

# Genetic Diversity among Selected Elite CIMMYT Maize Hybrids in East and Southern Africa

Benhildah P Masuka,<sup>\*</sup> Angeline van Biljon, Jill E. Cairns, Biswanath Das, Maryke Labuschagne, John MacRobert, Dan Makumbi, Cosmos Magorokosho, Mainassara Zaman-Allah, Veronica Ogugo, Mike Olsen, Boddupalli M. Prasanna, Amsal Tarekegne, and Kassa Semagn

## ABSTRACT

Genetic gain within the CIMMYT Eastern and Southern Africa (ESA) hybrid maize (*Zea mays* L.) breeding program from 2000 to 2010 was recently estimated at 0.85 to 2.2% yr<sup>-1</sup> under various environmental conditions. Over 100 varieties were disseminated from CIMMYT to farmers in ESA, hence the need to check genetic diversity and frequency of use of parents to avoid potential narrowing down of the genetic base. Fifty-five parents from CIMMYT ESA used in the hybrids were fingerprinted using genotyping-by-sequencing. Data analysis in TASSEL and MEGA6 generated pairwise genetic distances between parents of 0.004 to 0.4005. Unweighted pair group method with arithmetic mean (UPGMA) analysis produced two clusters (I and II) with two subclusters each (A and B) and two sub-subclusters (IAi and IAii). Principal coordinate analysis produced three clusters where IAi and IIA from the UPGMA analysis formed independent clusters while IAii, IB, and IIB clustered together. Lines were separated by pedigree and origin. Ninety-five percent frequency of pairwise genetic distances ranged between 0.2001 and 0.4000. However, only four of the 55 parents (CML444, CML395, CML312, and CML442) were each used in 15 to 30 of the 52 hybrids evaluated in the genetic gain study. The remaining 51 were used in one to four hybrids. Frequent use of the four parents gave 29 to 58% of the hybrids a narrow genetic base, posing risk in case of pest or disease outbreaks. Parents evaluated do not represent the genetic base of CIMMYT ESA but parents of the best-performing hybrids selected from 2000 to 2010. Breeders should ensure a wide genetic base for released varieties to avoid breakdown in case of pest or disease outbreaks.

B.P. Masuka, J.E. Cairns, B. Das, J. MacRobert, C. Magorokosho, M. Zaman-Allah, and A. Tarekegne, International Maize and Wheat Improvement Centre (CIMMYT), PO Box MP163, Harare, Zimbabwe; A. van Biljon and M. Labuschagne, Dep. of Plant Sciences, Univ. of the Free State, PO Box 339, Bloemfontein 9300, South Africa; B. Das, D. Makumbi, V. Ogugo, M. Olsen, B.M. Prasanna, and K. Semagn, CIMMYT, United Nations Ave., Gigiri PO Box 1041, Village Market-00621, Nairobi, Kenya. Received 7 Sep. 2016. Accepted 13 July 2017. <sup>\*</sup>Corresponding author (trimasuka@gmail.com). Assigned to Associate Editor Margaret Worthington.

**Abbreviations:** CKL, CIMMYT Kenya Line; CML, CIMMYT maize line; CZL, CIMMYT Zimbabwe line; ESA, Eastern and Southern Africa; GBS, genotyping-by-sequencing; MLN, maize leaf necrosis; MSV, *maize streak virus*; OPV, open-pollinated variety; PCoA, principal coordinate analysis; PCR, polymerase chain reaction; PC, principal component; SNP, single-nucleotide polymorphism; SSA, Sub-Saharan Africa; SSR, simple sequence repeats; TASSEL, trait analysis by association, evolution, and linkage; UPGMA, unweighted pair group method with arithmetic mean; ZEWA, Zimbabwe Early White A Population; ZEWB, Zimbabwe Early White B population; ZM, Zimbabwe maize.

As the world population is growing, total crop production globally is not meeting the rising demand for food. The rate of growth in global crop production is below what was recommended to cope with the rising demand (Pingali and Pandey, 2001; Ray et al., 2013). In Sub-Saharan Africa (SSA), where >300 million people, mostly resource-poor farmers, rely on maize (*Zea mays* L.) for food, feed, and livelihood (Bänziger and Aurus, 2007; M'mboyi et al., 2010; Rovere et al., 2010; Tefera et al., 2011), grain yield is estimated at <1.8 t ha<sup>-1</sup>, which is among the lowest in the world (Smale et al., 2011; Cairns et al., 2012; FAO, 2013). Such low production is a result of a number of bottlenecks, including

Published in Crop Sci. 57:2395–2404 (2017).  
doi: 10.2135/cropsci2016.09.0754

© Crop Science Society of America | 5585 Guilford Rd., Madison, WI 53711 USA  
This is an open access article distributed under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

periodic drought, high incidence of biotic stresses (diseases, insect pests, and weeds), poor soil fertility, scarcity and high cost of irrigation, and farmers' inability to access and afford quality seeds and fertilizers (Semagn, 2014).

