Full Length Research Paper

Genetic diversity in South African maize (*Zea mays* L.) genotypes as determined with microsatellite markers

Mienie, CMS[1]* and Fourie, AP[2]

[1]North West University, Private Bag X6001, Potchefstroom 2520, South Africa.
[2]Agricultural Research Council-Grain Crops Institute, Private Bag X1251, Potchefstroom 2520, South Africa.

One thousand and forty three (1043) maize genotypes including white and yellow maize inbred lines as well as hybrids from the public germplasm collection were characterized with 80 microsatellite markers distributed throughout the genome. A total of 1874 alleles were amplified and used in the genetic diversity analysis. Principal coordinate analysis confirmed the geographical distribution of the breeding lines. Cluster analysis using Rogers distance measures placed the breeding lines in several clusters and corresponded well with known pedigrees. Lines with mixed origin were classified in separate clusters and duplicate entries in the collection were identified. These mixed lines could not be placed in known heterotic groups, but could rather be used to identify new groups to be used in the breeding program. The genetic distances determined in the study can be applied to plan a more focused breeding program.

Keywords: Genetic distances, maize, microsatellites.

INTRODUCTION

Maize is one of the most widely grown crops worldwide and certainly an important staple food in the African diet. It is used as human food, animal feed and provides an important source of income and employment for a large proportion of the population. Maize breeding is monopolized by large multi-national private companies, with little attention given to more specific needs of local markets. The public sector breeding program is essential in breeding and providing varieties for this niche market as well as providing inbred lines to local private companies. Due to ever decreasing research funds, it is however, essential to scale down the amount of testcrosses and the size of current breeding programs.

Several heterotic groups are currently identified and used in the local breeding program. The locally is being used extensively in crosses. The I heterotic group is considered to be the most advantageous for yield adaptation in South Africa. In hybrid combination with the I-group, the US Cornbelt lines, related to the Lancaster line, Mo17, seems to be the preferred choice. Smaller heterotic groups include the K64 group.

Genetic diversity among inbred lines has traditionally been based on morphological data such as endosperm type, pedigree record and amount of heterosis expressed by the hybrid. This method has several limitations as morphological characters do not always reliably portray genetic relationships due to environmental interactions. Testcross designs with numerous testers are extremely expensive, labour intensive and require large-scale field evaluations. Molecular analyses can lower the cost involved in plant breeding. It also provides a means for determining the purity of hybrids produced in the seed industry and can save seed companies and farmers money in terms of ensuring the quality of the seed.

Knowledge of the relationships among lines would also

*Corresponding author. E-mail: charlotte.mienie@nwu.ac.za. Tel: +27182992319.

help identify a set of inbreds that have maximal diversity for the analysis of the effects of genetic background (Liu et al., 2003). The ideal marker system is highly polymorphic, co-dominant, accurate, reproducible, high-throughput and low cost (both in terms of capitalinvestment and cost per assay). For decades, simple sequence repeats (SSR) or microsatellites, have been the genetic markers of choice, because they are economical to score, have high allelic diversity and are usually selectively neutral (Smith et al., 1997). Large numbers of single nucleotide polymorphisms (SNPs) are already available and have been used extensively in genotyping maize. However, it is still not cost effective in South Africa to genotype large numbers of inbreds for the purpose of association analysis and it is thus necessary to identify a set of lines that capture the maximum number of alleles or haplotypes and determine population structure. Hamblin et al. (2007) compared the use of SSR and SNP markers in maize. The different mutational properties of these two classes of markers resulted in differences in heterozygosities and allele frequencies, but SSRs performed better in clustering germplasm into populations. SSRs provided more resolution in measuring genetic distance based on allele-sharing. Their results suggested that large numbers of SNP loci would be required to replace highly polymorphic SSRs in studies of diversity and relatedness (Hamblin et al., 2007).The relationship between mid-parentheterosis 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). Xia et al. (2004) concluded from their study of CIMMYT sub-tropical maize inbreds that the SSR based genetic distances, calculated with a modified Roger's distance measure, in combination with field evaluations, provided a solid basis for the detection of heterotic groups.

The objectives of our research were to: 1) fingerprint local maize genotypes according to the most appropriate and cost-effective procedure for determining genetic distances of breeding lines and 2) compile a database of local maize breeding material according to DNA data.

MATERIALS AND METHODS

Plant material

The pedigrees of hybrids studied in this paper are commercial intellectual property. Therefore, detailed information is not allowed. Seeds from breeding lines and hybrids were obtained from plant breeders at ARC-GCI in Potchefstroom and Cedara, South Africa and planted in small pots for germination. Young leaves from approximately 20 plants from each line were harvested and two bulks made up from equal amounts of leaf material from ten plants each. We assumed that the bulk sample will capture all heterozygosity present in the genotypes. Leaf material was freeze dried and ground to a fine powder.

SSR genotyping

DNA was extracted from each bulk sample using a modified CTAB extraction technique (Saghai-Maroof et al., 1984) and diluted to 50 ng/µl. Microsatellite marker genotyping involved the use of 96 fluorescently labelled SSR selected from published markers (www.agron.missouri.edu/ssr.html) to be evenly distributed throughout the maize genome (Figure 1).

