Plant Breeding, **136**, 41–49 (2017)

doi:10.1111/pbr.12439

Genetic divergence in two tropical maize composites after four cycles of reciprocal recurrent selection

Adesike O. Kolawole[1,2,3], Abebe Menkir[1], Melaku Gedil[1], Essie Blay[2], Kwadwo Ofori[2] and Jennifer G. Kling[4]

[1]International Institute of Tropical Agriculture (IITA), Oyo Road, PMB 5320, Ibadan, Nigeria; [2]West Africa Centre for Crop Improvement (WACCI), University of Ghana, PMB 30, Legon, Ghana; [3]Ladoke Akintola University of Technology, PMB 4000, Ogbomoso, Nigeria; [4]Oregon State University, Crop Science Building 107, Corvallis, OR 97331-3002, USA; Corresponding author, E-mail: a.menkir@cgiar.org

With 3 figures and 7 tables

Received August 24, 2016 / Accepted November 5, 2016
Communicated by T. Lübberstedt

Abstract

Two tropical maize composites were subjected to four cycles of reciprocal recurrent selection to develop divergent inbred lines with good combining ability. This study was conducted to examine the extent of genetic diversity, changes in allele composition and genetic structure, of 100 randomly selected S_1 lines each from the original (C_0) and advanced (C_4) selection cycles of TZL COMP3 and TZL COMP4, genotyped using single nucleotide polymorphism (SNP) markers. Results revealed that the proportion of alleles at both low and high frequencies decreased from C_0 to C_4, whereas those at intermediate frequencies increased at C_4 in the two composites. More unique and other alleles were lost at C_4 in TZL COMP3 relative to those in TZL COMP4. The changes in different measures of genetic diversity were either small or negligible with selection in the two composites. The proportion of markers departing from Hardy–Weinberg equilibrium (HWE) decreased with selection, whereas the total number of pairs of markers in linkage disequilibrium increased with selection in the two composites. Examination of changes in population structures using a model-based approach as well as cluster and multivariate analyses found a high degree of concordance in stratifying the 400 S_1 lines into four non-overlapping groups corresponding to the two selection cycles each within the reciprocal composites. The observed molecular-based divergence between cycles within the same composite and the clear differentiation between the complementary composites highlight the importance of reciprocal recurrent selection for preserving genetic diversity for long-term selection. This increases the potential of the advanced selection cycles to sustain genetic gain in productivity of hybrids adapted to the savannas in West and Central Africa.

Key words: genetic structure — genetic diversity — hybrid-oriented populations — single nucleotide polymorphism

Maize (*Zea mays* L.) has emerged as an important staple food crop and source of income for small holder farmers in sub-Saharan Africa (Smale et al. 2013). Both the area planted to maize and grain production have increased significantly in West and Central Africa (WCA) as a result of expanded use of the crop for food, animal feed, and industrial products. The introduction of improved maize cultivars adapted to the diverse agro-ecological zones has contributed to the phenomenal increase in maize production in the various countries in this region (Alene et al. 2009). Most of the area in WCA is planted to improve open-pollinated maize varieties (Rusike and Eicher 1997) mainly because private seed companies are not well developed in many countries. Auta et al. (2001) suggested that the development and accelerated deployment of maize hybrids can allow greater increases in maize yields in the major maize-producing countries in WCA. Studies have demonstrated that hybrids can increase farmers' maize yields by more than 40% in favourable growing environments and by more than 30% even under stressful conditions (Byerlee and Jewell 1997).

The International Institute of Tropical Agriculture (IITA) started a hybrid breeding programme in 1979 to strengthen involvement of the private sector in the production and marketing of hybrid maize in WCA (Kim 1997). This programme focused on generating hybrids with high yield potential and resistance to specific biotic and abiotic stresses for achieving greater and dependable yields in the major production zones in this region (Efron et al. 1989, Kim 1997). Considering the importance of having hybrid-oriented populations and application of selections schemes that maximize expression of heterosis in hybrids, IITA utilized results of diallel studies and promising heterotic patterns of tropical germplasm described by Wellhausen (1978) and Goodman (1985) as the basis to create two late-maturing composites known as TZL COMP3 and TZL COMP4 for a long-term reciprocal recurrent selection (RRS) programme following the comprehensive breeding approach proposed by Eberhart et al. (1967). The two composites have been subjected to four cycles of RRS without infusion of new germplasm to boost agronomic performance of hybrids formed from inbred lines derived from advanced selection cycles (Hallauer and Eberhart 1970, Falconer and Mackay 1996). Although the RRS programme has been underway for more than two decades, the changes in genetic diversity and divergence within and between the two composites have not been documented.

Comstock et al. (1949) originally proposed RRS as a cyclical breeding procedure for increasing the frequency of favourable alleles in two distinct populations to enhance the general and specific combining ability of lines derived from them. RRS minimizes inbreeding, promotes genetic recombination within populations, maintains adequate genetic variation and allows gradual improvement in future selection cycles (Hallauer 1985). As RRS changes allele frequencies, levels and distribution of the genetic variability and the genetic structure of the populations (Pinto et al. 2003a), information on the shifts in allele frequencies in the genome resulting from selection is important to assess future genetic progress without excessive loss of genetic diversity.

Molecular markers have been used to examine the extent of loss and fixation of favourable alleles in populations after improvement (Pinto et al. 2003a). Different types of molecular

markers have been used to examine the levels of genetic diversity and genetic changes in maize populations subjected to recurrent selection (Stuber et al. 1980, Heredia-Diaz et al. 1996, Labate et al. 1999, Pinto et al. 2003b, Hinze et al. 2005, Solomon et al. 2010). Labate et al. (1999) used 82 restriction fragment length polymorphism (RFLP) loci to determine the temporal changes in allele frequencies in two reciprocal maize populations after 12 cycles of selection and found that nearly 30% of the alleles were extinct and 10% of them were near fixation in each population. Also, mean expected heterozygosity decreased, while genetic variation increased in these populations. Pinto et al. (2003b) used 30 simple sequence repeat (SSR) markers to investigate the effects of modified RRS on genetic structures of improved cycles of two tropical maize populations and found that most of the alleles that were in low frequency in the original populations were lost after one cycle of RRS, while at the same time the number of alleles belonging to the extreme classes of frequencies increased. Ordas et al. (2015) also used SSR markers to determine changes in two maize (*Zea mays* L.) synthetics subjected to RRS using alternative methods and found that the genetic distance between reciprocal populations increased possibly due to a more efficient increase in favourable alleles with additive effects. More recently, single nucleotide polymorphism (SNP) markers have been extensively used for genetic diversity analysis, linkage map construction, marker-assisted selection (MAS) and marker–trait association (Pinto et al. 2003a, Hinze et al. 2005). SNP markers are biallelic in nature and occur at a much higher frequency in the genome than other markers (Zhu et al. 2003), and their genotyping can be easily automated (Jones et al. 2007). Very few studies have been conducted to assess changes in allele frequency and the genetic structure of populations improved through RRS using SNP markers. The present study was therefore conducted to examine the extent of genetic variation present within and between selection cycles of the two tropical broad-based composites (TZL COMP3 and TZL COMP4) and to determine the effect of RRS on changes in allele frequency and population structure of randomly selected S_1 lines from these composites using SNP markers.