The CIMMYT maize breeding program in SSA, using conventional pedigree methods, has been developing both inbred lines that are used as parents for hybrid development and open-pollinated varieties (OPVs) as byproducts of the hybrid breeding pipeline. The CIMMYT germplasm is a public good and is freely available to all partners. Breeders from the National Agricultural Research Systems (NARS) and the private sector in several countries in SSA use CIMMYT inbred lines for developing, testing, and releasing hybrids in their own countries. Over 150 maize hybrids and OPVs have been released by CIMMYT and partners in several countries in the region (CIMMYT, 2015; Fisher et al., 2015).

Genetic gain within the CIMMYT ESA maize breeding program was recently estimated. The annual genetic gain in grain yield over a period of 11 yr (2000–2010) was 109 kg ha<sup>-1</sup> yr<sup>-1</sup> or 1.4% under optimal conditions, 20.9 ha<sup>-1</sup> yr<sup>-1</sup> or 0.6% under low-nitrogen conditions, 141.3 ha<sup>-1</sup> yr<sup>-1</sup> or 2.2% under *maize streak virus* (MSV) conditions, and 22 and 32.5 ha<sup>-1</sup> yr<sup>-1</sup> or 0.9% under both managed and random drought stress, respectively (Masuka et al., 2017a). The CIMMYT produces OPVs as byproducts of the hybrid breeding pipeline; therefore, the genetic gain for OPVs was also evaluated and recently reported. Genetic gains in the early maturity group OPVs under optimal, random drought, low N, and MSV were estimated at 1.76, 1.21, 3.11, and 4.62% yr<sup>-1</sup>. In the intermediate to late maturity OPVs group, genetic gains under optimal, random drought, low N, and MSV were estimated at 1.35, 2.09, 1.74, and 2.32% yr<sup>-1</sup>. In terms of actual yield, gains per year yield progress in the early maturity OPVs group was 109.9, 25.3, 84.8, and 192.9 kg ha<sup>-1</sup> yr<sup>-1</sup> under optimal, random drought, low N, and MSV, respectively. In the intermediate to late maturity OPVs group yield, progress was estimated to be 79.1, 51.1, 53.0, and 108.7 kg ha<sup>-1</sup> yr<sup>-1</sup> under optimal, random drought, low N, and MSV (Masuka et al., 2017b). To help explain such considerable annual increase in genetic gain, it is essential to understand the extent of genetic differences and patterns of relationships of the parental lines used to make the hybrids that contributed towards heterosis in these crosses and the population byproducts.

Genetic diversity is important in a breeding program. A wide genetic base provides variation that enables breeding for solutions when problems like pest and disease outbreaks arise. When the genetic base narrows down, variation becomes limited and breeders may fail to find relevant variation to address the new or arising problems. It is therefore essential to maintain a wide genetic base and constantly check for diversity in a breeding program.

For the CIMMYT ESA program, hybrids are produced by crossing lines from different heterotic groups for maximum heterosis. There are three main heterotic groups used in CIMMYT ESA: A, B, and AB. Grouping is done by crossing new (introduced or developed) materials to know A and B testers. Based on observed combining ability, the new material is classified as A or B if they combine well with testers B and A, respectively, or AB if they combine well with both the A and B testers. Information on heterotic grouping is available in the CIMMYT maize line (CML) handbook (CIMMYT, 2005).

Single-nucleotide polymorphism (SNP) markers are a powerful tool for many genetic applications, including genetic diversity, genetic relationships, and population structure. Single-nucleotide polymorphism data can be obtained using one of the numerous uniplex or multiplex SNP genotyping platforms or using genotyping-by-sequencing (GBS). Advances in next-generation sequencing technologies have significantly reduced the costs of DNA sequencing that allows the development of GBS, which is being increasingly adopted for discovery applications (Elshire et al., 2011; Poland and Rife, 2012). Genotyping-by-sequencing is a 96 to 384 multiplexed system for constructing reduced representation libraries for the Illumina next-generation sequencing platform. It reduces sample handling by multiplexing up to 384 samples using unique and inexpensive barcoding, which in turn reduces polymerase chain reaction (PCR) and purification steps and avoids size fractionation. The method uses restriction enzymes to reduce genome complexity, avoids the repetitive fraction of the genome, and produces huge marker data points (Elshire et al., 2011). The objective of this study was to determine the genetic distance and relationship of parental lines of 52 hybrids released in SSA from 2000 to 2010 and understand why some of the inbred lines were more frequently used in hybrid formation than others.

