UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

CHAPTER 4

ASSOCIATION BETWEEN AFLP- BASED GENETIC DISTANCE AND HYBRID PERFORMANCE IN THE TANZANIAN MAIZE INBREDS

The contents of this Chapter have been published in the *Plant Breeding Journal*

Manuscript no W240 Kiula et al

Accepted 21 May 2007

Abstract

The identification of best inbred combinations for the development of commercial hybrid maize varieties is still the main challenge to maize breeders. Several strategies including crossing parents from different heterotic groups, pedigrees, and/or testers have been attempted. The main pitfall of these methods is that they are mainly based on phenotypic data, which are confounded by environmental variations. Thus the main aim of this study was, therefore, to study the associations between amplified fragment length polymorphisms (AFLPs) based genetic distance (GD) and F_1 phenotypic data. Furthermore, the efficacy of grouping lines into genetically similar clusters was investigated. This research used 21 inbreds, which were fingerprinted using AFLP analysis and their 210 F_1 progenies were planted in the field. Results of the analysis were applied to predict the best line combinations for commercial maize hybrid production. Joint data analyses revealed a tighter association between GD and the F_1 traits or mid parent heterosis (MPH) in the intergroup than in the intragroup crosses. Intergoup crosses should be field tested before their release. Intragroup and intergroup crosses showing low GD-MPH might be discarded to avoid field-testing costs. Better F_1 hybrid performance predictions can be achieved by integrating molecular and F_1 phenotypic data.

Keywords: Genetic distance, intragroup, intergroup, maize, phenotypic traits, heterosis.

Introduction

Genetic diversity among breeding materials is very important in plant breeding, as it largely determines the future prospects of the breeding program (Hallauer *et al.*, 1988). Before 1970, genetic diversity in crop plants was determined based on pedigree data (Lübberstedt *et al.*, 2000) and morphological traits (Yee *et al.*, 1999). To date, some of the genetic diversity study methods applied in crops include Malecot's (1948) coefficient of coancentry (*f*). This method however, is not commonly used because of several assumptions that are not always fulfilled (Messmmer *et al.*, 1993). Distance from geographic origin (Hallauer *et al.*, 1988; Moll *et al.*, 1965) has also been used to compare genetic distances between genotypes. The problem with the distance from geographic origin approach is that more germplasm from one geographic area are mixed or contaminated. Lastly, tester lines A and B is the most commonly and widely used method to assign lines into heterotic groups. But some of the pitfalls of the testers are that (a) some lines show heterosis when crossed to both testers, (b) many lines fall into more than one heterotic group, (c) the two testers are not sufficient to represent the genetic diversity present in the maize heterotic groups, and (d) including more testers can be expensive and time consuming (Warburton *et al.*, 2002).

In crop breeding programs, information on genetic relationships within species is used to organize germplasm collections, identify heterotic groups and to facilitate selections of breeding materials (Lee, 1995; Karp *et al.*, 1996). Several studies have tried to study and relate hybrid performance in maize using phenotypic markers and quantitative traits (Goodman, 1972). This was followed by genetic markers in the form of gene products such as isozyme diversity (Cox *et al.*, 1988; Melchinger *et al.*, 1990) and storage proteins (Cox *et al.*, 1988). However, the use of isozymes in these studies had limitations. The major limitations of the isozyme studies could be due to the low number of loci examined and low number of polymophisms detected (Melchinger *et al.*, 1990; Senior *et al.*, 1998). Direct DNA markers (Genetic markers) such as restriction fragment length polymorphisms (RFLPs) in these studies were suggested and applied to overcome the constraints associated with isozymes (Burr *et al.*, 1983).

The choice of inbred lines as sources of favourable genes for hybrid production and performance is of major concern to breeders. Different germplasm sources are screened for the presence of favourable genes. Pedigree, morphology and agronomic characters have been used to establish the description of a genotype (Jarman and Pickett, 1992; Martin *et al.,* 1998). However, phenotypic data may poorly reflect actual levels of genetic diversity among genotypes (Yee *et al.*, 1999). This is because such methods are subjective and are often influenced by environmental conditions (Russell *et al.*, 1997; Rafalski *et al.,* 1996). There is also greater level of heterosis expressed when genetically dissimilar parents are crossed (Martin *et al.*, 1998; Cheres *et al.*, 2000).

The DNA marker systems are useful tools for assessing genetic diversity among germplasm groups. The development of molecular markers started with the RFLP marker system in the 1980's (Bostein *et al.*, 1980). RFLP markers were regarded as the first shot in the genome revolution (Dodgson *et al.,* 1997). This technique, however, is not commonly used as it is too complex, require large amount of DNA and relatively more expensive than polymerase reaction (PCR) based (Tragoonrung *et al.*, 1992) marker systems. The PCR based marker systems like AFLP and simple sequence repeats have provided new tools for heterosis prediction and to investigate correlations between parental GDs and the performance of F_1 hybrids or MPH. AFLPs have been tested and confirmed as being high quality markers, highly reproductive and highly informative (Vos *et al.*, 1995; Liu and Cordes, 2004).

Information about the relationship among breeding materials and the genetic diversity in the available germplasm is important for the choice of parents in the plant breeding programs which apply particularly to hybrid breeding, where recognition and exploitation of heterotic patterns between different types of germplasm are important for success. How divergent are the lines from the same and different germplasm pools is one of the most fundamental questions for breeders.