Materials and Methods

Genetic materials: Diallel crosses of late-maturing maize populations were evaluated in a trial in multiple locations in 1988 (MIP 1996). The performance of population crosses observed in this trial followed the heterotic response of crosses between the well-known Tuxpeno dent and Caribbean flint races of maize (Wellhausen 1978, Goodman 1985). The populations belonging to the Caribbean heterotic group, namely TZB-SR and Suwan 1-SR, were then crossed to form a broad-based composite known as TZL COMP3 C_0. Populations representing the Tuxpeno heterotic group, namely TZPB-SR, POP 43-DMRSR and POP 21-SR, were intercrossed to form the second broad-based composite referred to as TZL COMP4 C_0 (MIP 1996). Details about the component populations and the steps followed to constitute the two late-maturing composites have been described in the Maize Improvement Program report (MIP 1996). A detailed outline of the RRS scheme for the two composites can be found in Menkir et al. (2015).

DNA extraction: More than 300 S_1 lines were extracted each from C_0 and C_4 of TZL COMP3 and TZL COMP4. Of these, 100 S_1 lines were selected at random from each selection cycle of the composites (TZL COMP3 C_0, TZL COMP4 C_0, TZL COMP3 C_4 and TZL COMP4 C_4) and planted in single rows at Ibadan (3°58′E, 7°22′N, altitude of 150 m.a.s.l.) in Nigeria in 2014. Young leaves were collected from at least 20 maize seedlings of each S_1 lines grown in the field for 2 weeks to form bulk samples that were kept in a perforated bag and stored in −80°C freezer. Leaf tissue samples lyophilized in console dry system from Labconco (Labconco Inc., Kansas City, Missouri, USA) were ground, and DNA was extracted using a modified CTAB protocol of Saghai-Maroof et al. (1984). The quality of the DNA for genotyping by sequencing (GBS) was ascertained by digesting the DNA with restriction enzyme *Hin*dIII. The genomic DNA was then transferred into a 96-well plate, properly sealed with rubber plate covers and sent to Institute for Genomic Diversity at Cornell University, Ithaca, NY, for genotyping.

SNP genotyping: All molecular information was obtained from the Genomic Diversity Facility in Cornell University. GBS libraries were prepared and analysed as described by Elshire et al. (2011), using the enzyme *Ape*KI for digestion and creating a library with unique barcodes for each genotype. Raw reads from the sequenced GBS library were called in the GBS analysis pipeline TASSEL version 3.0.166 (Glaubitz et al. 2014). The filtered sequences were aligned to the maize reference genome B73 REFGEN v1 (Schnable et al. 2009) using the Burrows-Wheeler alignment tool (BWA). This procedure produced 121 910 HapMap SNPs (filtered) covering the 10 chromosomes in the maize genome. To reduce the potential errors due to missing data, filtering alignment was set at minimum allele frequency of 0.05 and loci with more than 90% missing data were removed prior to further analysis, leaving 10 579 high-quality polymorphic markers. The SNP filtering was completed using TASSEL 5.2.1 software (Bradbury et al. 2007).

Diversity analysis: The genetic diversity parameters including the number of alleles, observed heterozygosity (Ho), Shannon–Weaver diversity index, gene diversity as described by Weir (1996), inbreeding coefficient (fixation index) and percent polymorphic loci (% P) were computed at each locus for each selection cycle within the two composites using POWERMARKER V3.25 (Liu and Muse 2005). Subsequently, the mean and its corresponding standard error were computed across marker loci for each parameter and each selection cycle using SAS (SAS Institute 2010), Paired *t*-test was also computed to determine the level of significance in the difference between the two selection cycles for each diversity parameter using SAS (SAS Institute 2010). Deviations from Hardy–Weinberg equilibrium (HWE) at individual loci were tested with an exact test described by Guo and Thompson (1992) using GENALEX software (Peakall and Smouse 2012). Linkage disequilibrium (LD) between all pairs of loci was estimated on each chromosome for each selection cycle within the two composites with 10 579 SNPs in TASSEL 5.2.1. Standardized disequilibrium coefficients (D′) and squared allele frequency correlations (r^2) were calculated in TASSEL 5.2.1 with a sliding window (Bradbury et al. 2007) to assess the significance ($P < 0.01$) of D′ and r^2 on each chromosome.

An ADMIXTURE model-based clustering method (Alexander et al. 2009) was used to assess population structure of the 400 S_1 lines. Pairwise genetic distance matrix was calculated from 10 579 SNP markers in PLINK (Purcell et al. 2007). A Ward's minimum variance hierarchical cluster dendrogram was then built from the identity-by-state matrix using the analyses of phylogenetics and evolution package (Paradis et al. 2004) implemented in R (R Development Core Team 2010). To validate the clustering pattern obtained from ADMIXTURE and the hierarchical clustering algorithms, we carried out discriminant analysis of principal components (DAPC) as described by Jombart et al. (2010) using the R package 'ADEGENET' 2.11 version of the R program (Jombart 2008). Bayesian information criteria (BIC) were used to define the number of groups where the *k* with the lowest BIC value represented the most probable number of groups for the set of data analysed. According to Jombart et al. (2010), the best BIC is indicated by an elbow in the curve of BIC values matching with the smallest BIC and the 'true' number of cluster. *F* statistics (F_{ST}) were calculated for pairs of cycles within and between the two reciprocal composites using ADMIXTURE to estimate the observed genetic divergence. To compare the relative importance of the molecular variation between selection cycles within the two tropical maize composites, analysis of molecular variance (AMOVA) was computed as proposed by Excoffier et al. (1992). The GENALEX V6.5 software was also used to estimate the pairwise PhiPT values, an analogue of F_{ST}

(Wright 1978), to determine the extent of differentiation between selection cycles of the two composites.

