduce the danger of narrowing down the genetic basis in tropical and subtropical maize (CIMMYT 1998). We investigated using molecular markers 20 of these pools and populations (further referred to as populations) (Table 1), which had previously been included in published field experiments.

Field experiments

Experiment 1 comprised a complete diallel of five subtropical early-maturity and two temperate populations (Pop46, 48, and Pool27, 28, 30, 40, 42) described in detail by Vasal et al. (1992a). Experiment 2 included a complete diallel of seven intermediate-maturity subtropical populations (Pop33, 34, 42, 45, 47 and Pool31, 34) and two temperate adapted populations (Pool39, 41) published by Beck et al. (1991). Experiment 3 comprised factorial crosses (Design-II, Comstock and Robinson 1948) of four intermediate-maturity subtropical populations (Pop42, Pop45, Pop47, Pool34) mated with four tropical populations (Pop22, Pop25, Pop32, Pop43) described in detail by Vasal et al. (1992c). In addition to the hybrid populations, all parent populations were included in each experiment. Experiment 1 was evaluated in five subtropical (four Mexican, one Turkish) and 17 temperate (16 U.S., one Canadian) environments. Experiment 2 was tested in five subtropical environments in Mexico. Experiment 3 was evaluated in six environments in Mexico and Colombia. The experimental design for the three experiments was a randomized complete block design with three replications in each environment. All crosses in both reciprocal forms were produced at Poza Rica, Mexico, in the 1985 winter season 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 ascertain Hardy-Weinberg equilibrium.

The parents and their crosses were evaluated for grain yield. In the subtropical environments, the experimental unit consisted of two 5-m rows spaced 75 cm, and a plant density of approximately 53,333 plants ha^{-1}. In temperate environments, plot size and plant density varied; at most sites, the experimental unit was two rows either 3.05 or 6.10 m in length, spaced either 0.76 or 0.91 m apart. Final stands ranged from 53,333 to 87,700 plants ha^{-1}. For the subtropical environments all rows were hand-harvested and grain yield (mg ha^{-1}) was calculated at 80% of the ear weight adjusted to 155 g kg^{-1} of moisture. For all temperate environments, plots were machine harvested and shelled grain weight was adjusted to 155 g kg^{-1} of moisture.

SSR analyses

Twenty one randomly chosen individuals from each of the 16 subtropical and temperate populations, and 48 individuals from the four tropical populations, were analyzed separately. DNA was extracted from plants grown from seed increases of the original populations tested in the field trials.

DNA was extracted employing a modified CTAB procedure (Saghai-Maroof et al. 1984). The 83 SSR markers used in the study were chosen from the MaizeDB database (http://nucleus.agron.missouri.edu/cgi-bin/ssr_bin.pl) based on the repeat unit and bin location to provide uniform coverage of the entire maize genome. Primers and PCR conditions were described in detail by Warburton et al. (2002). Briefly, 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 (Elder and Southern 1987); allele identity was assigned using Genotyper 2.1 (Perkin Elmer/Applied Biosystems) and the two inbred lines CML51 and CML292 as a control. Data have been stored in the MaizeDB database (http://nucleus.agron.missouri.edu/cgi-bin/ssr_bin.pl).

Statistical analyses

The three experiments were analyzed separately. Analyses of variance (ANOVA) for grain yield were computed for each mega-environment (ME) separately (Experiment 1: subtropical and temperate MEs; Experiment 2: subtropical ME; and Experiment 3: tropical, subtropical, and transition/mid-altitude MEs). Analyses III of Gardner and Eberhart (1966) were carried out for Experiment 1 and 2 and a Design II analysis (Comstock and Robinson 1948) for Experiment 3.

Entry mean squares were tested for significance by *F*-tests by using the corresponding entry × environment mean squares. Entry × environment mean squares were tested for significance by using the pooled error mean square. PMPH of each cross was calculated as the difference between the F_1 mean and the respective midparent mean for each ME.

Table 1 Description of the 20 CIMMYT maize populations used in this study

Population/Pool	Cycle	Experiment	Germplasm description
Tropical			
Pop22	6	3	Includes Tuxpeño and ETO Blanco germplasm, and germplasm from Central America
Pop25	0	3	Is composed of white flint selections from crosses among germplasm from Mexico, Columbia, the Caribbean, Central America, India, Thailand and the Philippines
Pop32	5	3	Is based on white flint germplasm from South America, Cuba, Mexico and the U.S.Cornbelt
Pop43	5	3	Is a Tuxpeño synthetic composed of 16 S_1 lines
Subtropical intermediate-maturity			
Pop33	2	2	Contains mainly Argentinian (Cateto) flints
Pop34	5	2	Includes Cuban flints, ETO, Tuxpeño, and germplasm from the U.S. Cornbelt, India and Nepal
Pop42	4	2 and 3	Is an advanced generation of ETO selected for short-plant type and crossed with Illinois Cornbelt components
Pop45	3	2 and 3	Includes U.S. Cornbelt germplasm, Tuxpeño, Cuban flints, Puerto Rico composite, and collections from the Dominican Republic
Pop47	2	2 and 3	Consists largely of Tuxpeño germplasm plus some U.S. Cornbelt lines
Pool31	14	2	Is a broadbased pool including white flint segregates from Ecuador, Argentina, India, Mexico, Pool32, and Pool33, but contains also germplasm from Mexico, U.S. Cornbelt, Brazil, Uruguay, Argentina, China, Pakistan, Yugoslavia, Lebanon, Guatemala, Venezuela, Peru, Cuba, and the Dominican Republic
Pool34	20	2 and 3	Includes germplasm from the Mexican lowlands and highlands, the U.S. Cornbelt, southern USA, Puerto Rico, Pakistan, Hungary, China, Peru, Pakistan, Lebanon, Nicaragua and Guatemala
Subtropical early-maturity			
Pop46	1	1	Represents a superior flint fraction (240 half-sib families) of Pool 29, which is based on germplasm from Europe, Lebanon, U.S. Cornbelt, China, Indonesia and South America.
Pop48	5	1	Is composed of dents from U.S. Cornbelt germplasm, southern European germplasm and 54 half-sib families from Pool 30
Pool27	20	1	Includes flint germplasm from the USA, China, Lebanon, Pakistan and several European countries
Pool28	14	1	Is based on crosses between white dent segregates from Pool27 and Hungarian germplasm from Pool 30, and various other germplasms
Pool30	15	1	Made up of dent germplasm from Europe, China, Lebanon, Mexico, South America and the U.S. Cornbelt
Temperate			
Pool39	12	2	Contains germplasm from the tropical lowlands and highlands, subtropical and temperate areas
Pool40	12	1	Is based on germplasm from Europe
Pool41	12	2	Includes predominantly U.S. Cornbelt germplasm plus germplasm from China, Korea and Lebanon
Pool42	12	1	Is based on germplasm from Mexico, Peru, Bolivia,Pakistan, Hungary, the USA and Yemen