The associations between genetic distances of maize inbred lines and the performance of F_1 phenotypic traits of agronomic importance including GLS scores, where a score of one refers to maize genotypes which are highly resistant to the GLS disease,

whereas nine corresponds to maize genotypes that highly succumbed to the GLS pathogen (Donahue *et al.,* 1991). These factors in maize are important driving forces for the production of commercial hybrid varieties. The relationships of these factors as determinants in maize breeding have not been studied or described in detail, neither have their applications for maize improvement in breeding programs. Also, this study assessed the efficacy of AFLP markers for grouping lines into genetically similar clusters and tested the genetic diversity of these lines in the intragroup and intergroup crosses. Furthermore, pairwise genetic distances of the inbred lines were compared, genetic distances of the lines with performance of phenotypic traits of their crosses were correlated, and finally, the intra and intergroup crosses were characterized in order to generate important information which could help maize breeders identify best line combinations for the production of commercial maize hybrid varieties.

Materials and methods

Plant material and Field design

Twenty-one inbred lines from the High Altitude Maize Improvement Program of Tanzania, (HAMIPT), which is located in the Mbeya region, in the Southern Highlands zone of the country (Table 1) were used in this study. The twenty-one lines were crossed in a complete diallel mating design. The field work was conducted for three consecutive years (2003-2005) across three locations (nine environments). There were 225 hybrids in total (210 single crosses, two controls and 13 filler materials) which were planted using a 15 x 15 triple lattice, and the inbred lines were planted in a randomized complete block design with three replications at each site. Planting was done in the mid December every year in all the sites. The plot size was a single row plot, 5.1 m long with inter row spacing of 0.75 m and intra row spacing of 0.30 m. Two kernels were planted in each hill, thereafter plants were thinned to one plant per hill, and this resulted in a plant population size of 44,400 plants/ha.

Morphological traits

The performance of F_1 phenotypic traits recorded included: yield, days to 50 % silking, height to ear, ear length, kernels/ear, rows/ear and GLS score. The grain

yield/plot was measured in terms of kilograms/ha at grain moisture content of 12.5 percent. Days to 50 % silking was obtained by counting number of days from the day of planting to the day when 50 % of the plants in each plot have silks. Height to ear was measured as the average height in centimeters from the ground level to the first ear in each plot. Ear length was measured as the average length in centimeters of all ears in each plot. The number of rows/ear was obtained by counting the number of rows/ear for all ears in a plot. The average number of rows/ear was taken as the overall number of rows/ear of that plot. The number of kernels/row was obtained by counting the number of kernels/row for all ears of a plot. The average number of kernels/row was taken as the overall number of kernels of that plot. Gray leaf spot scoring was recorded on a scare of 1-9 according to Donahue *et al.* (1991) with an increment of 0.25. A genotype with a score of one was highly resistant, while a score of nine was highly susceptible.

Furthermore, this study used 45 intergroup crosses (five crosses from each of the following line combinations: CML11 x Population 62 lines, CML11 x K530 lines, CML11 x K3758 lines, CML37 x Population 62 lines, CML37 x K3758 lines, Ecuador 509642 line x Population 62 lines, Ecuador 509642 line x K530 lines and Ecuador 509642 line x K3758 lines) and 45 intragroup crosses (15 intergroup crosses from each of the following line x line crosses; Population 62 x Population 62 lines, K530 x K530 lines and K3758 x K3758 lines) to study the associations between the estimated GD of lines (Table1) and the performance of F_1 phenotypic traits of line x line pairwise crosses (Tables 5 and 6).

Genotyping

The twenty-one inbred lines from the HAMIPT were analysed genetically using AFLP fingerprinting. Three seeds of each inbred line were sown in a plastic pot containing approximately 1.5 kg mixture of four parts sterilized clay soil: and 1 part sandy soil (4:1). Approximately 120 grams of a compound fertilizer of N: P: K (2: 3: 2) was applied in each plastic pot. This pot trial was conducted at the University of Pretoria Experimental farm during the 2000/2001-rain season. At the three to four leaf stage

(when plants were about two weeks old), the youngest fresh leaves of each line were harvested and immediately put on ice for DNA extraction.

DNA Extraction

Genomic DNA was extracted from leaves of each of the lines as described by Doyle and Doyle (1987). Total genomic DNA was extracted from approximately 100 mg of leaf tissue using 5 % (w/v) Cetyltrimethyl ammonium bromide (CTAB) [0.1 M Tris-HCL (pH 8.0), 1.4 M NaCl, 20 mM EDTA (pH 8.0), 1 % (w/v) soluble Polyvinylpyrrolidone (PVP) and 0.2 % (v/v) 2-Mercaptoethanol). Genomic DNA of each inbred was precipitated by using either ice-cold isopropanol or 95 % (v/v) ethyl alcohol. The genomic DNA pellets were dissolved in either 100 μl double distilled water or in low TE buffer [10 mM Tris-HCL (pH 8.0), 1 mM EDTA] and then stored at –20 ^{O}C. Genomic DNA was quantified spectrophotometically.

DNA analysis

The genetic analyses used the AFLP protocol (Vos *et al.*, 1995) with minor modifications. Genomic DNA of each maize inbred line (approximately 250ng of DNA) was digested at 37 ^{O}C for 2h using 1.25 U *Eco*RI and *Mse*I restriction endonucleases. Following adapter ligation, preselective and selective amplification reactions were performed. The preselective amplifications were carried out in a 35 μl volume containing 10 μl diluted ligation product, AFLP pre-amplification primer mix (10 μM *Eco*RI and 10 μM *Mse*I primers, 2.5 mM dNTPs each), 10x PCR buffer, 1.5 mM $MgCl_2$, 1 U Taq DNA polymerase (Promega Molecular Chemicals) for 30 cycles with the following cycle profile: adapter extension for 10 s at 72 ^{O}C followed by denaturation for 10 s at 94 ^{O}C, annealing for 30 s at 56 ^{O}C, and extension for 1 min at 72 ^{O}C with a 1 s per cycle increase in extension time. The final selective stage had the *Eco*RI/*Mse*I extensions of: ACA/CCC, ACA/CTG, ACA/CAG, ACA/CCG and ACA/CGC. AFLP fragments were resolved in 8 % long Ranger^{TM} polyacryalamide gels, using the LI-COR IR automated DNA analyser (LI-COR, Lincoln, NE, USA). The gel images were scored in a binary matrix that recorded the presence of bands as a plus (+) and absence of bands as minus (-). Semi automated scoring was performed

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

with SAGA^{MX} software (version 3.23, LI-COR, NE, USA). Scores were also manually edited to make corrections to the automated score where appropriate.