Results

Genetic diversity

A total of 10 579 informative SNPs were used for analysis of the extent of genetic variability and population structure in 100 S_1 lines each derived from initial (C_0) and advanced (C_4) selection cycles of the two tropical composites. Of these markers, 7368 showed allele frequency of less than 0.1, whereas 2810 had allele frequency of greater than 0.2. Also, there were 401 SNPs with alternative alleles of equal frequencies of 0.5. The 10 579 SNP markers produced 10 allelic pairs: four were homozygous and six were heterozygous. The proportion of A : G, A : C, A : T, C : G, T : G and C : T alleles at C_4 decreased in TZL COMP3, but changed very little in TZL COMP4 (Table 1). In contrast, the proportion of G : G, C : C, T : T and A : A alleles at C_4 increased in the two composites with the increase in TZL COMP4 being greater (Table 1).

The distributions of allelic frequencies both before and after selection showed a U-shape with higher frequencies having the extreme values (Fig. 1a,b). The proportion of alleles at both low and high frequencies decreased from C_0 to C_4, whereas those at intermediate frequencies increased at C_4 in the two composites (Table 2). The changes in allele frequencies at C_4 were greater in TZL COMP3. The proportion of fixed alleles increased at C_4 in the two composites with about 4% of the alleles got fixed in these composites at C_4. The percentage of unique allele detected in C_0 was about 11% more in TZL COMP3 relative to TZL COMP4 (Table 2). By C_4, almost all the unique alleles were lost in TZL COMP3, whereas nearly 2% were retained in TZL COMP4. Considering all the alleles detected in C_0, 17% were lost in TZL COMP3 and 13% were lost in TZL COMP4 at C_4 (Table 2).

The average number of alleles per locus decreased significantly at C_4 in the two composites, although the decrease was 11% in TZL COMP3 and only 2% in TZL COMP4 (Table 3). There was a significant increase in the average minor allele frequency at C_4 in TZL COMP 3, but not in TZL COMP4. Even though the average observed heterozygosity decreased significantly at C_4 only in TZL COMP3, the decrease was very small. The average gene diversity at C_4 significantly increased in TZL COMP3, whereas it decreased significantly in TZL COMP4. However, the changes in gene diversity were small in the two composites (Table 3). Shannon–Weaver diversity index decreased by 15% in TZL COMP3 and by 6% in TZL COMP4 at C_4. Also, percent polymorphic loci decreased by 22% in TZL COMP3 and by 4% in TZL COMP4 after four cycles of RRS (Table 3).

The fixation index, which is an indicator of excess of homozygosity, varied from −1.0 to 1.0 in both C_0 and C_4 of the two composites (data not shown). The total number of SNPs showing positive fixation index decreased by 26% in TZL COMP3 and by 19% in TZL COMP4 at C_4, whereas those showing negative fixation index decreased by 21% in TZL COMP3, but changed very little in TZL COMP4 at C_4 (Table 4). The sign of the fixation index over all loci in the two composite was mainly negative because the SNP markers tend to move towards heterozygosity. Less than 10% of the 10 579 SNP markers showed a significant ($P < 0.01$) deviation from Hardy–Weinberg equilibrium (HWE) in each selection cycle within the two composites (Table 4). The number of markers deviating from HWE decreased by 11% in TZL COMP3 and by 8% in

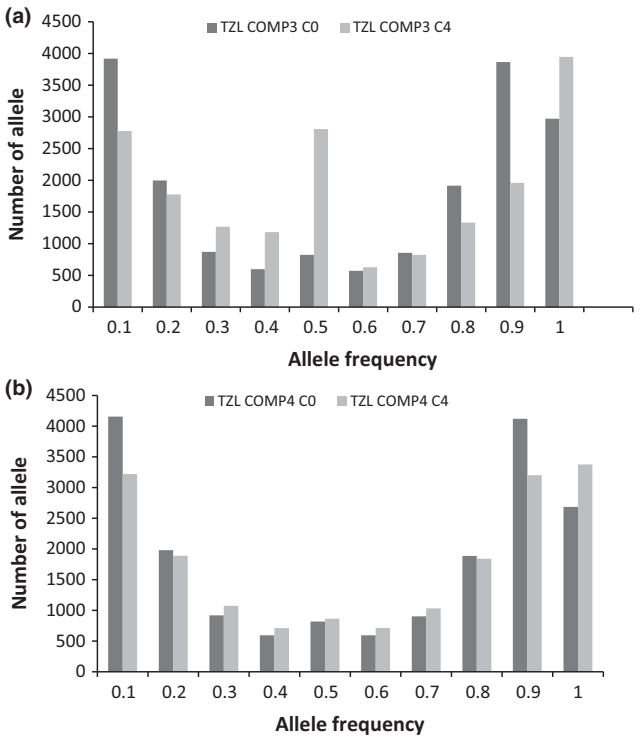

Fig. 1: (a, b) Distribution of the number of alleles according to allelic frequencies at C0 and C4 selection cycles for both TZL COMP3 (a) and TZL COMP4 (b)

Table 1: Percentage of allelic composition in two tropical maize composites subjected to four cycles of reciprocal recurrent selection

Alleles	TZL COMP3 C_0	TZL COMP3 C_4	TZL COMP4 C_0	TZL COMP4 C_4
A:G	26.98	21.58	27.41	26.24
A:C	10.65	8.68	10.76	10.47
A:T	4.24	3.34	4.28	4.12
C:G	16.19	12.86	16.52	15.63
T:G	10.69	8.76	10.89	10.37
C:T	27.33	21.95	27.82	26.69
G:G	1.65	8.82	0.85	2.62
C:C	1.54	8.52	1.00	2.47
T:T	0.40	2.77	0.23	0.69
A:A	0.33	2.73	0.24	0.71

Table 2: Percentage distributions of allele frequencies in C_0 and C_4 of TZL COMP3 and TZL COMP4

	TZL COMP3		TZL COMP4	
	C_0	C_4	C_0	C_4
	%			
Allele at low frequency (<0.25)	39.73	34.45	39.99	37.40
Allele at intermediate frequency (0.25–0.50)	11.08	24.71	11.16	12.92
Allele at high frequency (>0.50)	34.82	22.29	35.97	33.17
Allele at fixation (1.0)	14.37	18.55	12.88	16.51
Unique alleles	15.60	0.88	4.52	1.64
Lost alleles	–	17.07	–	13.12