The SSR primers were labelled with four different dyes [6FAM (blue), HEX (green), NED (yellow) and PET (red)] and combined in 12 multiplex groups (Table 1) with each containing seven to ten primer pairs according to colour and size avoiding overlapping of the same colour. PCR amplifications were carried out on a Techne 384-well thermal-cycler in a final reaction volume of 5 to10 µl, containing 50 ng genomic DNA, 1X multiplex PCR master mix (Qiagen, Valencia, USA) containing HotStarTaq DNA polymerase, multiplex PCR buffer, 2 mM $MgCl_2$, 250 mM of each dNTP (dATP, dCTP, dGTP, dTTP), and 0.05 µM of each multiplex fluorescent primer mix. The PCR products were diluted 1:20 and 1 µl was added to HiDi-formamide (Life Technologies) containing 0.12 µl GeneScan-LIZ500 as size standard and electrophoresed on an ABI Prism3130xl sequencer. Fragments were analyzed with Gene mapper 4.0 (Applied Biosystems) software and allele sizes was verified. A list of microsatellite loci and their chromosomal locations is given in Table 1.

Data analysis

Genetic distances of breeding lines and hybrids tested were calculated with the Powermarkerver 3.25 (Liu and Muse, 2005) using the Rogers (1972) parameter. The polymorphic information content (PIC) for each marker was determined using Powermarkerver 3.25 software. Average linkage [unweighted paired group method using arithmetic averages (UPGMA,)] clustering was calculated based on Rogers distance (RD) estimates between pairs of inbred lines for the yellow and white lines and hybrids separately. To evaluate the robustness of the UPGMA dendrogram, the cophenetic correlation was calculated (Sneath and Sokal, 1973) utilizing NTSYS ver 2.21 software (Rohlf, 2009).The data was transformed with the dcenter module and Eigenvectors calculated. Principal coordinates analysis (PCoA) was carried out on the calculated distances usingNTSYS ver. 2.21 software (Rohlf, 2009).

RESULTS

A total of 96 SSR primer sets were used to fingerprint the breeding lines and hybrid samples. The SSR markers with more than 20% missing values were discarded and data from 80 markers were used in the final analysis. The PIC values of markers varied between 0.2 and 0.9. The average number of alleles per locus was 12.7 for a total of 1043 genotypes tested, with a total of 1874 alleles amplified (Table 1). The number of alleles amplified varied between 6 and 36 per locus for the total population screened. When looking at the separate populations tested, the mean number of alleles per locus was the highest for the yellow maize breeding lines, which also had the highest entries in the screening program (Table 2).

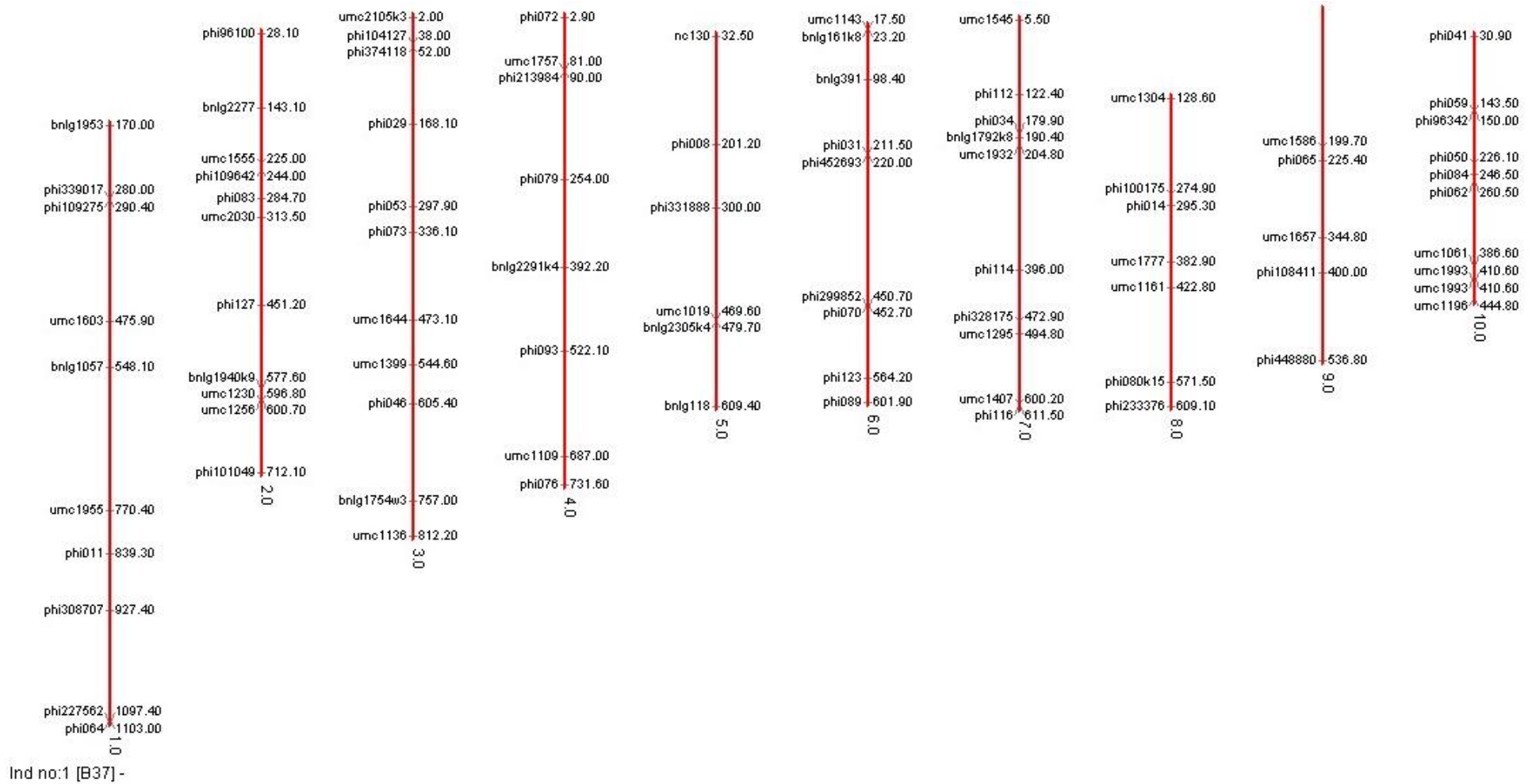In PCoA of the RD of the breeding lines, the white and