Table 1. Origin, pedigree and heterotic patterns of maize lines used in this study.

No	Line name	Kernel type	GLS Rating	Germplasm source	Pedigree	Established heterotic gro
1	P62145	Flint	1..5	Population 62	$P62s_7145\text{-}95$	A/B
2	K53015213	Flint	3..25	K205 x K230	$K530s_71521\text{-}3$	A
3	K375891	Dent	2..50	K337 x K358	$K3758s_79\text{-}1\text{-}1$	A
4	K37581011	Dent	3..25	K337 x K358	$K3758_{S6}810\text{-}1\text{-}1$	A
5	K5301482	Flint	3..00	K205 x K230	$K530s_514821(98)\text{-}1$	A
6	P627733	Flint	2..00	Population 62	$P62s_677\text{-}3Pap3$	A
7	K3758L36	Dent	2..25	K337 x K358	$K3758s_7L\text{-}3\text{-}6$	B
8	K53014821	Flint	2..25	K205 x K230	$K530s_51482\text{-}1$	A
9	CML37	Dent	2..25	Population 32	Pop32c4Hc128-1-1-B-H5	A
10	P629521	Dent	2.75	Population 62	$P62s_795\text{-}2\text{-}1$	A
11	CML11	Dent	3..00	Population 21	Popc5hc219-3-2-2-3#7-1B-4-10b	A
12	P621621	Flint	1.75	Population 62	$P62s_7162\text{-}1Pap2$	A
13	P621111	Flint	2..50	Population 62	$P62s_711\text{-}1\text{-}1$	A/B
14	P62L50	Flint	2..25	Population 62	$P62S_6$ 50	A
15	P628495	Flint	1.75	Population 62	$P62s_684\text{-}95Blk$	A/B
16	P62103	Flint	8..00	Population 62	$P62103s_86\text{-}95$	A
17	P62101	Flint	3..50	Population 62	$P62s_6101\text{-}95$	A
18	P62148	Flint	2..50	Population 62	$P62s_714\text{-}8$ -95	B
19	E50932815	Dent	2.50	E250 x E393	E5093s5Pap28-1-5	B
20	E5093642	Dent	3.50	E250 x E393	$E5093s_764$	B
21	P621321	Flint	1.75	Population 62	P62s713-2-1	A

E=Ecuador

Data analysis

Phenotypic correlation coefficients (r) between GDs and F_1 phenotypic data were computed using the formula by Dale van Vleck *et al.* (1987). Mid parent heterosis (MPH) and high parent heterosis (HPH) were calculated by the formula of Falconer and Mackay (1996). The experimental F_1 phenotypic grand mean of each trait was compared to the line cross means for producing the phenotypic data matrix (+ and -). When the line cross mean was equal or more than the trait grand mean then a + sign (band presence) plus was assigned to it. Whereas a - minus sign (band absence) was used if the line cross was less than the grand mean (Table 2). Then the phenotypic matrix was added to the inbred line polymorphic bands which were also scored as plus (+) for the presence of the band and minus (-) for the absence of the band and then were converted to 1 or 0 when compiled into a data matrix. The data matrix was used to perform cluster analysis (dendrogram) and pairwise genetic distances of the 21 inbred lines using the NCSS 2001 software program (Jerry, 2001) or the pairwise line

genetic distances were computed within each set of the AFLP marker system using the formula given by Nei and Li (1979), which is $GD_{ij} = (N_i + N_j - 2N_{ij})/(N_i + N_j)$; where GD_{ij} is the genetic distance between two inbred lines i and j. N_{ij} is the number of common bands between line i and j, and N_i and N_j are the total number of bands in line i and j, respectively, with regard to all primer pair combinations of AFLP. Thus, GD reflects the proportion of bands in common between the two inbred lines and it may range from 0 (identical profiles for two lines) to 1 (no shared bands). The "goodness of fit" of the clustering to the data matrix was determined by calculating the cophenetic correlation coefficients between dissimilarity and cophenetic matrix derived from ther dendrogram (Sneak and Sokal, 1973). Polymorphism information content (PIC) which provides an estimate of the discriminatory power of the locus or loci, by taking into account, not only the number of alleles that are expressed, but also the relative frequencies of those alleles was computed for each marker assay unit. PIC values in this study were calculated using the algorithm: PIC or (the diversity index of Nei, 1973) = $1 - \sum_{i=1}^{n} fi^2$ where f is the frequency of ith allele averaged across loci. The PIC also provides information on the informativeness of a marker system. Percentage polymorphism was calculated using the following formula: number of polymorphic bands/total number of bands in that assay unit x 100.

Table 2. The experimental grand means of F_1 phenotypic data of each trait used to produce the F_1 phenotypic data matrix in this study.