Table 3: Average diversity measures and their corresponding standard errors at C_0 and C_4 of TZL COMP3 and TZL COMP4 after four cycles of RRS computed from 10 579 SNP markers

Selection cycles	Average number of allele	Average minor allele frequency	Average observed heterozygosity	Average gene diversity	Average Shannon–Weaver information index	Percent polymorphic loci
TZL COMP3 C_0	1.95 ± 0.003	0.14 ± 0.001	0.21 ± 0.002	0.21 ± 0.001	0.33 ± 0.002	95
TZL COMP3 C_4	1.74 ± 0.005*	0.23 ± 0.002*	0.19 ± 0.002*	0.24 ± 0.002*	0.28 ± 0.003*	74
TZL COMP4 C_0	1.97 ± 0.002	0.14 ± 0.001	0.21 ± 0.002	0.22 ± 0.002	0.34 ± 0.002	97
TZL COMP4 C_4	1.93 ± 0.003*	0.14 ± 0.001	0.21 ± 0.002	0.21 ± 0.001*	0.32 ± 0.003*	93

*Significantly different ($P < 0.05$) from C_0 using *t*-test.

TZL COMP4 at C_4 (Table 4). Marked differences were detected in the number of markers deviating from HWE across chromosomes with chromosome 1 having the highest number within the two composites. The percentage of pairs of markers with significant ($P < 0.001$) LD changed from 1.8 to 2.8% in TZL COMP3 and from 1.9 to 2.2% in TZL COMP4 (Table 5). As a consequence, LD increased by 54% in TZL COMP3 and by 17% in TZL COMP4 at C_4 (Table 5). The change in the number of SNP pairs with significant LD at C_4 of each composite did not follow any consistent trend across chromosomes. The distributions of correlations (r^2) between alleles at two loci were found to be moderate with a mean value of 0.40 at C_0 and 0.5 at C_4 of the two composites.

Population structure and genetic divergence

Results of population structure analysis of the 400 S_1 lines using ADMIXTURE that was run by varying K from 2 to 10 are shown in Fig. 2. The most significant drop at ΔK occurred when $K = 4$, indicating that the S_1 lines were assigned into four clearly defined groups corresponding to the two selection cycles each within the tropical composites. The results of distance-based dendrogram also assigned the S_1 lines into four distinct groups, supporting the consistency of the result of ADMIXTURE-based clustering method (Fig. 3). To further assess the reliability of the groups obtained using the model-based population structure analysis and hierarchical clustering algorithms, we conducted discriminant analysis of principal components and found four genetically divergent groups (Fig. S1a, b). Partitioning the total SNP variation using analysis of molecular variance revealed significant ($P < 0.001$) differences between selection cycles and tropical composites (Table 6). When the total SNP variance was partitioned using the four groups, 80% of the total variance was attributable to the differences among the S_1 lines within selection cycles of the two composites, whereas the remaining 20% of the total variance was ascribed to differences between selection cycles within the two composites (Table 6). Further assessment using the F_{ST} test also found significant ($P < 0.001$) genetic differentiation between selection cycles within the two complementary composites accounting for 20% of the total variation among the 400 S_1 lines. The divergence between C_0 and C_4 accounted for 10.5% of the total variation in TZL COMP3 and 13.5% of the total variation in TZL COMP4 (Table 7). The difference between the original cycles (C_0) of the two composites explained 8.5% of the total molecular variance, whereas those between TZL COMP4 C_4 and TZL COMP3 C_0 and between TZL COMP3 C_4 and TZL COMP4 C_0 explained 8.1 and 15.5% of the total variation, respectively. The divergence between C_4 of the two composites represented the lowest value, accounting for 5.5% of the total observed molecular variance (Table 7).

Discussion

Maize breeders at IITA developed two broad-based tropical maize composites, which were subjected to four cycles of RRS. The present study was conducted to examine the extent of genetic diversity and changes in allele composition and genetic structure of 100 randomly selected S_1 lines each from the original and advanced selection cycles of the two tropical composites using 10 579 SNP markers. The different measures of genetic diversity, including average number of alleles at each locus, observed heterozygosity, gene diversity, Shannon–Weaver diversity index and percent polymorphic loci, either changed slightly or stayed the same after four cycles of RRS in TZL COMP 4, whereas there was a significant reduction in TZL COMP3 for most of the different measures. The substantial genetic diversity retained in the advanced selection cycle of each composite could arise from the effect of large effective population size used and reasonable selection intensity applied during improvements of the two composites. In our RRS programme, evaluation of 165–300 testcrosses of lines derived from the two composites followed by selection of the highest yielding 12–17% of the testcrosses for recombination during each selection cycle (Menkir et al. 2015) could slow the rate of loss in genetic diversity after four cycles of selection.

The observed changes in allele composition at C_4 of each composite could be attributed to the effect of the RRS method used. After four cycles of RRS, the proportion of alleles occurring at low frequencies decreased with selection consistent with the report of Pinto et al. (2003a). Although most of the unique alleles detected at C_0 in each composite were lost at C_4 possibly due to selection for favourable combination of alleles, their presence at low frequencies in advanced selection cycle of particularly TZL COMP4 is desirable for sustainable response from selection (Lucas et al. 2013) and as important source of novel alleles of desirable traits that can be used to broaden the genetic base of maize breeding populations (Pinto et al. 2003a). In contrast, Labate et al. (1997) found that a large fraction of unique RFLP alleles remained at reasonable frequencies in Iowa Stiff Stalk Synthetic and Iowa Corn Borer Synthetic after 12 cycles of RRS.