## MATERIALS AND METHODS

### DNA Extraction and Genotyping

Fifty-five parents—53 inbred lines and two F<sub>2</sub> populations listed in Table 1—that were used as parents in developing 52 of the 67 three-way and double-cross hybrids released in SSA between 2000 and 2010, recently evaluated for genetic gain (Masuka et al., 2017a), were analyzed in this study. Parents analyzed could further be classified according to the program in which they were developed; that is, CKL for CIMMYT Kenya prerelease CML lines developed in Kenya, CZL for CIMMYT Zimbabwe prerelease CML lines developed in Zimbabwe, and CML for released CIMMYT Maize Lines developed or introduced in the Kenya or Zimbabwe breeding programs. Each parent was submitted for analysis from the program that developed them, except for CKL05003 sourced from the Mexico genebank and CML539c from the Zambia National Programme that were already in the GBS database (analyzed), for which data was

**Table 1. Pedigrees, heterotic grouping, source, heterogeneity, and frequency of use of parental lines and F<sub>2</sub> populations in 52 of the 67 hybrids studies in the genetic gain study of the CIMMYT Eastern and Southern Africa maize breeding program for the 2000 to 2010 era.**

No.	Line or population	Heterotic group†	Source	Percentage heterogeneity	Proportion of hybrids with line
				%	
1	CML444	B	Zimbabwe	4.7	58
2	CML144		Zimbabwe	0.2	2
3	CML181		Zimbabwe	10.5	2
4	CML202		Zimbabwe	0.9	6
5	CML312	A	Zimbabwe	0.4	31
6	CML395	B	Zimbabwe	0.1	31
7	CML442	A	Zimbabwe	3.0	29
8	CML443	AB	Zimbabwe	4.5	2
9	CML445	AB	Zimbabwe	4.2	6
10	CML488	AB	Zimbabwe	1.1	6
11	CML489	B	Zimbabwe	0.7	4
12	CZL00001‡		Zimbabwe	7.6	8
13	CZL00003		Zimbabwe	12.6	4
14	CZL02012		Zimbabwe	2.6	4
15	CZL03002		Zimbabwe	13.3	2
16	CZL03003		Zimbabwe	0.3	2
17	CZL03004		Zimbabwe	1.4	2
18	CZL03007		Zimbabwe	1.1	2
19	CZL03018		Zimbabwe	3.7	2
20	CZL03021		Zimbabwe	0.4	2
21	CZL04001		Zimbabwe	20.1	2
22	CZL04002		Zimbabwe	3.1	2
23	CZL04006		Zimbabwe	1.9	6
24	CZL04021		Zimbabwe	0.6	4
25	CZL054	A	Zimbabwe	3.9	2
26	CZL057		Zimbabwe	0.5	2
27	CZL0610	B	Zimbabwe	1.7	2
28	CZL0617		Zimbabwe	0.2	2
29	CZL0619	A	Zimbabwe	2.8	2
30	CZL0713	B	Zimbabwe	2.2	2
31	CZL99014		Zimbabwe	2.6	2
32	CML539b	A	Zimbabwe	25.0	2
33	CML181-dent		Zimbabwe	2.6	2
34	CML536	A	Zimbabwe	0.8	2
35	CML159		Zimbabwe	0.3	2
36	CML440	AB	Zimbabwe	0.3	8
37	CKL05003§	B	Mexico	2.2	6
38	CZL00002	B	Mexico	0.1	2
39	CKL08001	A	Kenya	1.2	2
40	CKL08006	A	Kenya	2.2	2
41	CML216	AB	Kenya	0.1	4
42	CZL04005		Kenya¶	0.4	2
43	CML539c	A	Kenya	0.3	2
44	CKL05004	B	Kenya	0.3	4
45	CKL05007	B	Kenya	0.3	6
46	CKL05019	A	Kenya	0.5	2
47	P100C6-200		Kenya	2.1	4
48	CML78	A	Kenya	0.4	4
49	ZEWBc1F2	B	Kenya	25.5	2
50	CML197		Kenya	0.7	4
51	CKL05017	A	Kenya	0.3	2
52	CKL05018	A	Kenya	0.5	4
53	CKL05022	A	Kenya	0.4	4
54	ZEWAac1F2	A	Kenya	0.6	2
55	CML539	A	Zambia	0.7	2

† Heterotic group A combines well with B, B with A, and AB with both A and B. Where group is not indicated, the grouping was not reported in the official CIMMYT maize line (CML) handbook.

‡ CZL, CIMMYT Zimbabwe Line (advanced line developed in Kenya that may be released as a CML).

§ CKL, CIMMYT Kenya Line (advanced line developed in Kenya that may be released as a CML).

¶ CZL04005 was developed in Zimbabwe but submitted for analysis by the Kenyan program as one of the parents in their breeding program. Likewise, some CZL and CKL lines were submitted from the CIMMYT Mexico gene bank along with other materials from this study or earlier for other studies, and the genotyping-by-sequencing data was retrieved from the library for analysis.

retrieved from the library for analysis. Each parent was represented by bulking of approximately equal amounts of leaf tissue from 10 greenhouse-grown seedlings at the three- to four-leaf stage. Genomic DNA was extracted using a modified version of the cetyl trimethyl ammonium bromide (CTAB) method of CIMMYT protocol, as described elsewhere (Semagn, 2014). DNA concentration was measured using the Quant-iT PicoGreen dsDNA assay kit (Invitrogen) and the Tecan Infinite F200 Pro Plate Reader and normalized to 50 ng  $\mu\text{L}^{-1}$  by adding the required volume of 0.1 TE (10 mM Tris-HCl, pH 7.5 and 0.1 mM ethylenediaminetetraacetic acid [EDTA], pH 8.0). The quality of the extracted DNA was checked by digesting 250 ng of the genomic DNA from eight randomly selected samples with 3.6 units of ApeKI restriction enzyme (New England Biolabs) at 75°C for 3 h. Digested DNA samples, along with Lambda DNA digested with Hind III, were run on a 1% agarose gel containing 0.3  $\mu\text{g mL}^{-1}$  GelRed (Biotium) and visualized. Fifty microliters of the normalized DNA was transferred into a twin.tec PCR 96-well plate (Eppendorf, Hauppauge) and shipped to the Institute for Genomic Diversity (Cornell University, Ithaca, NY) for genotyping. DNA samples were genotyped using GBS at Cornell University as described by Elshire et al. (2011).