Figure 1. Maize genomic map indicating SSR markers used in the study (based on IBM2 2008 distances).

yellow groups were clearly separated in two clusters, with relatively few overlapping genotypes (Figure 2). When regarding the two localities where the breeding lines originated, it was clear that the white lines could be separated into three clearly distinguished groups, with the lines from Potchefstroom grouping in two separate clusters. The yellow breeding lines were much more heterogeneous between the two localities, with

Table 1. Multiplex grouping of SSR markers.

Panel	Marker	Dye	Bin	Repeat
M1	umc2162	PET	6.06	(CCG)5
	umc1993	6-FAM	10.06	(CT)17
	phi233376	6-FAM	8.03	CCG
	phi046	VIC	3.08	ACGC
	phi331888	VIC	5.04	AAG
	phi062	VIC	10.04	ACG
	phi089	NED	6.08	ATGC
	phi339017	NED	1.03	ACC
M2	umc1256	VIC	2.09	(CAT)5
	phi101049	VIC	2.10	AGAT
	phi108411	NED	9.05	AGCT
	phi96342	NED	10.02	ATCC
	umc2217	6-FAM	1.04	(TG)6
	umc1279	PET	9.00	(CCT)6
	phi104127	PET	3.00	ACCG
	phi96100	PET	2.00	ACCT
M3	bnlg118	6-FAM	5.07	(AG)n
	phi076	6-FAM	4.11	AGCGGG
	phi032	6-FAM	9.04	AAAG
	phi114	VIC	7.03	GCCT
	umc1196	NED	10.07	CACACG
	phi050	PET	10.03	AAGC
	phi123	PET	6.07	AAAG
M4	umc1304	6-FAM	8.02	(TCGA)4
	phi053	6-FAM	3.05	ATAC
	umc1295	VIC	7.00	(AT)6
	umc1657	VIC	9.00	(GACGG(4
	phi396160	VIC	5.02	AGGCG
	umc1757	NED	4.01	(TCC)7
	umc1109	PET	4.10	(ACG)4
	phi059	PET	10.02	ACC

Table 1. Contd.

M5	phi308707	6-FAM	1.10	AGC	
	umc1644	6-FAM	3.06	(TTG)8	
	phi072	VIC	4.00	AAAC	
	phi011	VIC	1.09	AGC	
	phi008	NED	5.03	GGC	
	phi079	NED	4.05	AGATG	
	phi093	NED	4.08	AGCT	
	bnlg391	PET	6.01	-	
	phi112	PET	7.01	AG	
M6	phi109275	6-FAM	1.03	AGCT	
	phi041	6-FAM	10.00	(CAT)5	
	umc1399	VIC	3.07	9CTAG)5	
	umc1122	VIC	1.06	(CGT)7	
	phi034	NED	7.02	CCT	
	phi031	NED	6.04	GTAC	
	umc1126	PET	2.00	ACC	
	phi127	PET	2.08	AGAC	
M7	umc1019	6-FAM	5.06	(CT)17	
	umc1161	6-FAM	8.06	(GCTGGG)5	
	bnlg1057	6-FAM	1.00	(AG)17	
	phi083	VIC	2.04	AGCT	
	phi070	NED	6.00	AGCTG	
	umc1545	PET	7.00	(AAGA)4	
	phi029	PET	3.04	AG/AGCG	
	phi213984	PET	4.01	ACC	
M8	umc1061	6-FAM	10.06	(TCG)6	
	bnlg1792k8	6FAM	7.00	(AG)16	
	phi452693	VIC	6.04	AGCC	
	phi227562	VIC	1.11	ACC	
	phi299852	NED	6.07	AGC	
	phi073	NED	3.05	AGC	
	phi328175	PET	7.04	AGG	
	phi080k15	PET	8.00	AGGAG	

Table 1. Contd.

M9	umc1407	6-FAM	7.00	(GGC)6			
	nc130	6-FAM	5.00	AGC			
	phi374118	6-FAM	3.02	ACC			
	phi064	VIC	1.11	ATCC			
	phi116	VIC	7.06	ACTG/ACG			
	umc1143	NED	6.00	AAAAT			
	umc1136	NED	3.10	(GCA)5			
	umc1777	PET	8.00	AGCC			
	phi084	PET	10.04	GAA			
M10	phi100175	6-FAM	8.03	AAGC			
	phi109642	VIC	2.00	ACGG			
	phi448880	VIC	9.06	AAG			
	bnlg2291k4	VIC	4.06	(AG)17			
	umc1555	NED	2.03	(TTCA)7			
	phi014	NED		GGC			
	bnlg2305k4	NED	5.06	(GA)6			
	phi065	PET	9.03	CACTT			
M11	bnlg161k8	VIC	6.00	(AG)n			
	bnlg1754w3	PET	3.09	(AG)20			
	umc2105k3	PET	3.00	(CCAT)4			
	bnlg1940k9	NED	2.08	(AG)18			
	bnlg2277	6-FAM	2.02	(AG)19			
	umc1932	NED	7.00	(GA)8			
M12	bnlg1953	NED	1.02	(AG)17			
	umc1586	NED	9.03	(ATC)5			
	umc1955	PET	1.08	-			
	umc1603	PET	1.00	(GCC)4			
	umc1230	PET	2.00	(TAA)8			

only a few lines falling in definite clusters. The lines originating from CIMMYT, Mexico, clustered together with the Cedara material for both white and yellow lines. These lines were used in the Cedara breeding program over a long period of time and therefore the result was as expected.