The similar direction in allele frequency changes observed in TZL COMP3 and TZL COMP4 can be explained by increase in alleles having positive effect on traits under selection, while at the same time a decrease in those with negative effect on target traits. Delaney and Bliss (1991) also reported that the frequency of alleles with large effects should increase or decrease faster than the frequency of alleles with relatively small effects if substantial genetic variability is available in the population under selection. However, the changes in allele frequencies and loss of unique and other alleles at C_4 were greater in TZL COMP3. TZL COMP3 was formed by crossing an orange maize variety

Table 4: The number of markers deviating from Hardy–Weinberg equilibrium has positive and negative fixation indices in C_0 and C_4 of two tropical maize composites in each chromosome

	TZL COMP3 C_0				TZL COMP3 C_4				TZL COMP4 C_0				TZL COMP4 C_4			
		Fixation index				Fixation index				Fixation index				Fixation index		
Chromosomes	Significant from HWE	Pos	Neg	Mean	Significant from HWE	Pos	Neg	Mean	Significant from HWE	Pos	Neg	Mean	Significant from HWE	Pos	Neg	Mean
1	172	462	720	−0.02	147	329	626	−0.03	169	451	736	−0.01	153	344	774	−0.03
2	109	284	461	−0.01	90	212	384	−0.04	105	282	472	−0.02	84	200	526	−0.05
3	89	253	503	−0.02	91	176	393	−0.04	109	253	509	−0.02	90	224	505	−0.03
4	86	241	417	−0.02	74	165	319	−0.03	101	266	406	0.00	82	191	451	−0.04
5	99	265	496	−0.02	94	193	328	−0.03	118	283	507	−0.02	100	248	511	−0.03
6	80	189	422	−0.02	78	180	292	−0.03	82	209	398	−0.03	90	200	389	−0.02
7	104	225	380	0.00	74	156	321	−0.02	73	223	388	−0.01	87	176	414	−0.04
8	77	233	364	−0.02	83	190	297	−0.03	87	220	392	−0.03	67	161	418	−0.06
9	83	195	294	0.00	70	136	242	−0.04	71	177	326	−0.03	77	147	347	−0.04
10	67	198	309	−0.01	56	147	268	−0.02	68	182	350	−0.01	79	168	332	−0.03
Total	966	2545	4366	−0.14	857	1884	3470	−0.31	983	2546	4484	−0.18	909	2059	4667	−0.37

Pos, positive fixation index; Neg, negative fixation index.

from Thailand (Suwan 1-SR) with an adapted variety (TZB-SR) developed at IITA (Menkir et al. 2015). As the objective of the long-term RRS programme was to develop a pair of source populations for white maize inbred lines with superior combining ability, only white lines with desirable agronomic traits were selected for testcross evaluation and recombination beginning with the formation of C_2 in TZL COMP3. The selection against ears with yellow to orange kernel colour could lead to elimination of an array of alleles, contributing to the observed more changes in allelic frequency and loss of alleles at C_4 in TZL COMP3. Genetic drift could also contribute to changes in allele frequencies.

Some SNPs that were not in HWE were found in the selection cycles of both TZL COMP3 and TZL COMP4, consistent with reports of Labate et al. (2000) and Reif et al. (2004). The number of markers that were not in HWE decreased at C_4 in the two composites, with the reduction being greater in TZL COMP3 possibly due to loss of alleles associated with yellow to orange kernel colour during selection of superior progeny for recombination in this composite. The most probable cause of departure from HWE of the markers is the positive assortative mating resulting from the enormous genetic variability in flowering time present in the two composites (Labate et al. 2000). The deviations from HWE in the two composites tended to be more towards excess heterozygosity possibly because selection favoured the heterozygous plants over the homozygous plants as reported by Peña-Asin et al. (2013). Intermating heterozygous plants maximize the probability of effective crossover and reduce the danger of fixing unfavourable alleles (Weber 1983).

The substantial amount of LD blocks in C_0 of both composites was probably generated during the development of the two composites. As TZL COMP 3 and TZL COMP4 were formed from crosses of broad-based populations with diverse genetic origin (Menkir et al. 2015), their original selection cycle retained significant LD mainly because of the additional need for several cycles of recombination to dissipate LD. The number of pairs of SNPs in LD varied considerably across chromosomes possibly due to differences in recombination rates among chromosomes as detected in Dent and Flint maize populations by Bauer et al. (2013). The total number of pairs of markers in LD increased with selection in the two composites, consistent with the results reported for other populations improved through RRS (Romay et al. 2012, Peña-Asin et al. 2013). Selection, hitchhiking effects of alleles linked with selected alleles, selection of favourable combination of alleles (epistasis) and random genetic drift have been implicated in generating increased LD after selection (Falconer and Mackay 1996, Labate et al. 2000, Butrón et al. 2005, Romay et al. 2012, Peña-Asin et al. 2013). The increase in r^2 values at C_4 in the two composites provides an indication of more frequent co-occurrence of allele at two loci (Hill and Robertson 1968). However, the LD generated after four cycles of improvements of the two composites will not limit further genetic gain from RRS with efficient recombination procedures that promotes selection for favourable combination of alleles (Falke et al. 2007).

The RRS programme that has been underway for more than 20 years has created a clearly defined structure within and between TZL COMP3 and TZL COMP4. The model-based approach as well as cluster and multivariate analyses showed a high degree of concordance in assigning the 400 S_1 lines into four non-overlapping groups, consistent with the results reported by Labate et al. (1999) who found a clear separation of individuals belonging to different populations and cycles of selection into

Table 5: Summary of LD blocks per chromosome as measured by r^2 using 10 579 SNP markers in 400 tropical maize S_1 lines

		Number of blocks[1]							
Chromosomes	Number of SNP pairs	TZL COMP3 C_0	r^2	TZL COMP3 C_4	r^2	TZL COMP4 C_0	r^2	TZL COMP4 C_4	r^2
1	60076	1204	0.42	1776	0.58	1452	0.42	1247	0.49
2	39151	608	0.42	1013	0.46	698	0.39	827	0.46
3	39351	592	0.44	831	0.53	566	0.44	653	0.48
4	34751	558	0.46	1029	0.43	631	0.45	798	0.54
5	40351	982	0.44	783	0.46	950	0.42	1152	0.47
6	31801	604	0.45	1099	0.50	590	0.45	867	0.46
7	31551	483	0.44	846	0.49	594	0.45	518	0.49
8	31701	510	0.46	1121	0.48	387	0.44	576	0.53
9	25851	500	0.38	757	0.42	339	0.38	519	0.55
10	27051	449	0.44	741	0.47	519	0.42	744	0.55
Total	361635	6490	4.35	9996	4.82	6726	4.26	7901	5.02
Average	36164	649	0.40	1000	0.50	673	0.40	790	0.50

[1]An LD block consists of a sequence of markers for which all pairs of adjacent loci are significant ($P < 0.01$).