Trait	(+) sign	(-)sign
1. GLS	Score ≤ 3.5 GLS rating	Score > 3.5 GLS rating
2. Yield	Score ≥ 4.8 MT	Score < 4.8 MT
3. Days to 50 % silking	Score ≥ 92.7 days	Score < 92.7 days
4. Ear height	Score ≥ 103.5 cm	Score < 103.5 cm
5. Ear length	Score ≥ 18.5 cm	Score < 18.5 cm
6. Rows/ear	Score ≥ 13.2	Score < 13.2
7. Kernels/row	Score ≥ 39.8	Score < 39.8

Results

The degree of polymorphism was measured by the number of polymorphic bands, percentage polymorphism and polymorphic index content (Table 3). The average PIC

value for the five primer combinations collectively was 0.28, respectively. Results showed that the number of fragments (Polymorphic and monomorphic bands) per primer pair combination varied from a minimum of 42 (primer pair combination *EcoR*I-ACA+*Mse*I-CGC) to a maximum of 74 fragments (primer pair combination *EcoR*I-ACA+*Mse*I-CCC). The total number of all the AFLP markers in this study was 278 fragments, of which 238 fragments were polymorphic with the fragment sizes ranged from 50 base pairs to 456 base pairs (Primer pair combination not shown). Furthermore, results showed that the highest PIC value (0.34) was revealed by the primer pair combination *EcoR*I-ACA+*Mse*I-CAG and primer pair combination *EcoR*I-ACA+*Mse*I-CGC exhibited the lowest PIC (0.20) value. The PIC values for the remaining primer pair combinations were as follows: Primer pair combination *EcoR*I-ACA+*Mse*I-CCC was 0.32, primer pair combination *EcoR*I-ACA+*Mse*I-CTG was 0.28 and primer pair combination *EcoR*I-ACA+*Mse*I-CCG was 0.24.

Table 3. Degree of polymorphism and information content for five primer combinations applied to 21 Tanzanian inbred lines.

*EcoR*1	*Mse*1	Total fragments	Polymorphic bands	% Polymorphism	PIC
1. ACA	CCC	74	68	91	0.32
2. ACA	CTG	48	39	81	0.28
3. ACA	CAG	56	44	78	0.34
4. ACA	CCG	58	52	89	0.24
5. ACA	CGC	42	35	83	0.20
Total		278	238	-	-
Average		56	48	82	0.28

% Polymorphism = (Polymorphic bands/ Total fragments) x 100

Pairwise genetic distance estimates for the 21 inbred lines (i.e. 210 F_1 line combinations) ranged from a minimum of 0.30 (intragroup crosses) to a maximum of 0.63 (intergroup crosses) (Table 4). Genetic distance estimates for line combinations within population 62 lines ranged from 0.33 (line 20 x line 16, line 6 x line 12) to 0.51 (line 1 x line 6) with a mean GD of 0.0.42. While the genetic distance estimates for the K3758 line series showed a minimum GD of 0.31 (line 7 x line 10) and a maximum of GD 0.46 (line 3 x line 4). The mean GD for the K3758 line series was 0.35. Furthermore, for the K530 line series, their GD mean was 0.43, with a maximum GD of 0.57 (line 13 x line 14) and a minimum GD of 0.32 (line 18 x line 14). Finally, data in Table 3 exhibited that such as line 9 x line 8; line 11 x line 12; line 13 x line

12; line 14 x line 16; line 17 x line 16; line 5 x line 6; line 1 x line 2; line 21 x line 20 and line 14 x line 15 recorded the highest line x line pairwise GDs. Whereas combinations like line 15 x line 10; line 7 x line 10; line 3 x line 10; line 4 x line 10; line 12 x line 6; line 17 x line 4; line 17 x line 10; line 18 x line 14; line 19 x line 15 and line 20 x line 16 exhibited the lowest line x line pairwise GD in this study.

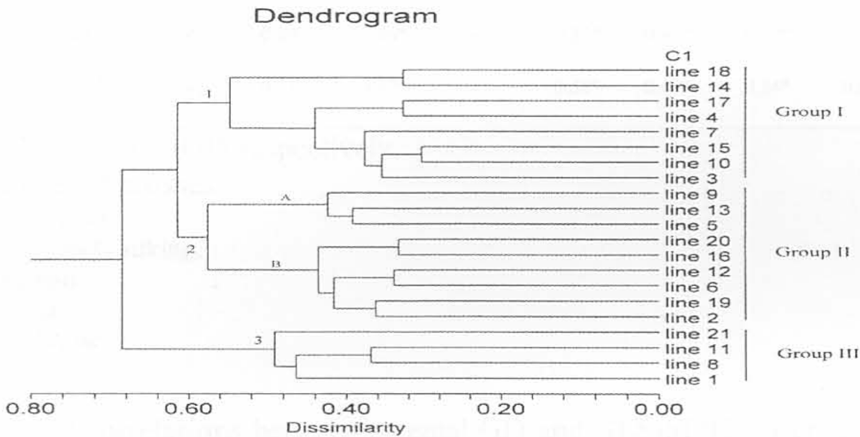
Table 4. Pairwise genetic distance coefficients of lines using five AFLP primer pair combinations on twenty-one lines analyzed by NCSS 2001software program.