We calculated the modified Roger's distance (MRD) between two populations (Wright 1978, pp 91; Goodman and Stuber 1983) as:

$$\mathrm{MRD} = \sqrt{\frac{1}{2m}\sum_{i=1}^{m}\sum_{j=1}^{a_i}(p_{ij} - q_{ij})^2}. \quad (1)$$

Here, p_{ij} and q_{ij} are the allele frequencies of the jth allele at the ith marker in the two populations under consideration, a_i is the number of alleles at the ith marker, and m refers to the number of markers. Standard errors of MRD estimates were obtained by using a bootstrap procedure with re-sampling over markers and individuals within populations. Following Melchinger et al. (1990), the squared modified Roger's distance (MRD^2) was partitioned into general ($GMRD^2$) and specific squared modified Roger's distances ($SMRD^2$) analogous to the subdivision of agronomic data into GCA and SCA effects. Pearson correlation coefficients (r) were calculated for MRD^2 and $SMRD^2$ with F_1 performance, PMPH and SCA effects. Significance tests of r were performed by using tabulated values based on Fisher (1921) z transformation. The polymorphic-index content (PIC) for each SSR marker was determined as described by Smith et al. (1997).

A principal coordinate analysis (PCoA, Gower 1966) was calculated separately for each experiment based on the matrix of MRD values. Heterotic groups were defined by using the k-means clustering algorithm (Hartigan and Wong 1979), which assigns populations to k clusters such that the within-cluster sum of squares is minimized. The predefined number k of clusters was choosen based on: (1) pedigree information, (2) information from breeders, and (3) the results from PCoA. All analyses were carried out with the

Table 2 Means (above diagonal) and panmictic midparent heterosis (PMPH, below diagonal) for grain yield in different mega-environments (ME) and modified Roger's distance (MRD) between populations (above diagonal) and their standard error (SE, below diagonal) of seven CIMMYT's maize populations and their crosses evaluated in Experiment 1

Pop.	Pop46	Pop48	Pool27	Pool28	Pool30	Pool40	Pool42
	Subtropical ME Mg ha^{-1}						
per se	4.50	4.69	4.88	4.99	4.41	3.73	3.26
Pop46		4.89	4.82	4.95	5.17	4.33	4.32
Pop48	0.29		5.42	5.26	4.93	4.45	4.40
Pool27	0.13	0.64		4.92	5.18	4.25	4.37
Pool28	0.21	0.42	–0.02		5.15	4.18	4.39
Pool30	0.72	0.38	0.54	0.45		4.46	4.38
Pool40	0.22	0.24	–0.05	–0.18	0.39		3.80
Pool42	0.44	0.43	0.30	0.27	0.55	0.31	*0.42*[a]
	Temperate ME Mg ha^{-1}						
per se	3.70	4.93	3.80	4.30	4.95	3.80	3.45
Pop46		4.98	4.06	4.28	4.88	3.88	3.90
Pop48	0.67		4.82	5.03	4.96	4.56	4.38
Pool27	0.31	0.46		4.45	4.70	4.23	4.10
Pool28	0.28	0.42	0.40		4.71	4.27	4.17
Pool30	0.56	0.02	0.33	0.09		4.70	4.24
Pool40	0.13	0.19	0.43	0.22	0.33		3.66
Pool42	0.33	0.19	0.48	0.30	0.04	0.04	*0.34*[a]
	MRD (above diagonal) and SE (below diagonal)						
Pop46		0.294	0.234	0.226	0.248	0.247	0.239
Pop48	0.022		0.301	0.257	0.224	0.247	0.256
Pool27	0.024	0.024		0.220	0.269	0.253	0.239
Pool28	0.021	0.023	0.024		0.217	0.214	0.220
Pool30	0.021	0.022	0.022	0.022		0.213	0.232
Pool40	0.021	0.020	0.023	0.021	0.020		0.222
Pool42	0.021	0.023	0.021	0.024	0.022	0.022	

[a] LSD (0.05) of the means

Plabsim software (Frisch et al. 2000), which is implemented as an extension to the statistical software R (Ihaka and Gentleman 1996).

An analysis of molecular variance (AMOVA) (Michalakis and Excoffier 1996) was performed to divide the molecular genetic variance into components attributable to the variance between and within populations using the software package Arlequin (Schneider et al. 2000).

Results

For all three experiments highly significant ($P < 0.01$) differences among the entries, parents, crosses, and parents vs crosses were observed in all MEs (Beck et al. 1991; Vasal et al. 1992a, c). GCA effects were highly significant ($P < 0.01$) in all cases except the transition/mid-altitude MEs in Experiment 3. SCA effects were significant ($P < 0.05$) in Experiment 1 for the temperate ME and highly significant ($P < 0.01$) in Experiment 3 for the subtropical ME.