## Data Analysis

Imputed GBS data was received from the Institute for Genomic Diversity, Cornell University, for 955,690 loci per line. Since GBS generates a large percentage of uncalled genotypes, the missing data were imputed by the Institute for Genomic Diversity using an algorithm that searched for the closest neighbor in small SNP windows across the maize database available at the Institute for Genomic Diversity (Romay et al., 2013). The GBS data were filtered using a minor allele frequency of 0.01 and a minimum count of 50 lines using trait analysis by association, evolution, and linkage (TASSEL) version 5.0.8 software (Bradbury et al., 2007). This filtering resulted in 258,038 polymorphic markers (27.0% of the initial loci) for further analyses. The proportion of heterogeneity (the number of markers that were not homozygous due to mixture of two homozygous genotypes or heterozygosity) and missing data after imputation were computed for each line. Identity-by-state-based genetic distance was calculated between each pair of lines using TASSEL and used for neighbor-joining clustering analysis implemented in molecular evolutionary genetics analysis (MEGA) version 6 (Tamura et al., 2013). Principal coordinate analysis (PCoA) was performed on the genetic distance matrix using DARwin version 5 (Perrier and Jacquemoud-Collet, 2006), and the first two principal components were plotted for visual examination of the clustering pattern of the lines.

## RESULTS

The number of polymorphic markers per chromosome varied from 18,086 on chromosome 10 to 41,139 on chromosome 1, with a mean of 25,804 (Fig. 1). The proportion of missing data per line across the 258,038 polymorphic markers varied from 0.4 to 12.8% and the average was 3.5% (Table 1).

Heterogeneity per parent varied from 0.1 to 25.5% and the average was 3.3% (Table 1). Approximately 87.3%

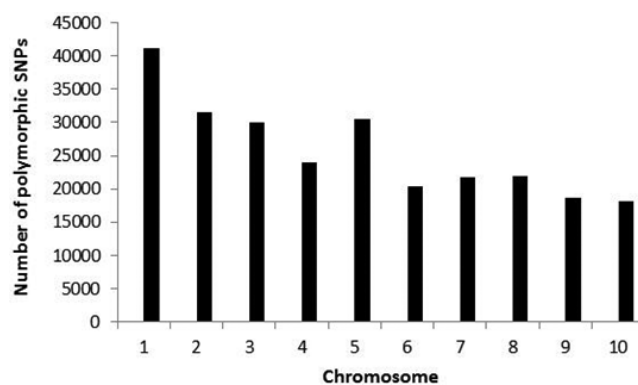


Fig. 1. The distribution of single-nucleotide polymorphism (SNP) markers per chromosome.

of the parents (48 out of 55 entries) had heterogeneity <5%. The remaining 12.7% of the entries that included one Zimbabwe maize (ZM)  $F_2$  population used in one of CIMMYT Kenya's top-cross hybrids (ZEWBc1F2), four prerelease CZLs (CZL00003, CZL03002, CZL04001, and CZL00001), and two released CMLs (CML539b and CML181) had heterogeneity ranging from 7.6 to 25.5%. For the  $F_2$  populations including ZEWBc1F2, high heterogeneity is expected, as they are constituted from at least 10 parents. Relatively high heterogeneity recorded among four CZL and two CML lines could be due to the bulk genotyping, a result of either residual heterozygosity or a combination of two homozygous SNPs, and can be reduced by selfing and selection to purify the lines.

Genetic distance between pairwise comparisons of the 55 parents ranged from 0.004 to 0.4005 (Table 2), with an overall average of 0.294. For 97.7% of the pairs of parents, genetic distance fell within 0.2001 and 0.3000. Only 0.5% of 1485 pairwise comparisons had genetic distances <0.1000. The six pairs of lines that showed <0.1000 genetic distance were CM144 vs. CML159, CZL04021 vs. CZL057, CKL08001 vs. CKL08006, CML539c vs. CZL539, CKL05017 vs. CKL05018, and CKL05017 vs. CKL05022. All the pairs with very low genetic distance (<0.1000) are of sister lines except for CML539 and CML539c. In this case, CML539 and CML539c are the

Table 2. Distribution of pairwise genetic distance calculated for 53 inbred lines and two  $F_2$  populations

Pairwise genetic distance	Frequency
	%
0.000–0.1000	0.5
0.1001–0.2000	1.7
0.2001–0.3000	63.1
0.3001–0.4000	34.6
0.4001–0.5000	0.1
Total	100.0
Summary	
Minimum distance	0.004
Maximum distance	0.4005
Mean distance	0.2944

same line but sourced from different breeders or breeding programs, that is from Zambia National and Kenya, respectively. It would be expected that CML539b, sourced from the Zimbabwe program where CML539 was developed, would have the same pairwise genetic distance as CML539 and CML539c of <0.1000. CML539b had higher pairwise genetic distances of 0.236 and 0.241 from CML539 and CML539c, respectively, mainly because it was from an old stock that was still segregating.