	Line 1	Line 2	Line 3	Line 4	Line 5	Line 6	Line 7	Line 8	Line 9	Line 10	Line 11	Line 12	Line 13	Line 14	Line 15	Line 16	Line 17	Line 18	Line 19	Line 20	Line 21
Line1	-	0.61	0.45	0.44	0.47	0.51	0.51	0.49	0.50	0.52	0.38	0.50	0.50	0.46	0.46	0.50	0.44	0.47	0.48	0.53	0.47
Line2		-	0.43	0.46	0.40	0.36	0.42	0.45	0.42	0.42	0.43	0.43	0.42	0.46	0.43	0.37	0.41	0.42	0.36	0.36	0.48
Line3			-	0.46	0.42	0.38	0.40	0.43	0.43	0.33	0.41	0.38	0.39	0.45	0.35	0.39	0.38	0.38	0.39	0.39	0.51
Line4				-	0.51	0.41	0.36	0.36	0.41	0.32	0.36	0.42	0.38	0.41	0.35	0.34	0.32	0.35	0.41	0.35	0.48
Line5					-	0.54	0.50	0.41	0.39	0.40	0.45	0.41	0.39	0.50	0.35	0.44	0.42	0.41	0.39	0.39	0.51
Line6						-	0.47	0.47	0.44	0.36	0.34	0.33	0.48	0.40	0.36	0.37	0.40	0.40	0.39	0.35	0.44
Line7							-	0.52	0.45	0.31	0.38	0.45	0.42	0.39	0.36	0.39	0.40	0.36	0.42	0.41	0.50
Line8								-	0.63	0.43	0.36	0.41	0.41	0.47	0.38	0.43	0.41	0.42	0.45	0.42	0.43
Line9									-	0.52	0.54	0.43	0.43	0.45	0.48	0.43	0.40	0.47	0.40	0.40	0.54
Line10										-	0.47	0.41	0.35	0.41	0.30	0.41	0.32	0.36	0.40	0.34	0.49
Line11											-	0.57	0.43	0.37	0.37	0.41	0.41	0.36	0.42	0.45	0.50
Line12													0.57	0.53	0.39	0.38	0.38	0.42	0.35	0.40	0.41
Line13													-	0.57	0.45	0.39	0.42	0.41	0.48	0.41	0.54
Line14														-	0.54	0.51	0.43	0.32	0.46	0.42	0.57
Line15															-	0.46	0.39	0.38	0.33	0.35	0.43
Line16																-	0.52	0.39	0.42	0.33	0.48
Line17																	-	0.48	0.49	0.35	0.52
Line18																		-	048	0.42	0.50
Line19																			--	0.43	0.46
Line20																				-	0.55
Line21																					-

The dendrogram or genetic similarity or dissimilarity tree produced in this study is presented in Figure 1. This tree revealed three main groups of inbred line-F_1 phenotypic data similarities. Group 1, was comprised of eight line-F_1 phenotypic data in total, of which six lines were K3758 and two lines of K530 series. Group II, however, was the largest one with nine inbred-F_1 phenotypic data. In this group five lines originated from Population 62 germplasm, three lines from K530 inbred series and one CIMMYT line. Furthermore, group II exhibited subgroups, A and B. The subgroup A was comprised of a K530 line, a CIMMYT line and a line from

population 62 germplasm. While subgroup B had five lines from Population 62 and one line belonging to K530 series. Finally, there was group III which had one CIMMYT line one line from K530 series, one line that originated from Population 62 germplasm and the Ecuador line. Inbred lines in group III revealed high genetical diversity as they were from different germplasm sources respectively.

Figure 1. The dendrogram showing the clustering of the 21 inbred lines used in this study.



Results of the associations between the estimated GD of the lines and the performance of F_1 traits of line x line pairwise crosses for the forty five intergroup crosses and the forty five intragoup crosses revealed that yield, GLS score, ear length rows/ear and kernels/row had tighter correlations in the intergroup crosses than in the intragroup crosses (Table 5). But height to ear and days to 50 % silking revealed higher associations in the intragroup crosses than in the intergroup crosses. Furthermore, Table 5 showed tighter correlations between line GD and heterosis for both GLS scores (i.e. GLS-MPH and GLS-HPH), GD and MPH for height to ear, GD and MPH

for days to 50 % silking, GD and MPH for rows/ear and finally GD and MPH for kernels/row in the intergroup than in the intragroup crosses respectively.

Table 5. Correlation coefficients between GD and morphological traits in the intragroup[1] and intergroup[2] crosses and correlation coefficients between GD and their mid parent/high parent heterosis of F_1 morphological traits in both intragroup3 and intergroup crosses[4] in this study.

Type of crosses	#	GD	Yield	GLS	A	B	C	D	E	GLS-HPH	GLS-MPH
Intragroup crosses [1]	45	0.19	0.28*	0.43*	0.59*	0.62*	0.48*	0.27*	0.38*	-	-
Intergroup crosses [2]	45	0.31	0.46*	-0.57*	0.42*	0.51*	0.63*	0.49*	0.56*	-	-
Intragroup Crosses [3]	45	0.19	0.46*	-	0.33*	0.42*	0.69*	0.42*	0.24*	0.21*	0.43*
Intergroup Crosses [4]	45	0.31	0.37*	-	0.51*	0.49*	0.34*	0.68*	0.61*	0.51*	0.48*

* significant at $p = 0.05$ respectively.
# = number of crosses.
A = height to ear.
B = days to 50 % silking.
C = ear length.
D = row/ear
E = kernels/row

The overall correlations between parental GD and GLS-MPH; parental GD and GLS-HPH in maize crop (both inter and intragroup crosses) are exhibited in Table 6. These calculations attempt to elucidate whether MPH or HPH had the tightest association with gray leaf spot resistance in maize. Results indicated that GLS-HPH is more tightly associated with GD than GLS-MPH for GLS resistance. Table 6 also show that correlations between inbred GD and F_1 grain yield, inbred GD and MPH for grain yield. Finally results in this table (Table 6) revealed that r(GD, F_1 grain yield) was 0.39. Both mid parent heterosis (for grain yield) and GLS-MPH exhibited an association of 0.1 with GD of inbred lines, but GLS-HPH showed a tighter correlation coefficient (0.27) with GD of lines.

Table 6. Correlation coefficient r (GD, MPH (yield); GD, GLS-MPH and GD.GLS-HPH.

	F_1 grain yield	MPH (yield)	GLS-MPH	GLS-HPH
r(GD)	0.39*	0.1*	-0.1*	-0.27*

- significant at $p = 0.05$ respectively.