Experiment 1

Average grain yield in the subtropical ME ranged for the parent populations from 3.26 Mg ha^{-1} (Pool42) to 4.99 Mg ha^{-1} (Pool28) and for the crosses from 3.80 Mg ha^{-1} (Pool40 × Pool42) to 5.42 Mg ha^{-1} (Pop48 × Pool27) (Table 2). PMPH for grain yield was maximum in Pop46 × Pool30 (0.72 Mg ha^{-1}) and minimum in Pool28 × Pool40 (–0.18 Mg ha^{-1}). In the temperate ME, average grain yields for parents and their crosses were 4.13 Mg ha^{-1} and 4.43 Mg ha^{-1}, respectively. Pool30 (4.95 Mg ha^{-1}) and Pop48 (4.93 Mg ha^{-1}) had the highest grain yields among the parents. Average grain yields for the crosses ranged from 3.66 Mg ha^{-1} (Pool40 × Pool42) to 5.03 Mg ha^{-1} (Pop48 × Pool28). PMPH ranged from 0.02 Mg ha^{-1} (Pop48 × Pool30) to 0.67 Mg ha^{-1} (Pop46 × Pop48) and averaged 0.29 Mg ha^{-1}.

Experiment 2

Average grain yield in the subtropical ME ranged from 4.61 Mg ha^{-1} (Pool41) to 7.21 Mg ha^{-1} (Pop42) for the parents and from 4.91 Mg ha^{-1} (Pool39 × Pool41) to 7.87 Mg ha^{-1} (Pop42 × Pop47) for the crosses (Table 3). PMPH averaged 0.38 Mg ha^{-1} with a maximum of 0.92 Mg ha^{-1} (Pop33 × Pop45) and a minimum of –0.09 Mg ha^{-1} (Pop45 × Pool39).

Experiment 3

Average grain yield for the crosses ranged from 5.62 Mg ha^{-1} (Pop32 × Pool34) to 7.43 Mg ha^{-1} (Pop43 × Pop42) for tropical ME, from 6.11 Mg ha^{-1} (Pop25 × Pool34) to 8.03 Mg ha^{-1} (Pop22 × Pop42, Pop43 × Pop42) for subtropical ME, and from 6.23 Mg ha^{-1} (Pop32 × Pool34) to 7.97 Mg ha^{-1} (Pop22 × Pop47) for the transition/mid-altitude ME (Table 4). PMPH averaged 0.59, 0.78 and 0.53 Mg ha^{-1} for the tropical, subtropical and transition/mid-altitude MEs, respectively.

Table 3 Means (above diagonal) and panmictic midparent heterosis (PMPH, below diagonal) for grain yield in the temperate mega-environment and modified Roger's distance (MRD) between populations (above diagonal) and their standard error (SE, below diagonal) of nine CIMMYT's maize populations and their crosses evaluated in Experiment 2

Pop.	Pop33	Pop34	Pop42	Pop45	Pop47	Pool31	Pool34	Pool39	Pool41
	Mg ha^{-1}								
per se	5.77	6.60	7.21	6.36	7.01	6.11	6.19	5.16	4.61
Pop33		6.77	6.89	6.98	6.96	6.12	6.17	5.64	5.64
Pop34	0.59		7.40	7.13	7.16	6.64	7.13	6.34	6.08
Pop42	0.40	0.50		7.47	7.87	7.03	7.13	6.38	6.57
Pop45	0.92	0.65	0.69		7.03	6.32	6.62	5.67	5.42
Pop47	0.57	0.36	0.76	0.35		7.06	6.87	6.31	6.47
Pool31	0.18	0.28	0.37	0.09	0.50		6.37	5.94	5.86
Pool34	0.19	0.74	0.43	0.35	0.27	0.22		5.79	5.62
Pool39	0.18	0.46	0.19	–0.09	0.23	0.31	0.11		4.91
Pool41	0.45	0.48	0.66	–0.07	0.66	0.50	0.22	0.03	*0.67*[a]
	MRD (above diagonal) and SE (below diagonal)								
Pop33		0.244	0.264	0.237	0.257	0.251	0.242	0.228	0.256
Pop34	0.024		0.236	0.292	0.268	0.272	0.281	0.277	0.305
Pop42	0.021	0.021		0.284	0.281	0.278	0.270	0.272	0.278
Pop45	0.021	0.023	0.022		0.273	0.245	0.223	0.229	0.230
Pop47	0.022	0.023	0.021	0.021		0.261	0.261	0.276	0.289
Pool31	0.026	0.027	0.025	0.027	0.027		0.269	0.264	0.278
Pool34	0.024	0.024	0.021	0.021	0.022	0.026		0.230	0.260
Pool39	0.021	0.021	0.020	0.019	0.021	0.024	0.022		0.212
Pool41	0.023	0.021	0.020	0.020	0.021	0.024	0.023	0.020	

[a] LSD (0.05) of the means

Table 4 Means and panmictic midparent heterosis (PMPH) for grain yield in different mega-environments (ME) and modified Roger's distances (MRD) between populations and their standard error (SE) of tropical × subtropical crosses and parents evaluated in Experiment 3