The white lines were compared to some anchor lines with known heterotic grouping (Figures 3 and 4). The lines with known pedigrees related to K64, clustered together (r=0.91), with only one line with larger RD from the rest of the group. A large number of white lines clustered with the Lancaster line Mo17. The rest of the groups could be separated into distinguishable clusters by the UPGMA derived dendrogram (Figure 4). PCoA and cluster

Table 2. Average and range of number of alleles per locus for 1043 maize genotypes.

Population	Number of genotype	Number of allele	Mean alleles / locus	Range allele
White breeding lines	207	575	7.2	2-20
Yellow breeding lines	643	770	9.5	2-24
Hybrids	193	529	6.61	2-19
Total	1043	1874	12.7	6-36
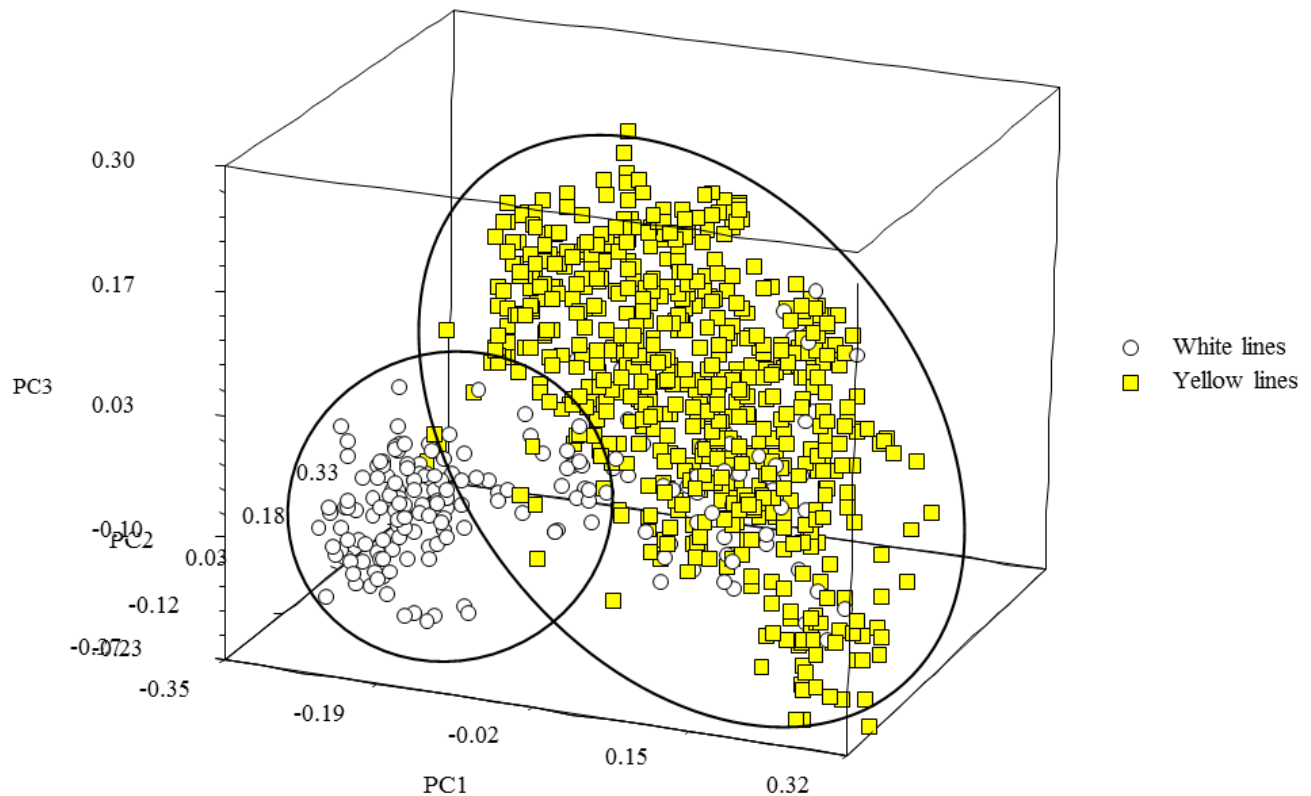
Figure 2. Principal coordinates analysis of white and yellow breeding lines from the two localities.

analysis of yellow breeding lines revealed a small group of lines clustering with Mo17 (Figures 5 and 6) (r=0.72), with the rest of the lines with mixed origin. The cluster analysis (Figure 6) identified several related groups of lines indicating close genetic relationships. These results show the extent to which these lines were interbred in the breeding program over several decades.

One hundred and ninety three (193) hybrids were included in the study to explore the diversity achieved and the genotypes were compared to locally cultivated maize cultivars produced by private companies. Five distinct groups of ARC hybrids could be identified. No overlap could be seen between ARC hybrids and those from the private companies tested (Figure 7). Cluster analysis (not shown) revealed several identical hybrids included in the entries (r = 0.88).