The neighbor-joining tree generated from the distance matrix grouped the 53 inbred lines and two F<sub>2</sub> populations into two main clusters (I and II, Fig. 2) with two subclusters each (A and B).

Lines in subcluster IA further subclustered to IAi and IAii. Cluster IA had lines from both CIMMYT's Eastern (Kenya) and Southern (Zimbabwe) Africa maize breeding program, whereas IB had the two populations ZEWAc1F2 and ZEWBc1F2 (prefixes stand for Zimbabwe Early White

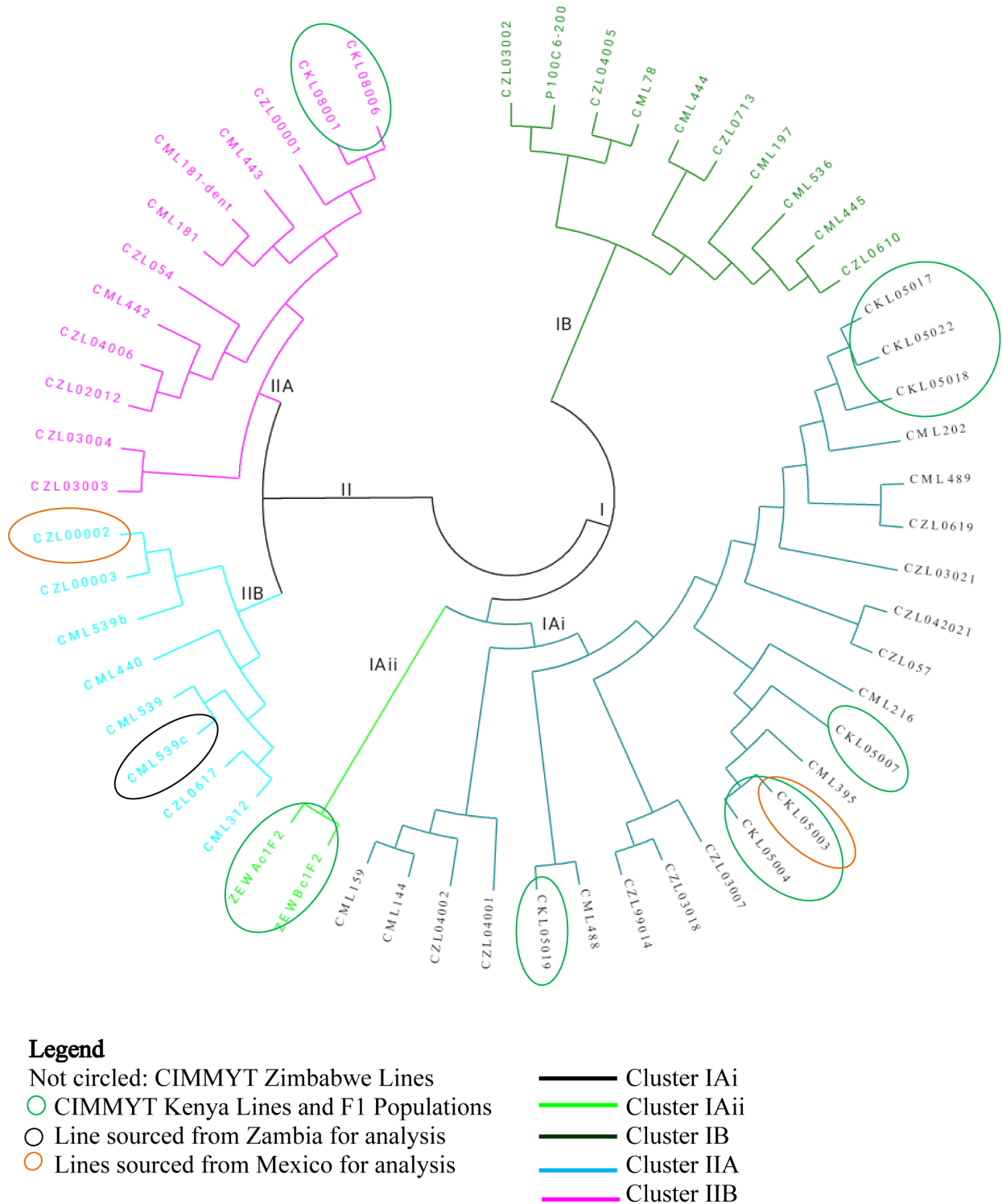


Fig. 2. Nearest-neighbor joining tree for the 53 inbred lines and two F<sub>2</sub> populations.