Pop.	Pool34	Pop42	Pop45	Pop47	*per se*	Pool34	Pop42	Pop45	Pop47
	Grain yield (Mg ha^{-1})					PMPH (Mg ha^{-1})			
Tropical ME									
Pop22	6.14	6.65	6.26	6.06	6.65	0.90	0.61	0.45	0.11
Pop25	5.64	6.79	6.28	6.45	6.27	0.59	0.94	0.66	0.69
Pop32	5.62	5.66	6.45	6.06	6.13	0.64	–0.13	0.90	0.37
Pop43	6.34	7.43	6.93	6.35	7.25	0.80	1.09	0.82	0.10
per se	3.84	5.44	4.97	5.25	*0.74*[a]				
Subtropical ME									
Pop22	7.38	8.03	6.99	7.27	7.56	1.18	0.68	0.38	0.07
Pop25	6.11	7.70	6.63	7.31	6.77	0.30	0.75	0.41	0.51
Pop32	6.87	7.09	7.17	7.12	6.00	1.45	0.52	1.34	0.71
Pop43	7.16	8.03	7.43	7.27	6.73	1.37	1.10	1.23	0.49
per se	4.85	7.14	5.67	6.83	*0.66*[a]				
Transition/mid-altitude ME									
Pop22	6.50	6.45	6.65	7.97	6.83	0.42	–0.41	0.69	1.34
Pop25	6.38	6.51	6.64	7.01	6.46	0.49	–0.17	0.86	0.57
Pop32	6.23	7.21	6.83	7.06	6.73	0.20	0.40	0.92	0.48
Pop43	6.68	7.64	6.55	7.06	6.71	0.66	0.84	0.65	0.49
per se	5.33	6.89	5.10	6.43	*0.92*[a]				
MRD/SE									
	MRD					SE			
Pop22	0.274	0.308	0.299	0.298		0.022	0.021	0.021	0.021
Pop25	0.290	0.300	0.294	0.284		0.024	0.023	0.024	0.023
Pop32	0.295	0.278	0.321	0.307		0.024	0.022	0.022	0.019
Pop43	0.324	0.326	0.327	0.316		0.025	0.021	0.024	0.023

[a] LSD (0.05) of the means

SSR marker data

The 83 SSR primers generated a total of 641 alleles in the 528 genotypes analyzed. The number of alleles per marker across the 20 populations was on average 7.7 and ranged from 2 to 17. PIC values for the SSR loci ranged from 0.10 to 0.85, with an average of 0.60. MRD between pairs of populations for Experiment 1, 2 and 3 averaged 0.241, 0.260, 0.303, and ranged from 0.213 (Pool30 × Pool40) to 0.301 (Pop48 × Pool27), 0.212 (Pool39 × Pool41) to 0.305 (Pop34 × Pool41), and 0.274 (Pop22 × Pool34) to 0.326 (Pop43 × Pop42), respectively (Tables 2, 3 and 4).

Table 5 Analysis of molecular variance (AMOVA) of the populations from the three experiments based on 83 SSR markers

Source of variation	*df*	Sum of Squares	Variance components	Percentage of variation
	Experiment 1			
Among populations	6	399.8	1.2	6.3
Within populations	285	5,000.5	17.5	93.7
Total	*291*	*5,400.3*	*18.7*	*100.0*
	Experiment 2			
Among populations	8	581.8	1.3	7.0
Within populations	369	6,474.4	17.5	93.0
Total	*377*	*7,056.2*	*18.8*	*100.0*
	Experiment 3			
Among populations	7	1,302.2	10.4	11.6
Within populations	544	10,274.5	22.3	88.4
Total	*551*	*11,576.7*	*32.8*	*100.0*

Table 6 Correlations of squared modified Roger's distance (MRD^2) and specific squared modified Roger's distance ($SMRD^2$) based on 83 SSR markers obtained for the parent populations in maize with various parameters (Y) from the analyses of generation means of the grain yield data for different mega-environments of three experiments

Parameter Y	Experiment					
	1 ST[a]	1 TR[a]	2 ST[a]	3 TR[a]	3 ST[a]	3 TM[a]
	r(MRD^2, Y)					
F_1 performance	0.43*	0.40	0.47**	0.64**	0.37	0.14
SCA effects[b]	0.34	0.28	0.35*	0.29	0.18	–0.07
PMPH[c]	0.37	0.56**	0.53**	0.33	0.43	0.18
	r($SMRD^2$, Y)					
F_1 performance	0.16	0.16	0.12	0.28	0.17	–0.08
SCA effects[b]	0.55**	0.47*	0.45**	0.51	0.31	–0.12
PMPH[c]	0.32	0.36	0.34*	0.40	0.19	–0.09

*, ** Significant at the 0.05 and 0.01 levels of probability, respectively
[a] TR, ST and TM refers to tropical, subtropical, and transition mid-altitude mega-environments, respectively
[b] Specific combining ability
[c] PMPH is the panmictic midparent heterosis

PCoA was peformed separately for each experiment (Fig. 1). In Experiment 1, principle coordinate (PC) 1 clearly separated: (1) Pool27 and Pop46, from (2) Pool30 and Pop48, whereas Pool28, Pool40 and Pool42 were positioned in between these two groups. In Experiment 2, PC1 separated: (1) Pop34 and Pop42, from (2) Pop33, Pop45, Pool34, Pool39 and Pool41. PC2 separated these two groups from Pop47 and Pool31. The populations investigated in Experiment 3 formed two clearly separated clusters: (1) Pop22, Pop25, Pop32 and Pop43, and (2) Pop42, Pop45, Pop47 and Pool34.

For all three experiments the AMOVA revealed only a small proportion ($\leq$11.6%) of the molecular variance among populations and the major proportion within populations (Table 5).

Correlations of MRD^2 and $SMRD^2$ with F_1 performance, SCA effects and PMPH estimated from the field data, were positive except for Experiment 3 in the transition/midaltitude ME (Table 6). SCA effects were more closely correlated with $SMRD^2$ than MRD^2. In contrast, PMPH was more closely related with MRD^2 than $SMRD^2$ and highly significant ($P < 0.01$) in two instances (Fig. 2).

Discussion

For hybrid breeding, Melchinger and Gumber (1998) recommended the following criteria for the choice of heterotic patterns: (1) high mean performance and large genetic variance in the hybrid population; (2) high per se performance and good adaption of the parent population to the target region(s); and (3) low inbreeding depression, if hybrids are produced from inbred lines. The main focus of this study was to investigate the use of SSR markers for the grouping of germplasm and the identification of promising heterotic patterns before evaluating the germplasm in intensive field trials.