DISCUSSION

Maize breeding is a never-ending challenge for developing new inbred lines and cultivars with improved yield, as well as tolerance to a diverse spectrum of biotic and abiotic stress conditions, such as diseases, pests, low soil fertility and drought. New cultivars must also conform to changing industry requirements such as grain quality for milling and ethanol production. Climatic changes such as global warming pose new challenges to breeders for developing cultivars that will perform well under changed conditions. Although seed companies produce their own hybrids, they remain dependent on foundation-seed suppliers for inbred lines and even sometimes locally adapted hybrids.
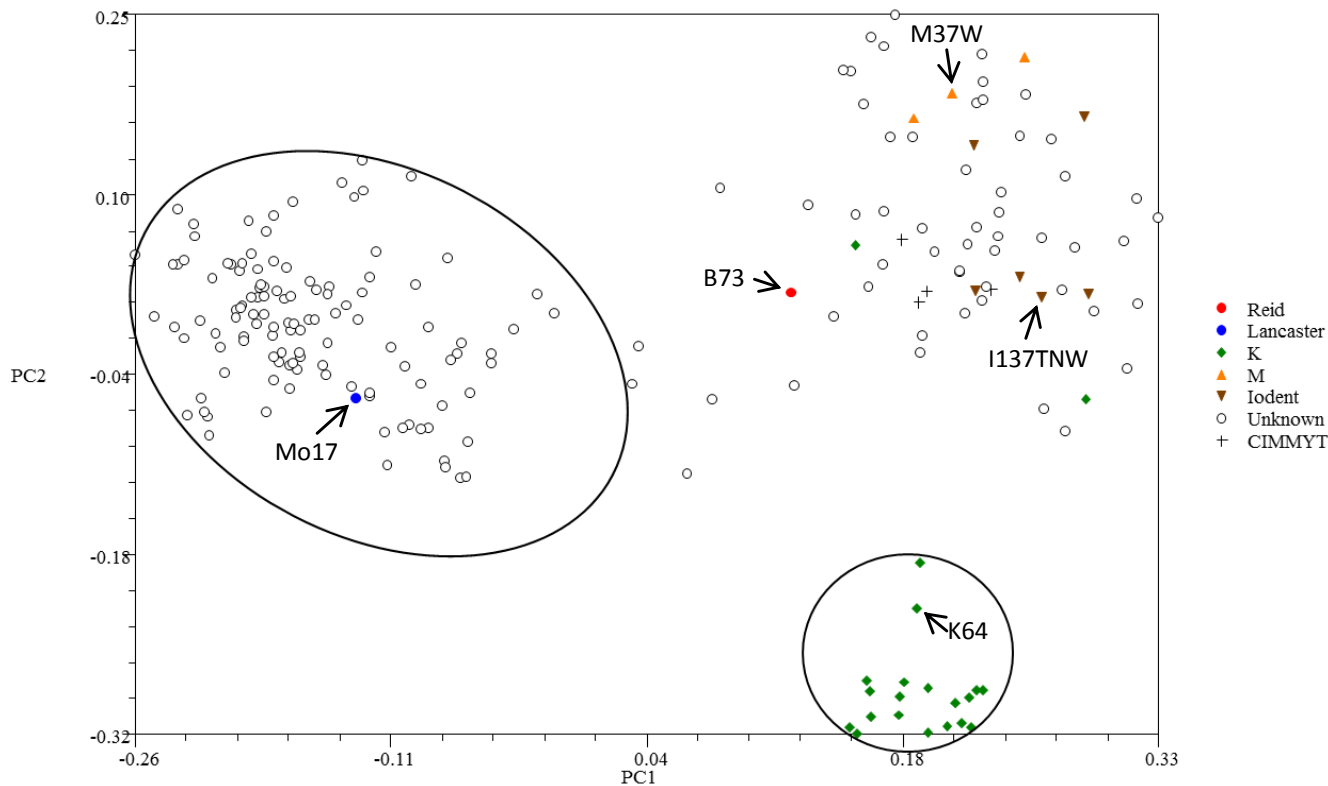
The ARC-GCI has a large germplasm collection of

Figure 3. Principal coordinates analysis of white breeding lines compared with anchor lines of known heterotic groups.

inbred lines, containing both white and yellow lines, at its disposal. These lines needed to be characterized on a molecular level to determine close relationships or genetic distances to narrow down the possibilities for utilizing this material in testcrosses. It also served the purpose of identifying duplicate lines, as well as genetic drift within well-known breeding lines.

During the past decade, the technology used in finger-printing of field crops developed from restriction fragment length polymorphism (RFLP) to amplified fragment length polymorphism (AFLP), SSR and SNP. The RFLP technique was laborious and time consuming and only single data points could be analyzed per reaction. The AFLP technique was less time consuming and labour intensive and rendered up to 100 data points per reaction, but it was difficult to duplicate results over time and between different laboratories. The SSR technique is a simple polymerase chain reaction (PCR) technique, which lends itself to high throughput of samples when multiplex reactions are used with an automatic genetic analyzer like the ABI3130xl. The technique is also easy to duplicate among different laboratories.

Although the SNP technique can analyze thousands of data points per run, one has to look at the application that you need the analysis for. The SSR technique can be utilized to give low coverage of the whole genome of maize, giving an overall indication of the genetic material present

in a specific population. Knowledge of the relationships among lines can help identify a set of specific inbreds that have maximal diversity for the analysis of the effects of background (Liu et al., 2003), to be used further in experimentation and association analysis of specific traits.

A similar African study was conducted by Legesse et al. (2007), to determine the diversity present in an African breeding population in Ethiopia and Zimbabwe. They found that they could distinguish between their inbred lines with as little as 27 SSR loci, which grouped their lines in groups corresponding to adaptation to different altitudes and consistent with known pedigree information. In our local line collection, the breeding program has used lines from different countries and sources and the end result is a totally different population from the original due to interbreeding between different heterotic groups. New grouping of the breeding lines became of vital importance for improvement of the crop.