A and B Populations). There were three sister lines from the Kenya program, CKL05017, CKL05018, and CKL05022, selected from a cross of CML387 to CML 390 that clustered together in IAi. CKL05019 shared the same parent and belonged to the same cluster IAi. CML489, CZL03021, CZL04021, CKL05007, CKL05003, and CKL05004, all in cluster IAi, shared the common parent CML202. CML144 and CML159 were extracted from the same populations. CZL0619, CZL03012, CZL99014, and CZL03018 were selections of crosses from different lines. In cluster IAii, CZL03002 and P100C6-200 were sister lines and clustered together. One of the parents in CZL0610 is a sister line to CML445. For CML536 as well, one of the parents is a sister line to CML445, and a single-cross parent has CML197. In cluster IB, ZEWAc1F2 and ZEWBc1F2 were F<sub>2</sub> populations used to form top-cross hybrids.

In Cluster IIA, CZL00002 was a sister line to CZL00003. The rest of the lines in the cluster are CML312 and those developed with CML312 as one of the parents (CZL0617 and CML539). CML539, CML539b, and CML539c are the same line from different sources—Zambia, Kenya, and Zimbabwe, respectively. The reason for testing the different stocks sourced from different breeders was because they appeared in different hybrids presented with different names, as CZL03014 in Kenyan hybrids, CML312-SR from the Zambian breeding program, and SYN312SR from the Zimbabwean program hybrids. CML539c was sourced from an old stock that was still segregating, hence the greater distance between the source and the other two CML539 sources, Zimbabwe and CZL03014 from Kenya. CML440 clustered close to CML312, and the CML312-derived linML312 and CML440 were extracted from populations.

In cluster IIB, CKL08001 and CKL08006 lines from Kenya clustered with a sister line CZL00001. All three lines had an INTA background or parents. The two lines from South Africa, CML181 and CML181, dent clustered together. The third subcluster had lines developed from ZM populations, that is, ZM621, ZM607, and ZM523 developed and released by CIMMYT Zimbabwe.

The first three principal components (PCs) from principal coordinate analysis (PCoA) explained 25.2% of the total SNP variation among samples. A plot of PC1 (12.3%) and PC2 (6.8%) formed three major groups (Fig. 3). In the PCoA, lines separated mainly based on parental material or the pedigree. Two of the groups were made up of pedigree lines developed through pedigree selections after crosses of different lines, whereas the other cluster was made up of OPVs and lines developed from OPVs.

Lines from Cluster IAi in the nearest neighbor cluster analysis formed the first cluster (top left) in the PCoA. The lines in the first PCoA cluster further subclustered according to origin and pedigree. Lines from Kenya in the first PCoA cluster formed two groups clustered according to

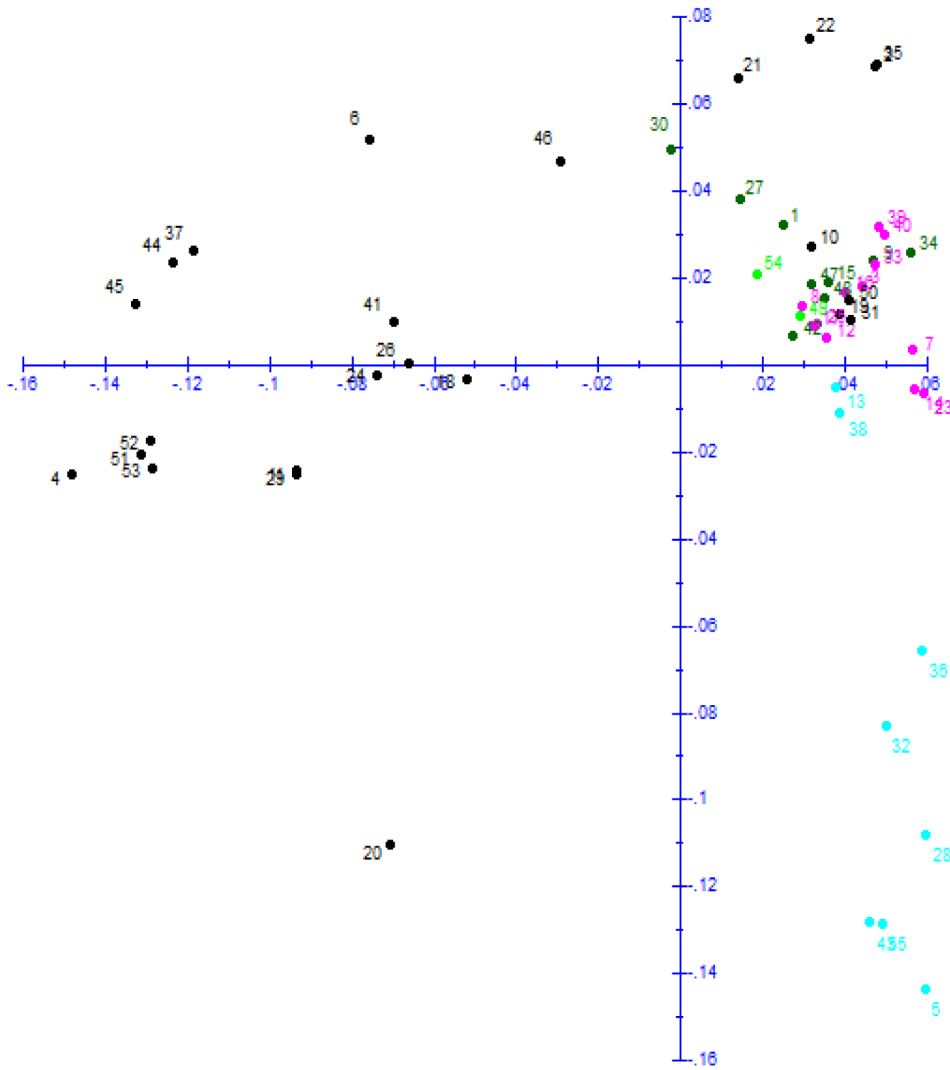
pedigree. The third subcluster had lines from Zimbabwe. The second cluster (top right) in the PCoA was made up of the lines from Clusters IAii, IB, and IIB in the nearest neighbor cluster analysis. Most of the lines clustered together despite their origin, but lines from Kenya were still close to each other.