Descriptive statistics

In this study, we found on average across the 20 populations a higher number of alleles per marker (7.7) than reported by Lu and Bernardo (2001) investigating 40 U.S. inbred lines with 83 SSR markers (4.9), and Senior et al. (1998) evaluating 94 elite U.S. maize inbreds with 70 SSR markers (5.0). This can be explained by the broad germplasm base captured in the 20 populations and the diverse origin of their ancestors (Table 1). In contrast to

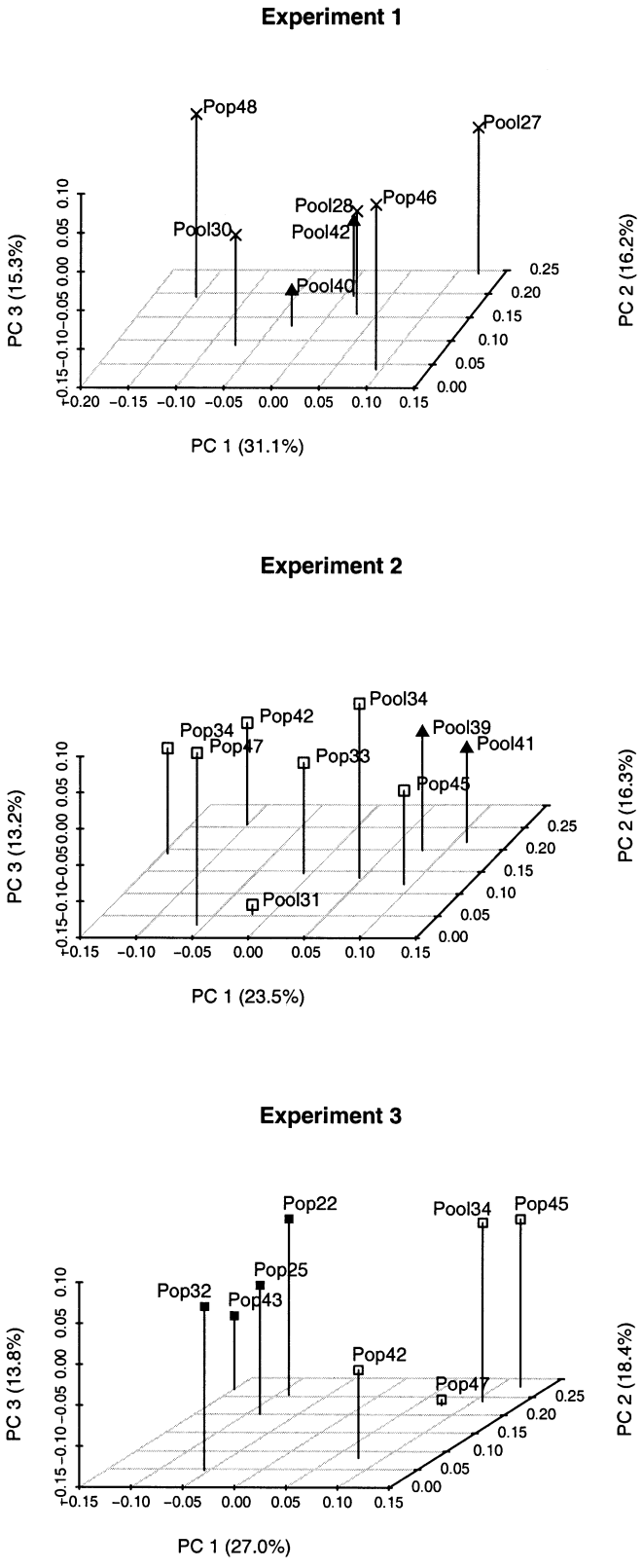

Fig. 1 Principal coordinate analysis based on the modified Roger's distance (MRD) between the populations (tropical ■, subtropical intermediate-maturity ❑, subtropical early-maturity ×, and temperate ▲ populations). PC1, PC2 and PC3 are the first, second and third principal coordinates, respectively

the high number of alleles per marker in our study, the average PIC value (0.60) was similar to those reported by Smith et al. (1997) (0.62) and Senior et al. (1998) (0.59). This can be explained by a high number of rare alleles in our study. The high within population variance revealed in the AMOVA (Table 5) can be explained by the high number of populations with a mixed origin (Table 1).

Correlation between MRD^2, $SMRD^2$ and PMPH, SCA and F_1 performance

We investigated the correlation between PMPH and MRD^2, because quantitative genetic theory suggests a linear relationship between both measures under simplifying assumptions (Falconer and Mackay 1996, pp 255). This is in harmony with related studies on mid-parent heterosis in crosses of inbred lines (see e.g., Melchinger et al. 1991; Boppenmaier et al. 1993), where the commonly employed Roger's distance is equal to MRD^2 (Melchinger 1993). A high correlation between PMPH and MRD^2 can be expected if: (1) a high association exists between heterozygosity at the marker loci and heterozygosity at quantitative trait loci (QTL), (2) heterozygosity at QTL is closely related to heterosis (Charcosset et al. 1991), (3) epistasis is absent, and (4) the populations are adapted to target environments (Moll et al. 1965).

In agreement with this expectation we found in Experiment 1 for the temperate ME a highly significant (r = 0.56**) correlation between PMPH and MRD^2 (Fig. 2). The relatively low correlation (r = 0.37) between both measures in the subtropical MEs can be explained by the non-significant SCA effects, adaption problems of crosses with the two temperate pools, and multiple alleles (Cress 1966). Nevertheless, for both MEs PMPH increased with increasing MRD^2. For Experiment 2, we observed a highly significant correlation between PMPH and MRD^2 (r = 0.53**). Here, a greater number of populations was adapted to the ME than in Experiment 1. The low correlations between both measures observed in Experiment 3 for all three MEs could be attributable to adaption problems of the parent and hybrid populations. The correlation of MRD^2 and PMPH in most experiments and MEs were higher than the correlation of MRD^2 and F_1 performance (Table 6), which is in accordance with the expectations from quantitative genetic theory (Charcosset and Essioux 1994).