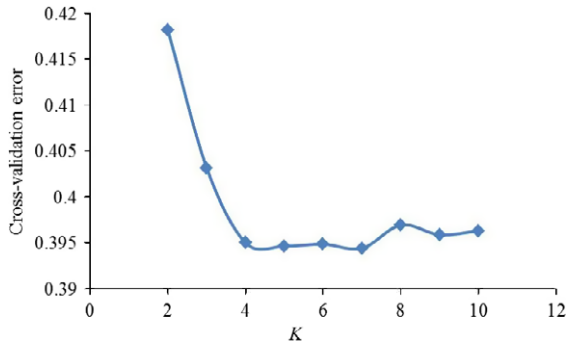

Fig. 2: Determination of the optimal number of clusters using ADMIXTURE with 10-fold cross-validation error rates for $k = 2$ to $k = 12$, showing that the least error rate was produced by $k = 4$ [Color figure can be viewed at wileyonlinelibrary.com]

distinct groups after 12 cycles of RRS. These results highlight the level of accuracy of the results of the diallel studies that were used as the bases to form the two distinct composites and the success of the selection procedures employed thereafter over the years. Also, the observed large proportion of the molecular variation within than among selection cycles was a reflection of the broad genetic base of the two composites, consistent with the results of Reif et al. (2003) and Berilli et al. (2011) who also found more than 85% of the variation among progeny within populations and less than 15% of the variation among maize populations. Our results showed that the advanced population in each composite substantially diverged after four cycles of RRS. The fact is that 25–50 S_1 lines derived from each composite were recombined to form new selection cycle in the two composites (Menkir et al. 2015); considerable genetic diversity was maintained across selection cycles. This is in agreement with

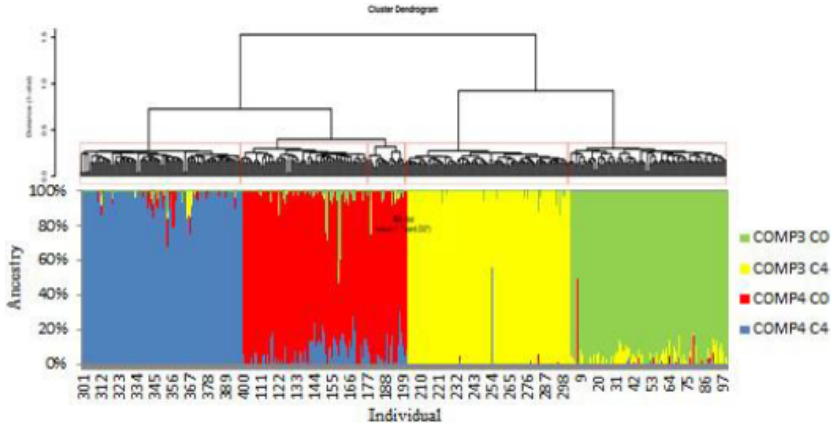

Fig. 3: Population structure of 400 S1 lines from original and improved cycles of TZL COMP3 and TZL COMP4 (Ward's minimum variance method) dendrogram and the individual ancestry estimated from ADMIXTURE analysis. Individuals are partitioned into segments corresponding to the inferred membership in $k = 4$ genetic clusters as indicated by the colours [Color figure can be viewed at wileyonlinelibrary.com]

Table 6: Hierarchical analysis of molecular variance (AMOVA) and Wright's fixation index (F_{ST}) for 400 S_1 lines derived from two reciprocal composites based on 10 579 SNP markers

Level of variation	df	Sum of squares	Mean squares	Variance components	Percentage of variation	F_{ST}
Among populations	3	102096.26	34032.09	326.90	20	0.20***
Within populations	396	531408.94	1341.94	1341.94	80	
Total	399	633505.20		1668.84	100	

***P-value <0.001.

Table 7: F_{ST} divergences estimates using 10 579 SNPs loci between selection cycles within and between the two composites

	TZL COMP4 C_4	TZL COMP4 C_0	TZL COMP3 C_4	TZL COMP3 C_0
TZL COMP4 C_4	0.000			
TZL COMP4 C_0	0.135	0.000		
TZL COMP3 C_4	0.055	0.155	0.000	
TZL COMP3 C_0	0.081	0.085	0.105	0.000

Guzman and Lamkey (1999) who emphasized the importance of using large superior progeny for recombination to prevent loss of genetic variability due to the potential effect of genetic drift that can limit further genetic advance in the selection programme.

The diversities observed in allele frequencies, clustering pattern, genetic divergence and population structure of the original cycles of TZL COMP3 and TZL COMP4 depict that they belong to two distinct heterotic groups (Tuxpeño dent and Caribbean flint). The maintenance of two separate genetic pools allows different alleles to be fixed in each composite and guarantees a heterozygous condition for these loci to develop inter-population hybrids in future. Further improvements of the two composites are expected to increase the probability of developing divergent new inbred lines with good combining ability whose single-cross hybrids will outperform the best commercial hybrids adapted to the savannas in WCA. The development of such divergent maize inbred lines will maximize heterosis in hybrids through expression of the dominance gene effects attributed to the heterozygous loci (Falconer and Mackay 1996, Lamkey and Edwards 1999).

In summary, the RRS that has been underway for more than two decades created significant divergence between cycles within the same composite and clear differentiation between the complementary composites. The changes in genetic structures of the two composites as a result of accumulation of favourable complementary alleles will be useful to increase mean performance and sustainable genetic gain in productivity of maize in WCA. This knowledge of the genetic composition and structure of TZL COMP3 and TZL COMP4 using SNPs enhances our understanding on the potential progress that can be achieved in the long-term RRS programme. As the improved selection cycles of the two composites are endowed with enormous genetic diversity, the prospect seems to be promising for attaining further progress in the development of divergent elite inbred lines from the two composites for use as parents of productive hybrids.

Acknowledgements

This report is a part of PhD thesis research funded by the Alliance for a Green Revolution in Africa (AGRA) at West Africa Centre for Crop Improvement (WACCI), University of Ghana, Legon, and the International Institute of Tropical Agriculture (IITA). The authors are grateful to all staff in the Bioscience Center of IITA for their contributions to this study.

References

Alene, A. D., A. Menkir, S. O. Ajala, B. Badu-Apraku, A. S. Olanrewaju, V. M. Manyong, and N. Abdou, 2009: The economic and poverty impacts of maize research in West and Central Africa. Agric. Econ. **40**, 535—550.

Alexander, D. H., J. Novembre, and K. Lange, 2009: Fast model-based estimation of ancestry in unrelated individuals. Genome Res. **19**, 1655—1664.