SSR markers were utilized in this study in the genotyping of the germplasm collection of the ARC-GCI from two different localities and were used at a density of eight markers per chromosome. The 80 markers utilized amplified a total of 1874 alleles giving an average of 23.4 alleles per locus, which is higher than observed by Van Inghelandt et al. (2010). Results were used to determine the genetic relationships among the inbred lines and
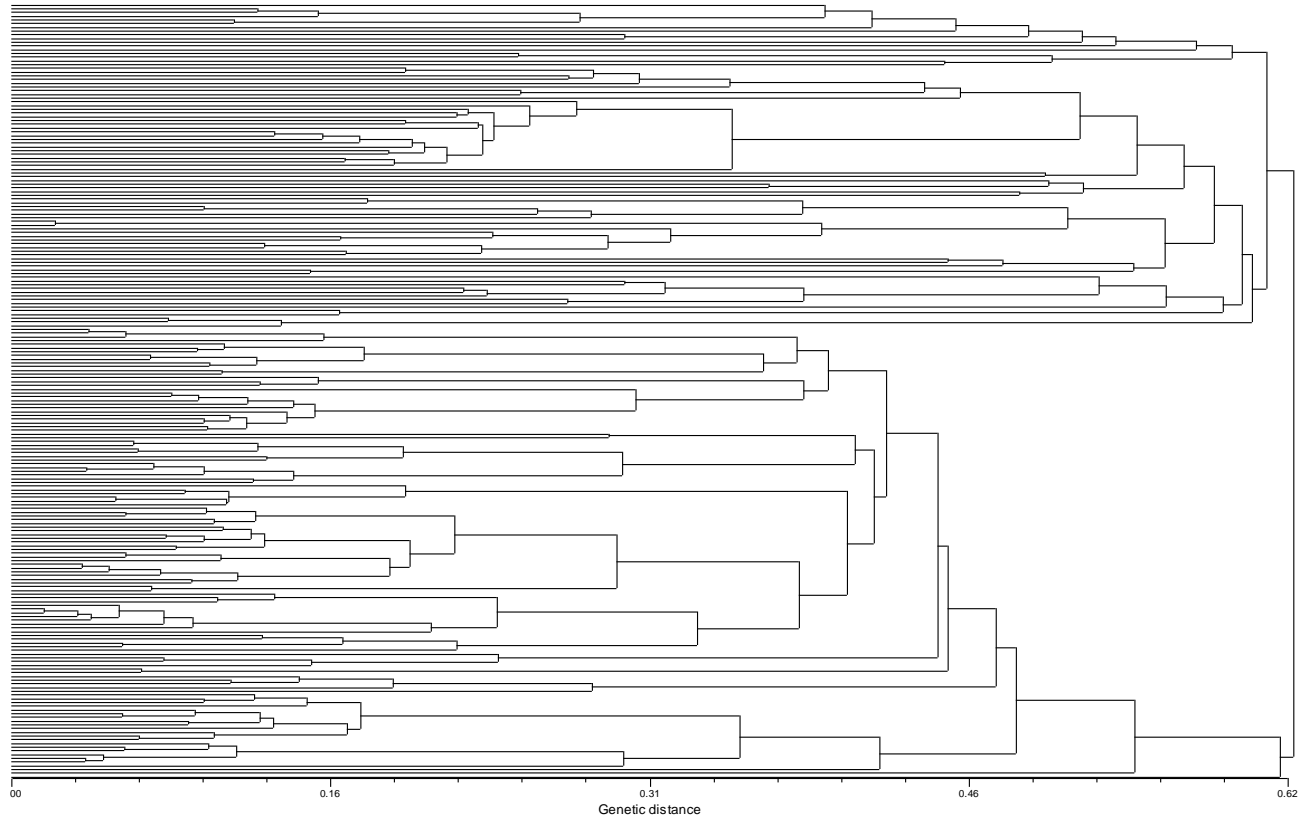
Figure 4. Associations among white inbred lines revealed by average linkage (UPGMA) cluster analysis based on Rogers distance.
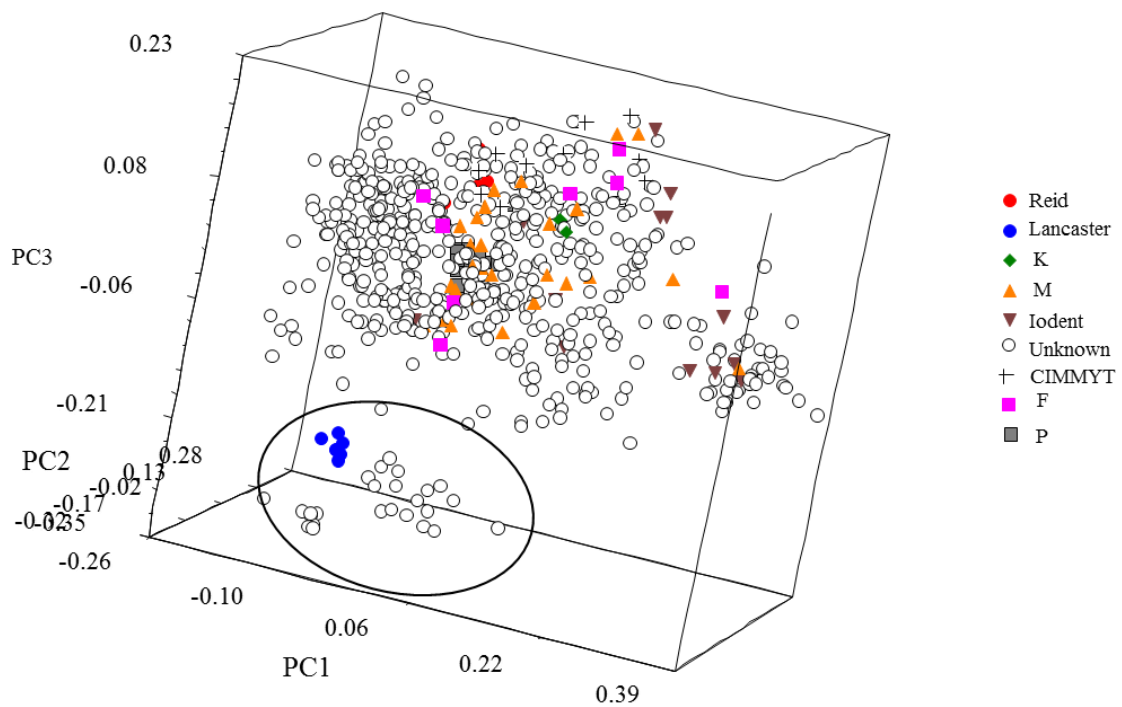