Discussion

In general each of the five primer pair combinations used in this study discriminated effectively most of the unrelated lines. There was a high average percent of polymorphic bands (82 %) per assay unit, and the PIC values were comparable to those of Lübberstedt *et al.* (2000). Similar results were reported for US maize germplasm (Pejic *et al.,* 1998), soybean (Powell *et al*., 1996) and wheat (Bohn *et al*., 1999). The higher PIC values of the AFLP markers in comparison with other marker systems like RFLPs, RAPDs Microsatellites, together with higher reproducibility shows that AFLP markers are a valuable tool for identification of maize inbred lines, plant variety protection and registration as well as germplasm patenting (Melchinger *et al*., 1998; Ajmone-Marsan *et al*., 1998; Lübberstedt *et al*., 2000). AFLPs have other advantages in terms of reliability, reproducibility, discrimination, standardization and cost effectiveness. Furthermore, these results showed that primer pair Eco*R*I-ACA+*Mse*I-CAG was the most discriminative or informative because it revealed the highest PIC value (0.30) and primer combination *EcoR*I-ACA+*Mse*I-CGC was the least discriminative as it had the lowest PIC value (0.22) (Table 3).

The dendrogram in this study revealed that the inclusion of phenotypic data produced different groupings from AFLP genotyping data alone when predicting best line combinations and hybrid heterosis (Kiula *et al*., 2006). This study showed that the grouping of inbred lines plus phenotypic data gave better predictions of hybrid performances since different dissimilar lines were grouped together (Figure 1). This could be based on the genetic theory that maximum heterosis of F_1 hybrids are products of crossing inbred lines which are genetically dissimilar. Hence genetic analyses of inbred lines with phenotypic data improved the predictions on combining ability for the production of commercial hybrid maize varieties.

This research also evaluated the Tanzanian inbred lines as potential parents for developing commercial hybrids. Thus the study included agronomically important traits such as GLS score, days to 50 % silking, kernels/row, rows/ear, height to ear, ear length and grain yield that have not been observed yet in studies in this context before. It is demonstrated that the GDs of genotypes from different germplasm groups

(intergroup crosses) have on average significantly greater GDs than combinations of genotypes from the same germplasm group (intragroup crosses) (Table 5). This observation is consistent with the current genetic theory and those of Melchinger *et al.* (1998) and Pejic *et al.* (1998) who reported that intergroup crosses exhibit higher genetic distances between lines and lower genetic distances for those lines in the intragroup crosses. Results in this study also showed that there are higher correlations between GD and F_1 or MPH data for intergroup crosses than in the intragroup crosses (Table 5). This observation complied with the expectation of Melchinger (1999) know as "group effects". The group effects occurred because both GD and MPH or F_1 phenotypic data were expected to increase from crosses among related lines to intragroup crosses and further to intergroup crosses. In addition to this, an increase in heterosis in intergroup crosses are due to a more favourable ratio of general combining ability (GCA) to specific combing ability (SCA) variances (Dhillon *et al.*, 1990). However, there are correlations between GD and MPH or F_1 in intragroups that were sometimes tighter than in the intergroup (Table 5). Reports of Charcosset *et al.* (1991) and Lübberstedt *et al.* (2000) supported the obtained results and these observations can be explained by the quantitative-genetic theory (Chacosset and Essioux, 1994) which states that in the intergroup crosses, the covariances between the specific genetic distance and specific combining ability (SCA) is a sum of positive and negative terms, which may cancel each other resulting in low or zero correlation between GD and F_1 or MPH. This is because the maternal and paternal gametic arrays may differ in the linkage phase for many quantitative trait loci-marker pairs (Chacosset and Essioux, 1994). The tighter associations of GD with MPH or F_1 in Table 5 in the intragroups than in the intergroup crosses could be due to firstly, the hidden relatedness between some parents considered as being related based on their pedigree data, and secondly because of the presence of some linkage phase between quantitative trait loci (QTLs) and the maternal loci in the maternal and paternal gametic arrays of the intragroup, which results in the positive covariance between GD and F_1 or MPH (Melchinger, 1999), and increases the GD versus F_1 or MPH in the intragroup crosses, respectively.

In both intragroup and intergroup crosses the correlation between GD and F_1 or MPH is expected to increase if (i) QTLs influencing heterosis are closely linked to markers

used in the calculation of GDs; and (ii) markers used to calculate GDs are linked to QTLs. So increasing the marker density alone will not necessarily improve the quality to predict MPH by GD estimates, rather markers must be selected for tight linkage to QTLs affecting heterosis of the target trait in the germplasm under study (Charcosset *et al.,* 1991). However, attempt to identify QTLs-marker associations by using regressing phenotypic values of fixed set of hybrids on the their coded marker genotypes must be regarded with great caution, since Lübberstedt *et al.* (1998) reported that QTLs regions affecting a given trait are not consistent across different germplasm. Results from this study revealed that maximum yields were mainly obtained from intergroup crosses whereas the lowest yields were produced when genetically related lines were crossed. For example, intergroup crosses such as line 9 x line 8 (GD = 0.63), line 1 x line 2 (GD = 0.61) and line 14 x line 13 (GD = 0.57) could produce yields of approximately 9 MT/ha, 7.42 and 6.20 MT/ha, respectively across three different sites, while an intragroup cross such as line 15 x line 10 (GD = 0.30), line 10 x line 4 (GD = 0.32) and line 17 x line 10 (GD = 0.32) produced only approximately 3.15, 2.9 and 2.3 MT/ha in multienvironments (data not shown). This is because an increase in GDs between the two parents will positively increase heterosis for the cross (Cheres *et al*., 2000) or in quantitative genetic theory heterosis is said to be a function of gene frequency differences between the two parents, the level of dominance and epistatic gene effects (Falconer and Mackay, 1996).