CZL04001, CZL04002, CML144, and CML159 formed a subcluster in the PCoA. Even though CZL00002 and CZL00003 appeared in Cluster IIA in the nearest-neighbor cluster analysis, they were in Cluster II of the PCoA, subclustering along with lines CZL02012 and CZL04006 from Cluster IIB. CZL00002 and CZL00003 were extracted from the drought-tolerant populations that are coded as DTPs that were developed and screened specifically for drought tolerance, whereas CZL02012 and CZL04006 were extracted from the ZM populations. The third PCoA cluster (bottom right) had CML312 and lines developed from CML312 (CML539, CML539b, CML539c, and CZL0617) that clustered due to pedigree, as well as CML440, all from Cluster IIA in the nearest-neighbor analysis. A few parents—CML395, CKL05019, CZL03021, and CZL0713—were lying outside the clusters but were aligned towards clusters with other parents that belonged to the same cluster as them in the nearest-neighbor analysis. Principal coordinate analysis more clearly separated the lines compared with the nearest-neighbor cluster analysis, but both methods separated lines based on their pedigrees and origin.

From an analysis of the frequency of use of parents in the 52 CIMMYT ESA hybrids, the lines CML444, CML395, CML312, and CML442 were used in 30, 16, 16, and 15 hybrids respectively (i.e., in 58, 31, 31, and 29% of the 52 hybrids, Table 1), whereas most of the lines were used in one (2%) to three hybrids (6%). The four lines CML444, CML395, CMZL442, and CML312 all had distances of >0.25 for all possible pairs among them. This is because CIMMYT ESA lines are classified into three main heterotic groups (A, B, and AB; Table 1) based on combining ability with testers from the A, B, or AB group, and lines are combined across heterotic groups for maximum heterosis. There are some established A, B, and AB line and single-cross testers used for classifying lines, as well as hybrid development. The single-cross CML444 and CML395 were used as a B tester, since both lines are in heterotic Group B. Lines with good combining ability with the CML444 and CML395 single cross are classified into heterotic Group A. The CML442 and CML312 single cross was commonly used as the A tester. CML442 and CML444, as well as CML312 and CML444 single crosses, were used as AB testers. Hence, these four lines dominated in the best-performing 52 CIMMYT ESA hybrids released from 2000 to 2010.

The frequently used lines, however, belonged to different clusters both in the nearest-neighbor analysis and

Fig. 3. Principal coordinate analysis for the 53 inbred lines and the two F<sub>2</sub> populations.



**Legend**

- Cluster IAi
- Cluster IAii
- Cluster IB
- Cluster IIA
- Cluster IIB

PCoA, with CML395 in cluster IAi, CML444 in IAii, CML312 in IIA, and CML442 in IIB. The four lines also had reasonable genetic distances ranging from 0.266 to 0.282 (Table 3).

**Table 3. Pairwise genetic distances for the four commonly used lines.**

	CML444	CML395	CML442	CML312
CML444				
CML395	0.271			
CML442	0.266	0.282		
CML312	0.272	0.270	0.276	

**DISCUSSION**

The 53 parental lines and two F<sub>2</sub> populations assessed in this study were all from the CIMMYT ESA breeding program. The stations have a mid-altitude tropical environment. The Southern Africa (Zimbabwe) station experiences a unimodal rainfall pattern (October–April), whereas the Eastern Africa (Kenya) station experiences a binomial rainfall pattern and lies close to the equator. All lines used in this program are improved for yield and adaptation to the mid-altitude environment. Unweighted pair group method with arithmetic mean (UPGMA) analysis separated lines into two clusters, each with two subclusters. Separation was mostly based on pedigrees and origin,

with lines developed using an OPV or from OPVs sub-clustering on their own. Some lines, including CML312, CML445 and its sister lines, CML202, CML216, CML442, and INTA, were used in line development as parents and influenced clustering with lines subclustering according to lines they were developed from (pedigree).