Figure 5. Principal coordinates analysis of yellow breeding lines compared with anchor lines of known heterotic groups.
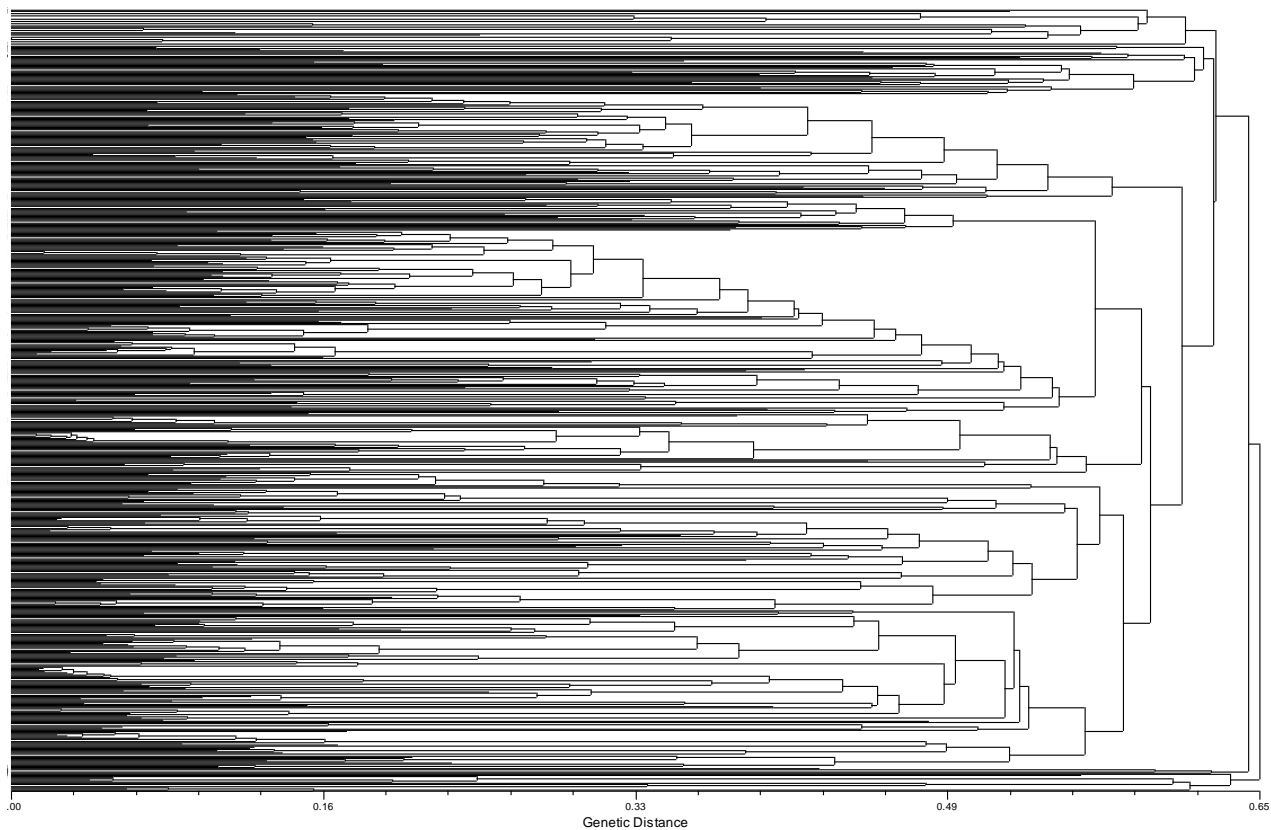
Figure 6. Associations among yellow inbred lines revealed by average linkage (UPGMA) cluster analysis based on Rogersdistance.
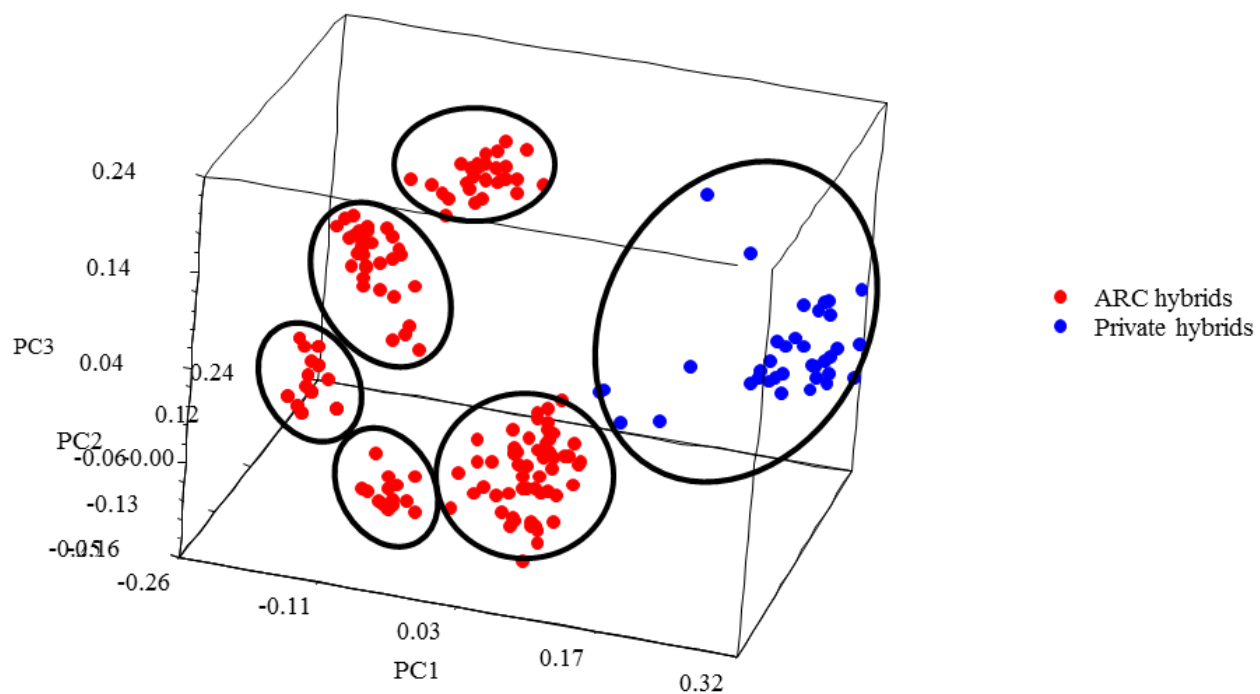