According to the quantitative–genetic theory, mid parent heterosis (Falconer and Mackay, 1996) is a linear function of dominace effects and high parent heterosis as a function of both additive and dominance effects. Computations of correlations between GD and GLS-MPH, GD and GLS-HPH in this study showed that there was a tighter association between GD estimated and the HPH than the MPH for GLS resistance (Table 6). Mahn (1977) reported that GLS resistance in maize is controlled by mainly additive genes with narrow sense heritability (h^2) of GLS resistance of 0.33. Narrow sense heritability is a function of additive genes as given by the formula that heritability is the ratio of additive variance over phenotypic variance. Based on this it follows that additive genes contribute to GLS resistance in maize. However, the calculated correlation coefficient between GD and GLS-HPH is greater than the correlation between GD and GLS-MPH. This implies that GLS-HPH mainly

contributes to additive genes for GLS resistance in maize and a small percent by a dominant gene, consistent with other published findings. Lastly, it was also observed that after addition of the absolute values of r(GD, GLS-HPH and GLS-MPH) gives 0.37. This value is also similar to the narrow sense heritability of GLS resistance, which is 0.33. Thus based on this observation it is suggested that the summation of the two correlation coefficients of GLS-HPH and GLS-MPH could potentially be useful in the calculation of narrow sense heritability of GLS resistance in maize.

Finally, it was observed that there was a tighter correlation in intergroup crosses than in the intragroup crosses. Thus, this study therefore recommends and supports that field-testing of particularly intergroup crosses is inevitable in order to confirm these findings. Also, the research strongly supports the idea that tightly linked DNA markers to QTLs which are responsible for F_1 heterosis of the target traits are desperately needed. The identification of such markers from linkage studies could provide accurate relationships between GDs and MPH or F_1 phenotypic data. In this way high yielding intergroup crosses could be easily identified from GD-MPH estimates and would reduce unnecessary field costs like diallel crossing and evaluations of these crosses in multienvironments.

Conclusion

In conclusion, DNA markers such as AFLPs are powerful tools for identification of genetically similar/different genotypes. Marker systems could reveal genetic relationships among breeding materials (and genetic resources) and such markers might assist breeders in the choice of genetically diverse parents for best line combinations to produce maize hybrids of commercial value and monitor the level of genetic diversity in germplasm pools. Finally, heterotic response (and hybrid performance) between these germplasm crosses could be predicted from GDs based on DNA markers and conformations of such predictions need field evaluations especially for intergroup crosses. Intragroup crosses with low GD-MPH could be discarded in order to reduce field costs. Lastly, it is our belief that better F_1 hybrid performance estimations can be achieved from using both molecular and F_1

phenotypic data than using only molecular data like GDs to predict hybrid performances.

Acknolewdgement

I thank the Tanzanian Ministry of Agriculture and the NRF-South Africa for funding this study. I also thank the University of Pretoria and the ARI-Uyole to allow me to use their facilities and Rose Andrews van Zyl for her assistance with the AFLP analysis.

REFERENCES

Ajmone-Marsan, P., P. Castiglion, F. Furasi, M. Kuiper, and M. Motto. 1998. Genetic diversity and relationship to hybrid performance in maize as revealed by RFLP, and AFLP markers. *Theor. Appl. Genet* 96: 219-227.

Burr, B., S.V. Evola, F.A. Burr, and J.S. Beckerman. 1983. The application of restriction fragment length polymorphism to plant breeding. In: J.K. Setlow and A. Hollaender (ed). Genetic principles and methods. Vol.5. Plenum Press London. pp.45-59.

Bohn, M., H.F. Utz, and A.E. Melchinger. 1999. Genetic similarities among winter wheat cultivars determined on the basis of RFLP, AFLPs and SSRs and their use for predicting progeny variance. *Crop Sci* 39: 228-237.

Bostein, D., R.L. White, M. Skolnick, and R.W. Davies. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphism. *Am. J. Hum. Genet* 32: 314-331.

Cox, T.S., J.P. Murphy, and L.G. Harrell. 1988. Isoelectric focusing patterns of kernel isozymes from 80 North American winter wheat cultivars. *Can. J. Plant Sci* 68: 65-72.

Cheres, M.T., J.F. Miller, J.M. Crane, and S.J. Knapp. 2000. Genetic distance as a predictor of heterosis and hybrid performance within and heterotic groups in sunflower. *Theor. Appl. Genet* 100: 889-894.

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

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

Dale van Vleck, L., E. J. Pollack, and E.A. Branford. 1987. Genetic for the animal sciences. W. H. Freeman and Company. New York.

Doyle, J.J. and J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull 19*: 11-15.

Donahue, P.J., E.L. Stromberg, and S.L. Myers. 1991. Inheritance of reaction to gray leaf spot in a Diallel cross of 14 maize inbreds. *Crop Sci* 31: 926-931.

Dodgson, J. B., H.H. Cheng, and R. Okimoto. 1997. DNA marker technology: a revolution in animal genetics. *Poult. Sci* 76: 1108-1114.

Dhillon, B.S., P.A. Gurrath, E. Zimmer, M. Wermke, W.G. Pollmer, and D. Klein. 1990. Analysis of diallel crosses of maize for variation and covariance in agronomic traits at linkage and grain harvests. *Maydica* 38: 245-248.

Falconer, D.S. and T.F.C. Mackay. 1996. Introduction to quantitative genetics. Longman group Ltd, London.

Goodman, M. M. 1972. Distance analysis in biology. *Syst. Zool* 21: 174-186.