To improve the low correlation of MRD^2 and SCA (Table 6), we partitioned MRD^2 into $GMRD^2$ and $SMRD^2$. Under the assumption of no epistasis and by using the parameter definitions of Gardner and Eberhart (1966), SCA can be shown to be a linear function of the $SMRD^2$ for the underlying QTL, provided all QTL have equal dominance effects (Melchinger et al. 1990). In accordance with these quantitative genetic expectations, SCA was in all instances more closely correlated with $SMRD^2$ than MRD^2 (Table 6).

The results of the first two experiments suggest that PMPH and its major component SCA increase with

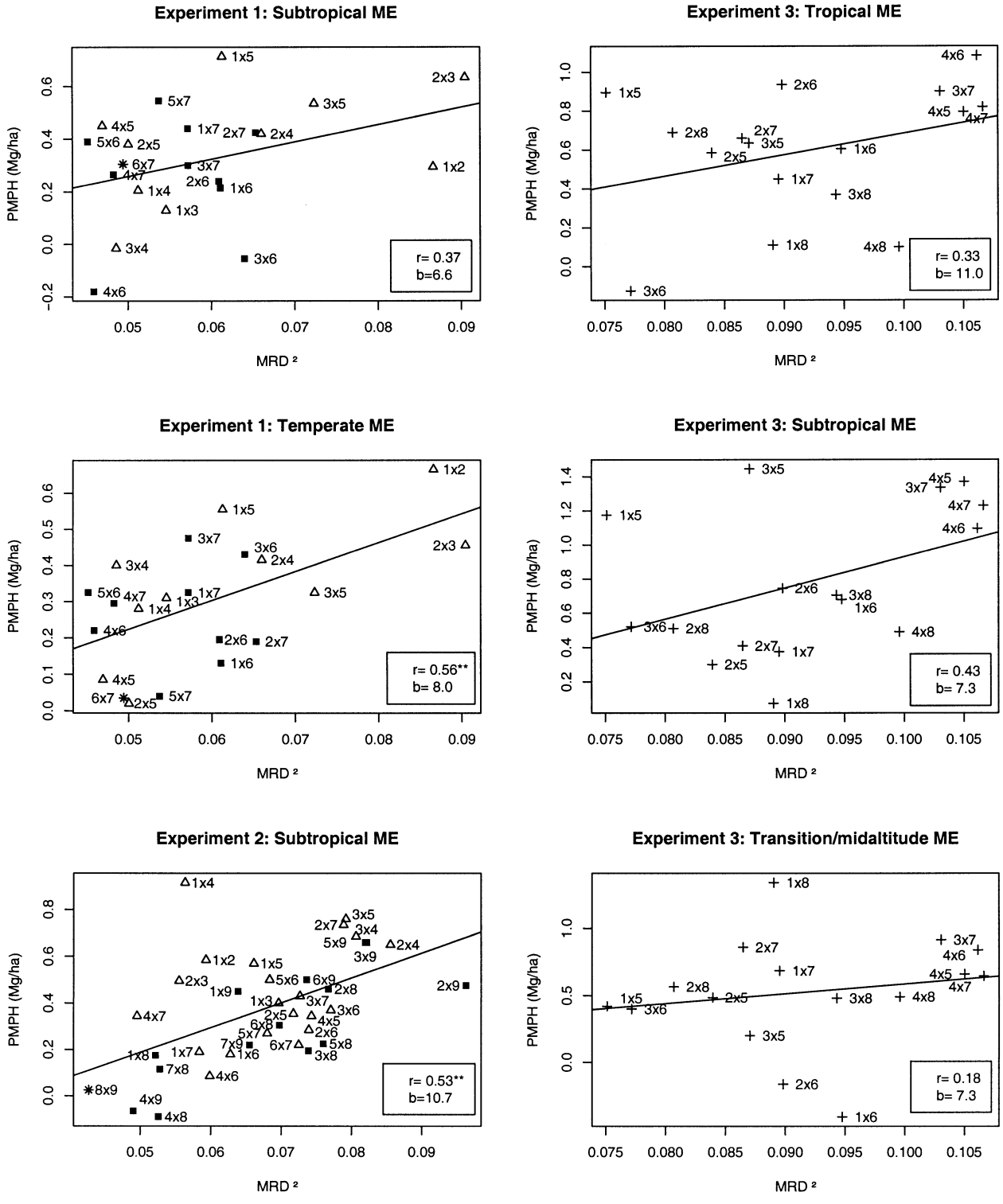

Fig. 2 Relation between squared modified Roger's distance (MRD^2) and panmictic midparent heterosis (*PMPH*) of grain yield for Experiment 1, 2, and 3 (** indicates significance at $P = 0.05$) evaluated in different mega-environments. Crosses between subtropical adapted populations ▲, between subtropical and temperate populations ■, between temperate populations *, and between subtropical × tropical populations +. Experiment1: 1 = Pop46, 2 = Pop48, 3 = Pool27, 4 = Pool28, 5 = Pool30, 6 = Pool40, 7 = Pool42; Experiment2: 1 = Pop33, 2 = Pop34, 3 = Pop42, 4 = Pop45, 5 = Pop47, 6 = Pool31, 7 = Pool34, 8 = Pool39, 9 = Pool41; Experiment3: 1 = Pop22, 2 = Pop25, 3 = Pop32, 4 = Pop43, 5 = Pool34, 6 = Pop42, 7 = Pop45, 8=Pop47

increasing genetic distance among the parent populations. Experiment 3 shows that adaption problems can cause deviations from this rule. Hence, if the populations are adapted to the target regions, genetic distance can be used as a further criterion in the search for promising heterotic patterns and groups.