Auta, S. J., J. G. Akpoko, H. Mamman, and H. A. Hassan, 2001: Production trends of small-scale maize farmers under the Sasakawa Global 2000 Programme in Kaduna State of Nigeria, In: B. Badu-Apraku, M. A. B. Fakorede, M. Ouedraogo, and R. J. Carsky (eds), Impact Challenges and Prospects of Maize Research and Development in West and Central Africa, 412—418. Proceedings of a regional maize workshop, WECAMAN/IITA, Cotonou, Benin Republic.

Bauer, E., M. Falque, H. Walter, C. Bauland, C. Camisan, L. Campo, N. Meyer, N. Ranc, R. Rincent, W. Schipprack, et al., 2013: Intraspecific variation of recombination rate in maize. Genome Biol. **14**, 1—17.

Berilli, A. P. C. G., M. G. Pereira, L. S. A. Gonçalves, K. S. Cunha da, H. C. C. Ramos, G. A. Souza Filho, and A. T. Amaral Júnior do, 2011: Use of molecular markers in reciprocal recurrent selection of maize increases heterosis effects. Genet. Mol. Res. **10**, 2589—2596.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler, 2007: TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics **23**, 2633—2635.

Butrón, A. R., P. Tarrio, A. Revilla, B. Ordás, and R. A. Malvar, 2005: Molecular changes in the maize composite EPS12 during selection for resistance to pink stem borer. Theor. Appl. Genet. **110**, 1044—1051.

Byerlee, D., and D. Jewell, 1997: The technological foundation of the revolution. In: D. Byerlee, and C. K. Eicher (eds), Africa's Emerging Maize Revolution, 127—143. Lynne Rienner Publishers Inc, Boulder, CO.

Comstock, R. E., H. F. Robinson, and P. H. Harvey, 1949: A breeding procedure designed to make maximum use of both general and specific combining ability. Agron. J. **41**, 360—367.

Delaney, D. E., and F. A. Bliss, 1991: Selection for increased percentage phaseolin in common bean. II. Changes in frequency of seed protein alleles with S_1 family recurrent selection. Theor. Appl. Genet. **81**, 306—311.

Eberhart, S. A., M. N. Harrison, and F. Ogada, 1967: A comprehensive breeding system. Theor. Appl. Genet. **37**, 169—174.

Efron, Y., S. K. Kim, J. M. Fajemisin, J. H. Mareck, C. Y. Tang, Z. T. Dabrowski, H. W. Rossel, I. W. Buddenhagen, and G. Thottappilly, 1989: Breeding for resistance to maize streak virus: a multidisciplinary team approach. Plant Breeding **103**, 1—36.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, and K. Kawamoto, 2011: A robust, simple Genotyping- by-Sequencing (GBS) approach for high diversity species. PLoS ONE **6**, e19379.

Excoffier, L., P. E. Smouse, and J. M. Quattro, 1992: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics **131**, 479—491.

Falconer, D. S., and T. F. C. Mackay, 1996: Introduction to Quantitative Genetics. 4th edn. Longman Sci. and Tech, Harlow.

Falke, K. C., H. P. Maurer, A. E. Melchinger, H. P. Piepho, C. Flachenecker, and M. Frisch, 2007: Linkage disequilibrium in two European F_2 flint maize populations under modified recurrent full-sib selection. Theor. Appl. Genet. **115**, 289—297.

Glaubitz, J. C., T. M. Casstevens, F. Lu, J. Harriman, R. J. Elshire, Q. Sun, and E. S. Buckler, 2014: TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS ONE **9**, e90346.

Goodman, M. M., 1985: Exotic germplasm: status, prospects and remedies. Iowa State J. Res. **59**, 494—527.

Guo, S., and E. Thompson, 1992: Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics **48**, 361—372.

Guzman, P. S., and K. R. Lamkey, 1999: Predicted gains from recurrent selection in the BS11 maize population. Maydica **44**, 93—99.

Hallauer, A. R., 1985: Compendium of recurrent selection methods and their application. CRC Crit. Rev. Plant Sci. **3**, 1—33.

Hallauer, A. R., and S. A. Eberhart, 1970: Reciprocal full-sib selection. Crop Sci. **10**, 315—316.

Heredia-Diaz, O., A. Alsirt, E. H. Darrah, and E. H. Coe, 1996: Allelic frequency changes in the MoSCSS maize synthetic in response to bi-

directional recurrent selection for rind penetrometer resistance. Maydica **41**, 65—76.

Hill, W. G., and A. Robertson, 1968: Linkage disequilibrium in finite populations. Theor. Appl. Genet. **38**, 226—231.

Hinze, L. L., S. Kresovich, J. D. Nason, and K. R. Lamkey, 2005: Population genetic diversity in a maize reciprocal recurrent selection program. Crop Sci. **45**, 2435—2442.

Jombart, T., 2008: Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics **24**, 1403—1405.

Jombart, T., S. Devillard, and F. Balloux, 2010: Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. **11**, 1—15.

Jones, E. S., H. Sullivan, D. Bhattramakki, and J. S. Smith, 2007: A comparison of simple sequence repeat and single nucleotide polymorphism marker technologies for the genotypic analysis of maize (*Zea mays* L.). Theor. Appl. Genet. **115**, 361—371.

Kim, S. K., 1997: Achievements, challenges and future direction of hybrid maize research and production in West and Central Africa. In: B. Badu-Apraku, M. O. Akoroda, M. Ouedraogo, and F. M. Quin (eds), Proceedings of a Regional Maize Workshop, 42—82. IITA, Cotonou, Benin Republic.

Labate, J. A., K. R. Lamkey, M. Lee, and W. L. Woodman, 1997: Molecular genetic diversity after reciprocal recurrent selection in BSSS and BSCB1 maize populations. Crop Sci. **37**, 416—423.

Labate, J. A., K. R. Lamkey, M. Lee, and W. L. Woodman, 1999: Temporal changes in allele frequencies in two reciprocally selected maize populations. Theor. Appl. Genet. **99**, 1166—1178.

Labate, J. A., K. R. Lamkey, M. Lee, and W. L. Woodman, 2000: Hardy-Weinberg and linkage equilibrium estimates in the BSSS and BSCB1 random mated populations. Maydica **45**, 243—255.

Lamkey, K. R., and J. W. Edwards, 1999: The quantitative genetics of heterosis. In: J. G. Coors, S. Pandey, M. V. Ginkel, A. R. Hallauer, D. C. Hess, K. R. Lamkey, A. E. Melchinger, and C. W. Stuber (eds), Proceedings of the international symposium on the genetics and exploitation of heterosis in crops, 31—48. CIMMYT, Mexico City, Mexico.