Hallauer, A.R., W.A. Russell, and K.R. Lamkey. 1988. Corn breeding. In: G.F. Sprague and J.W. Dudley (ed.) Corn and Corn improvement. 3rd edn. Agron. Monogr. 18. ASA, CSSA and SSSA, Madison, Wisconsin, pp. 453-564.

Jarman, R. J, and A.A. Pickett. 1992. Botanical descriptors of cereal varieties, 1992: supplement. National Institute of Agricultural Botany, Cambridge.

Jerry, H. 2001. Number Crunchier Statistical Systems (NCSS), Statistical software, Kaysville, Utah.

Karp, A., O. Seberg, and M. Buiatti. 1996. Molecular techniques in the assessment of botanical diversity. *Annals of Botany* 78: 143-149.

Kiula, B.A., A-M Botha, and N.G. Lyimo. 2006. The prediction of best line combiners and heterosis in the Tanzanian maize breeding lines through the use of Amplified fragment length polymorpohism. *S. Afric J. of plant and soil* 23 (4): 246-251.

Lee, M. 1995. DNA markers and plant breeding programs. *Adv. Agron*. 35: 265-344.

Liu, Z.L. and J.F. Cordes. 2004. DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* 238: 1-37.

Lübberstedt, T., A.E. Melchinger, S. Fahr, D. Clein, A. Dally, and P. Westhoff. 1998. QTLs mapping in testcrosses of flint lines of maize: III. Comparison across populations for forage traits. *Crop Sci* 38: 1278-1289.

Lübberstedt, T., A.E. Melchinger, C. Duble, M. Vuylsteke, and M. Kuiper. 2000. Relationships among early European maize inbreds: IV. Genetic diversity revealed with AFLP markers and comparison with RFLP, RAPD, and Pedigree data. *Crop Sci* 40: 783-791.

Martin, A.F., R.H. Busch, K. Khan, and L. Huckle. 1998. Genetic diversity and heterosis of spring wheat crosses. *Crop Sci* 38: 1108-1112.

Mahn N. Q. 1977. Inheritance of resistance to gray leaf spot in maize. M.S. Thesis. Virginia Polytechnic Institute and State University, Blacksburg, USA.

Malecot, G. 1948. "Les mathematiques de l Heredite". Masso net Cie, Paris.

Melchinger, A. E. 1999. Genetic diversity and heterosis. Monograph of ASA-CSSA-SSSA. Library of Congress Cat. Number: 99-73185. South Segoe Road, Madson, WI.53711. pp. 99-118.

Melchinger, A.E., M. Lee, R.K. Lamkey, and W.W. Woodman. 1990. Genetic diversity for restriction fragment length polymorphisms: Relation to genetic effects in maize inbreds. *Crop Sci* 30: 1033-1040.

Melchinger, A.E., R. K. Gumber, R.B. Leipert, M. Vuylsteke, and M. Kuiper. 1998. Prediction of test cross means and variances among F_3 progenies of F_1 crosses from test cross means and genetic distances of their parents in maize. *Theor. Appl. Genet* 96: 503-512.

Messmer, M.M., A.E Melchinger, R.G. Hermann, and J. Boppenmaier. 1993. Relationship among early European maize inbreds: II Comparison of pedigree and RFLP data. *Crop Sci* 33: 944-950.

Moll, R. H., J.H. Longuist, J.V. Fortuna, and E.C. Johnson. 1965. The relationship of heterosis and genetic divergence in maize. *Genetics* 52: 139-144.

Nei, M. 1973: Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci* USA. 70: 3321-3323.

Nei, M. and W. H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci* USA. 76: 5269-5273.

Pejic, L., Ajmone-Marsan, P. Morgante, M. Kozumplick, V. Castiglioni, P. Taramino, and M. Motto. 1998: Comparative analysis of genetic similarity among maize inbred lines tested by RFLPs, RAPDs, SSRs and AFLPs. *Theor. Appl. Genet* 97: 1248-1255.

Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, C. Tingey, and A. Rafalski, 1996. The comparison of RFLP, RAPD, AFLP and SSR markers for germplasm analysis. *Mol. Breed* 2: 225-238.

Rafalski, J.A., J.M. Vogel, M. Morgante, W. Powell, C. Andre, and S.V. Tingey, 1996: Generating and using DNA markers in plants. In: B. Birren and E. Lai (edn). Non mammalian Genomic Analysis: A practical guide. Academic Press Inc., San Diego. pp. 75-135.

Russell, J.R., J. D. Fuller, M. Macaulay, B.G. Hatz, A. Jahoor, W. Powell, and R. Wraught. 1997. Direct comparison of levels of genetic variation among barley accessions detected by RFLP, AFLP, and SSR markers for germplasm analysis. *Theor. Appl. Genet* 95: 714-722.

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.

Sneath, P.H.A. and R.R. Sokal. 1973. Numerical taxonomy. W.H. Freeman and Co., San Fransisco.

Tragoonrung, S., V. Kanazin, P. M. Hayes, and T.K. Blake. 1992. Sequence–tagged sites facilitated PCR for barley genome mapping. *Theor. Appl. Genet* 84: 1002-1008.

Vos, P., R. Hogers, M. Bleeker, M. Rejans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 23: 4407-4414.

Warburton, M. L., X. Xianchun, J. Crossa, J. Franco, A.E. Melchinger, M. Frisch, M. Bohn, and D. Hoisington. 2002. Genetic characterization of CIMMYT inbred maize lines and open pollinated populations using large scale fingerprinting methods. *Crop Sci* 42: 1832-1940.

Yee, E., K.K. Kidwell, G. R. Sills, and T.A. Lumpkin. 1999. Diversity among selected *Vigna angularis* (Azuki) accessions based on RAPD and AFLP markers. *Crop Sci* 39: 191-197.