Heterotic groups and patterns

Wellhausen (1978) described several heterotic patterns and identified four outstanding racial complexes: (1) Tuxpeño and related dents (Mexican, West Indian, Cuban, and Southern U.S. dents), (2) Cuban flints, (3) Coastal Tropical flints (Carribean flint), and (4) Cateto flint. He suggested to form two separate heterotic groups in the CIMMYT maize germplasm: (1) a dent composite, consisting of Tuxpeño and related dents, and (2) a flint composite consisting mainly of Cuban, Carribean and Cateto flints. However, instead of establishing two heterotic groups, CIMMYT maize breeders formed populations and pools mostly disregarding the natural heterotic patterns, which exist between the flint and dent germplasm complexes (Vasal et al. 1999), because this strategy seemed promising for breeding of OPVs. Nevertheless, some populations with a relatively pure genetic background are available (Table 1).

With the beginning of the hybrid development effort, CIMMYT conducted in the 1980s several diallel studies with different germplasm sources to detect heterotic patterns in the germplasm with mixed origin (Crossa et al.1990; Beck et al. 1991; Vasal et al. 1992a, b). Although promising heterotic patterns were suggested, it was too difficult to clearly define heterotic groups on the basis of the field data. This stimulated us to perform a combined analysis with field and molecular data for obtaining a clearer picture on promising heterotic patterns and groups.

Subtropical early-maturity germplasm

Under the preassumption of two groups, the k-means algorithm arrived for the subtropical early-maturity germplasm at the following subdivision: (1) Pool27, Pop46 and Pool28, and (2) Pool30 and Pop48. However, in the PCoA (Fig. 1) Pool28 was positioned midway between Pool27, Pop46 and Pool30, and Pop48, in accordance with pedigree information. Pop46 and Pool27 were both established using flint germplasm from the U.S., Lebanon and several European countries. Pool27 also contains white flints from Argentina. Pop48 was generated from 54 half-sib families of Pool30, which was established using dent germplasm from Europe, China, Lebanon, South America and the U.S. Cornbelt. In contrast, Pool28 was developed by mixing dent and flint germplasm from Pool30 and Pool27, respectively, which precludes their use for hybrid breeding.

Considering the field and molecular data, two heterotic groups could be formed in the subtropical early-maturity germplasm: (1) a flint composite consisting of Pop46 and Pool27, and (2) a dent composite consisting of Pop48 and Pool30.

Subtropical intermediate-maturity germplasm

With k = 3, the k-means algorithm based on MRD resulted in the following subdivision for the intermediate-maturity subtropical germplasm: (1) Pop34 and Pop42, (2) Pop33, Pop45 and Pool34, and (3) Pop47 and Pool31. These results are in accordance with the pedigree information. Pop42 and Pop34 contain ETO germplasm. The latter includes also Cuban flints and Tuxpeño germplasm. Pop33 was established using Cateto flints. Pop45 contains Cuban flints, but also Tuxpeño and a large diversity of other germplasm. Pop47 was established using 276 half-sibs of Pool32, which was established using germplasm from the same sources as Pool31. The mixed origin of Pop34, Pop45 and Pool31, Pool34 precludes their use for hybrid breeding. Hence, considering the molecular and field data two heterotic groups can be formed in the subtropical intermediate-maturity germplasm: (1) a flint composite consisting of Pop33 and Pop42, and (2) a dent composite consisting of Pop47.

In conclusion, SSR based technology offers a powerful tool for assessing the diversity among maize populations. The relationships between the populations obtained by using MRD and PCoA are in excellent agreement with the pedigree information. SSR based genetic distances in combination with field evaluation provide a solid basis for the detection of promising heterotic groups and patterns at the beginning of a hybrid breeding program.

Systematic introgression of exotic germplasm for hybrid breeding

With the increasing germplasm exchange between tropical, subtropical and temperate areas, greater options of germplasm sources are available for breeders. For hybrid breeding one has to consider the racial complexes and relationships between the populations to introgress exotic germplasm systematically in the existing heterotic groups. We investigated the use of SSR markers to achieve this goal. Considering the maturity type of the germplasm introgressed, we propose an exchange between early tropical and late subtropical, early subtropical and late temperate, germplasm and vice versa.

Pool42 was established to introduce tropical germplasm into temperate areas. The low hybrid performance of crosses with Pool42 in Experiment 1 (Table 2) suggested that it may not be of direct use for breeding programs in temperate environments. Pool39, 40 and 41 were designed to introgress temperate germplasm for the winter maize areas in the subtropics and tropics. Similar results were observed as for Pool42 (Tables 2, 3), which

indicated that they may not be valuable for breeding programs in subtropical environments. The low hybrid performance of the four pools can be explained by their low *per se* performance in all MEs.

In contrast, the high yield and PMPH of crosses between subtropical × tropical germplasm (Table 4) suggested that the exchange between both types of germplasm could benefit CIMMYT's hybrid breeding program. Aggregating all information about the relationships between the populations (Fig. 1 and Table 1) and considering the field data (Table 4), we propose an exchange of germplasm between both ETO-based Pop32 and Pop42 on one side, and the largely Tuxpeño-based Pop22 and Pop47 on the other side. Furthermore, genotypes with rare or absent SSR marker alleles in the other group and good test performance can be identified and used to systematically broaden the germplasm basis. Thus, useful alleles can be introgressed and benefit the respective breeding programs.

Acknowledgements The molecular marker analyses of this research were supported by funds from the German "Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung" Project No. 98.7860.4-001-01. Thanks to J. Crossa, G.C. Han, S. Pandey, and G. Srinivasan for providing the field data and seeds for performing this study. This paper is dedicated to Prof. Dr. h. c. F. W. Schnell on the occasion of his 90th birthday.