Figure 7. Principal coordinates analysis of hybrids tested.

hybrids. Principal coordinates analysis clearly separated white and yellow lines as expected from the pedigree data that were available from the breeding programs. These groups were further analysed separately. In the analysis of white breeding lines based on Rogers estimates, the lines clearly were separated in two different groups with PCoA analysis. A large group of lines were grouped together with Mo17, a Lancaster line, in accordance with pedigree data.

All the white lines could not be grouped into specific heterotic groups and some were of mixed origin. These lines were developed through crosses between the different groups in the breeding program and it is thus not possible to group them with currently known heterotic groups. The cluster analysis gave a more detailed grouping of lines (Figure 4): the K-group was derived from K64 crosses. K64 is a Kansas inbred known for fairly good drought tolerance.

The CB group derived from inbreeding US Corn belt hybrids clustered with Mo17, a Lancaster line. A 3rd group was identified as a mixed group containing combinations of Corn Belt material with Sahara or I-group lines as well as lines with maize streak virus resistance. The M37W group was derived from the Australian inbred 21A. Lines originating from CIMMYT also clustered in this group as a sub-group.

The I137TN group is the major South African group known for its superior combining ability for grain yield, especially in combination with US Corn Belt material. The group originated from the locally developed I137TN inbred. I137TN was selfed out of a yellow endosperm local variety cross between the two varieties Teko Yellow x.

ACKNOWLEDGEMENT

REFERENCES

Brummer EC (1999).Capturing heterosis in forage crop cultivardevelopment. Crop Sci. 39:943-954.

Charcosset A,Essioux L (1994).The effect of population structureon the relationship between heterosis and heterozygosity atmarker loci.Theor. Appl. Genet. 89:336-343.

Hamblin MT, Warburton ML, Buckler ES (2007).Empirical Comparison of Simple Sequence Repeats and Single Nucleotide Polymorphisms inAssessment of Maize Diversity and Relatedness.PLoS ONE 2(12): e1367. doi:10.1371/journal.pone.0001367.

Legesse BW, Myburgh AA, Pixley KV, Botha AM. (2007) Genetic diversity of African maize inbred lines revealed by SSR markers. Hereditas 144:10-17.

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

Melchinger AE, Messmer MM, Lee M, Woodmana WL, Lamkey KR(1991) Diversity andrelationships among U.S. maize inbredsrevealed by restriction fragment lengthpolymorphisms. Crop Science. 31:669–678.

Rogers JS (1972).Measures of genetic similarity and genetic distance.Studies in Genetics VII. University of Texas Publishing, 7213:145-153.

Rohlf F (2009). NTSYS pc.Numerical Taxonomy system. Exeter Publishing, Setauket.

Saghai-Maroof MA, SolimanKM, 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.

Smith JSC, Chin ECL, Shu H, Smith OS, WalLSJ, Senior ML, Mitchell SE, Kresovich S, Ziegle J (1997). An evaluation ofthe utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisonswith data from RFLPs and pedigree. Theor. Appl. Genet. 95:163-173.

Sneath PHA,Sokal RR(1973). Numerical taxonomy. Freeman, San Francisco.

Van Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010). Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor. Appl. Genet. 120:1289-1299.

Xia XC, Reif JC, Hoisington DA, Melchinger AE, Frisch M, Warburton ML (2004). Genetic diversity among CIMMYT maize inbred lines investigated with SSR markers: I. Lowland tropical maize. Crop Science. 44:2230-2237.