Liu, K., and S. V. Muse, 2005: Power Marker: an integrated analysis environment for genetic marker analysis. Bioinformatics **21**, 2128—2129.

Lucas, C. J., H. Zhao, M. Schneerman, and S. P. Moose, 2013: Genomic changes in response to 110 cycles of selection for seed protein and oil concentration in maize In: W. B. Philip (ed.), Seed Genomics, 1st edn, 217—236. John Wiley and Sons Inc, Ames, Iowa.

Menkir, A., J. G. Kling, B. Anjorin, F. Ladejobi, and M. Gedil, 2015: Evaluating Testcross Performance and Genetic Divergence of Lines Derived from Reciprocal Tropical Maize Composites. Maydica 60 - Electronic publication, Maydica 60-2015.

MIP, 1996: Maize Improvement Program archival report, 1989-1992. Part I: Maize population improvement CID IITA Ibadan Nigeria.

Ordas, B., R. A. Malvar, R. Díaz, and A. Butrón, 2015: Molecular changes in two maize (*Zea mays* L.) synthetics after reciprocal selection with two alternative methods. Mol. Breed. **35**, 111—119.

Paradis, E., J. Claude, and K. Strimmer, 2004: APE: analyses of phylogenetics and evolution in R language. Bioinformatics **20**, 289—290.

Peakall, R., and P. E. Smouse, 2012: GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update. Bioinformatics **28**, 2537—2539.

Peña-Asin, J., A. Alvarez, and B. Ordas, 2013: Molecular changes during intra and inter recurrent selection of two populations of maize: one adapted and one non adapted to the selection environment. Euphytica **193**, 359—376.

Pinto, L. R., M. L. C. Vieira, C. L. de Souza Jr, and A. P. de Souza, 2003a: Genetic diversity assessed by microsatellites in tropical maize populations submitted to a high-intensity reciprocal recurrent selection. Euphytica **134**, 277—286.

Pinto, L. R., M. L. C. Vieira, C. L. de Souza Jr, and A. P. de Souza, 2003b: Reciprocal recurrent selection effects on the genetic structure of tropical maize populations assessed at microsatellite loci. Genet. Mol. Biol. **26**, 355—364.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, et al., 2007: PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. **81**, 559—575.

R Development Core Team, 2010: R: A Language and Environment for Statistical Computing, Reference Index Version 2.12.1. R foundation for statistical computing, Vienna, Austria. ISBN 3-90005107-0. Available at: http://www.R-project.org.

Reif, J. C., A. E. Melchinger, X. C. Xia, M. L. Warburton, D. A. Hoisington, S. K. Vasal, G. Srinivasan, M. Bohn, and M. Frisch, 2003: Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. Crop Sci. **43**, 1275—1282.

Reif, J. C., X. C. Xia, A. E. Melchinger, M. L. Warburton, D. A. Hoisington, D. Beck, M. Bohn, and M. Frisch, 2004: Genetic diversity determined within and among CIMMYT maize populations of tropical, subtropical, and temperate germplasm by SSR markers. Crop Sci. **44**, 326—334.

Romay, M. C., A. Butrón, A. Ordas, P. Revilla, and B. Ordas, 2012: Effect of recurrent selection on the genetic structure of two broad-based Spanish maize populations. Crop Sci. **52**, 1493—1502.

Rusike, J., and C. K. Eicher, 1997: Institutional innovations in the maize seed industry. Emerging maize revolution. In: D. Byerlee, and C. K. Eicher (eds), Africa's Emerging Maize Revolution, 127—143. Lynne Rienner Publishers Inc, Boulder, CO.

Saghai-Maroof, M. A., K. M. Soliman, R. A. Jorgensen, and R. W. Allard, 1984: Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. Proc. Natl Acad. Sci. USA **81**, 8014—8018.

SAS Institute, 2010: Statistical Analysis Software (SAS). Users guide, SAS Inst Inc, Cary, NC.

Schnable, P. S., D. Ware, R. S. Fulton, J. C. Stein, F. Wei, S. Pasternak, C. Liang, J. Zhang, L. Fulton, T. T. Graves, et al., 2009: The B73 maize genome: complexity, diversity, and dynamics. Science **326**, 1112—1115.

Smale, M., D. Byerlee, and T. S. Jayne, 2013: Maize Revolutions in sub-Saharan Africa. In: K. Otsuka, and D. F. Larson (eds), An African Green Revolution, 165—195. Springer, Netherlands.

Solomon, K. F., I. Martin, A. Zeppa, and V. Keating, 2010: Temporal genetic structure patterns under reciprocal recurrent selection. Euphytica **176**, 239—249.

Stuber, C. W., R. H. Moll, M. M. Goodman, H. E. Schaffer, and B. S. Weir, 1980: Allozyme frequency changes associated with selection for increased grain yield in maize (*Zea mays* L.). Genetics **95**, 225—236.

Weber, W. E., 1983: The efficiency of repeated crosses within a population of self-fertilizing species. Z. Pflanzenüchtg. **91**, 129—139.

Weir, B. S., 1996: Genetic Data Analysis II: Methods for Discrete Population Genetic Data. Sinauer Associates Incorporated, Sunderland, MA.

Wellhausen, E. J., 1978: Recent development in maize breeding in the tropics. In: D. B. Walden (ed.), Maize Breeding and Genetics, 59—91. John Wiley and Sons, New York.

Wright, S., 1978: Evolution and the Genetics of Populations. Variability within and among natural populations. University of Chicago Press, Chicago.

Zhu, Y. L., Q. J. Song, D. L. Hyten, C. P. Van Tassell, L. K. Matukumalli, D. R. Grimm, S. M. Hyatt, E. W. Fickus, N. D. Young, and P. B. Cregan, 2003: Single-nucleotide polymorphisms in soybean. Genetics **163**, 1123—1134.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. (a) Determination of the optimal number of clusters using DAPC with Bayesian Information Criterion (BIC) estimates for *k*-means clusters ($k = 1$ to $k = 10$). (b) Scatter plot of individuals on the two principal components of DAPC based on the analysis of 400 S1 tropical maize lines using 10 579 SNP markers. The graph represents the individuals as dots and the groups as inertia ellipses. Eigenvalues of the analysis are displayed in inset. Group1 = TZL COMP4 C4, Group 2 = TZL COMP4 C0, Group 3 = TZL COMP3 C4, Group 4 = TZL COMP3 C0.