References

Beck DL, Vasal SK, Crossa J (1991) Heterosis and combining ability among subtropical and temperate intermediate-maturity maize germplasm. Crop Sci 31:68–73

Boppenmaier J, Melchinger AE, Seitz G, Geiger HH, Herrmann RG (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

Brummer EC (1999) Capturing heterosis in forage crop cultivar development. Crop Sci 39:943–954

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

Charcosset A, Essioux L (1994) The effect of population structure on the relationship between heterosis and heterozygosity at marker loci. Theor Appl Genet 89:336–343

CIMMYT (1998) A complete listing of maize germplasm from CIMMYT. Maize Program Special Report, Mexico DF, Mexico

Comstock RE, Robinson HF (1948) The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. Biometrics 4:254–266

Cress CE (1966) Heterosis of the hybrid related to gene-frequency differences between two populations. Genetics 53:269–274

Crossa J, Vasal SK, Beck DL (1990) Combining ability estimates of CIMMYT tropical late yellow maize germplasm. Maydica 35:273–278

Elder JK, Souther EM (1987) Computer-aided analysis of one-dimensional restriction fragments gels. In: Bishop MJ, Rawling CJ (eds) Nucleid acid and protein sequence analysis – a practical approach. IRL Press, Oxford, pp 165–172

Falconer DS, Mackay TF (1996) Introduction to quantitative genetics, 4th edn. Longman Group Ltd, London

Fisher RA (1921) On the "probable error" of a coefficient of correlation deduced from a small sample. Metron 1:1–32

Frisch M, Bohn M, Melchinger AEM (2000) Plabsim: software for simulation of marker-assisted backcrossing. J Hered 91:86–87

Gardner CO, Eberhart SA (1966) Analysis and interpretation of the variety cross diallel and related populations. Biometrics 22:439–452

Goodman MM, Stuber CW (1983) Races of maize. VI. Isozyme variation among races of maize in Bolivia. Maydica 28:169–187

Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–338

Hallauer AR, Russell WA, Lamkey KR (1988) Corn breeding. In: Sprague GF, Dudley JW (eds) Corn and corn improvement, 3rd edn. Agron Monogr 18, ASA, CSSA and SSSA, Madison, Wisconsin, pp 463–564

Hartigan JA, Wong MA (1979) A K-means clustering algorithm. Appl Stats 28:100–108

Ihaka R, Gentleman R (1996) A language for data analysis and graphics. J Computat Graphical Stats, Vol 5 3:299–314

Lamkey KR, Edwards JW (1999) Quantitative genetics of heterosis. Chapter 10. In: Coors JG, Pandey S (eds) The genetics and exploitation of heterosis in crops. CSSA, Madison, Wisconsin

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

Melchinger AE (1993) Use of RFLP markers for analysis of genetic relationship among breeding materials and prediction of hybrid performance. In: International Crop Science I. CSSA, Madison, Wisconsin, pp 621–628

Melchinger AE (1999) Genetic diversity and heterosis. Chapter 10. In: Coors JG, Pandey S (eds) The genetics and exploitation of heterosis in crops. CSSA, Madison, Wisconsin

Melchinger AE, Gumber RK (1998) Overview of heterosis and heterotic groups in agronomic crops. In: Lamkey KR, Staub JE (eds) Concepts and breeding of heterosis in crop plants, CSSA, Madison, Wisconsin, pp 29–44

Melchinger AE, Lee M, Lamkey KR, Woodman WL (1990) Genetic diversity for restriction fragment length polymorphisms: relation to estimated genetic effects in maize inbreds. Crop Sci 30:1033–1040

Melchinger AE, Messmer MM, Lee M, Woodman WL, Lamkey KR (1991) Diversity and relationships among US maize inbreds revealed by restriction fragment length polymorphisms. Crop Sci 31:669–678

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

Moll RH, Salhuana WS, Robinson HF (1962) Heterosis and genetic diversity in variety crosses of maize. Crop Sci 2:197–198

Moll RH, Longquist JH, Fortuna JV, Johnson EC (1965) The relation of heterosis and genetic divergence in maize. Genetics 52:139–144

Ron Parra J, Hallauer AR (1997) Utilization of exotic maize germplasm. Plant Breed Rev 14:165–187

Saghai-Maroof MA, Soliman KM, Jorgenson R, Allward RW (1984) Ribosomal DNA spacer length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. Proc Natl Acad Sci USA 81:8014–8018

Schneider S, Roessli D, Excoffier L (2000) Arlequin, ver 2.0: a software of population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland

Senior ML, Murohy MM, Goodman MM, Stuber CW (1998) Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. Crop Sci 28:63–67

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovitch S, Ziegle J (1997) An evaluation of the 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

Vasal SK, Srinivasan G, Crossa J, Beck DL (1992a) Heterosis and combining ability of CIMMYT's subtropical and temperate early maturity maize germplasm. Crop Sci 32:884–890

Vasal SK, Srinivasan G, Beck DL, Crossa J, Pandey S, Leon C de (1992b) Heterosis and combining ability of CIMMYT's tropical late white maize germplasm. Maydica 37:217–223

Vasal SK, Srinivasan G, Gonzalez F, Han GC, Pandey S, Beck DL, Crossa J (1992c) Heterosis and combining ability of CIMMYT's tropical × subtropical maize germplasm. Crop Sci 32:1483–1489

Vasal SK, Cordova HS, Pandey S, Srinivasan G (1999) Tropical maize and heterosis. Chapter 34. In: Coors JG, Pandey S (eds) The genetics and exploitation of heterosis in crops. CSSA, Madison, Wisconsin

Warburton ML, Xia XC, Crossa J, Franco J, Melchinger AE, Frisch M, Bohn M, Hoisington D (2002) Genetic characterisation of CIMMYT maize inbred lines and open-pollinated populations using large scale fingerprinting methods. Crop Sci (in press)

Wellhausen EJ (1978) Recent developments in maize breeding in the tropics. In: Walden DB (ed) Maize breeding and genetics. John Wiley and Sons, New York, pp 59–91

Wright S (1978) Evolution and genetics of populations, vol. IV. The University of Chicago Press