Pairwise genetic distances were all  $<0.4500$  for all the lines evaluated. Average diversity or genetic distance recorded was 0.2944, which was relatively low compared with findings reported by Legesse et al. (2007) and Li et al. (2004), but comparable with findings by Beyene et al. (2013). From an analysis of genetic distances and relationship among 703 doubled haploid lines using GBS, Beyene et al. (2013) reported distances ranging from 0.070 to 0.475 with an average of 0.355. Less than 5% of the distances were  $<0.100$ , whereas 69% were 0.300 to 0.475. Legesse et al. (2007) reported an average diversity of 59% (0.590) in an evaluation of lines from CIMMYT Zimbabwe and Ethiopia using simple sequence repeat (SSR) markers. In an evaluation of popcorn lines in China using SSR markers, genetic distances ranged from 0.125 to 0.730, averaging 0.477 (Li et al., 2004). Studies reported higher-diversity evaluated lines from different geographical regions with different climates, as in the study by Legesse et al. (2007) that evaluated highland maize materials from Ethiopia and mid-altitude materials from Zimbabwe, as well as Li et al. (2004) that analyzed materials from China and Italy. Different genotyping methods were used in all of these studies that could have, among other factors, contributed to the differences among the different studies. For instance, SSR markers do not cover the whole genome like GBS and may show higher distances, unlike GBS, which covers the whole genome and will pick all similarities and differences that may be missed when SSR analysis is used. Breeding objectives like the target quality and use of the grain may have contributed to these differences as well.

This study analyzed parental lines and populations of hybrids developed in a specific environment for adaptation to specific biotic and abiotic conditions and to meet specified grain quality preferences. The focus on quality was mainly on kernel color and texture, that is, flint to semi-flint or semi-dent for mainly flour and other products like samp (dehusked maize grain served as part of the main course of a meal or as a snack). Biotic factors commonly selected for included MSV, gray leaf spot [*Botryosphaeria zeae* (G.L. Stout) Arx & E. Müller], and *Exserohilum turcicum* (Pass.) K.J. Leonard & E.G. Suggs disease resistance, while abiotic factors included low N and drought stress tolerance. Selection pressure for defined traits can result in the narrowing down of the genetic base of a breeding program. This may not be immediately noted since, according to Duvick (1990), maize is so diverse that narrowing down of germplasm has not been noted to seriously affect performance or breeding efficiency. Challenges of a narrow

germplasm base occur when there are new disease or pest outbreaks, like more recently the development of maize lethal necrosis (MLN) disease in East Africa (CIMMYT, 2012a, 2012b). If most hybrids share common parents, they may succumb to such outbreaks.

Even though the pairwise genetic distance among the lines was high with 97.8% between 0.2001 and 0.4000, a few of the studied lines (CML444, CML395, CML312, and CML442) were used in most of the hybrids, giving the hybrids a narrow genetic base. This was mainly due to stability and adaptation of the lines or single-cross testers to the mid-altitude environments, as well as their good agronomic and stress tolerance traits that were opted for by breeders for use in hybrid formation over other traits. The lines and testers were also older than the new lines like CZL0617, CZL0713, and CKL05003 that were released later in 2006, 2007, and 2005, respectively. Hence, despite preference, the old lines were used in both old and new hybrids evaluated for genetic gain, released from 2000 to 2010. The slow turnover of female lines probably also resulted in the lines being used in a relatively high number of hybrids before new lines got into the system or pipeline.

The use of a few lines in most of the hybrids in a breeding program or region poses a risk of reducing diversity in the genepool of the parental lines, population, and hybrids that are cultivated in a region, exposing the varieties to common natural disasters such as disease outbreaks. The diversity of parental lines, populations, and hybrids therefore might have been restricted by the pedigree breeding method and preferential use of material adapted to the mid-altitude tropical environments, posing risk in cases of disease or pest outbreaks. The frequently used lines, however, belonged to different heterotic groups as well as different clusters, both in the nearest-neighbor analysis and PCoA, and had reasonable genetic distances.

Despite all four lines belonging to different clusters, the genetic base of the released hybrids is narrow, rendering susceptibility to disease or pest outbreaks that may arise in the region. The genetic base of the best 52 hybrids in this study released in the CIMMYT ESA program from 2000 to 2010, some of which have been disseminated to farmers throughout ESA through collaborators, could be narrow in terms of line use in hybrid formation and genetic distance among the lines used. There is need to ensure that, during selection of the best-performing varieties, the genetic base is not narrowed down. It is important that these commonly used lines be improved whenever they show weaknesses, like in the case of MLN outbreak in Eastern Africa in 2012 (CIMMYT, 2012a, 2012b), since they are already adapted to the geographical region. Improved versions of these lines like CML539, an improved MSV-resistant version of CML312, can be used in addition to the older varieties. Such improved lines add



diversity to the genetic base. Combining the commonly used old and adapted lines in single-cross testers with new lines from different clusters will provide a wider range of testers and a wider genetic base for hybrids that are developed and released.

## CONCLUSIONS

Diversity was low. Pairwise genetic distances were  $<0.4500$  with an average of 0.2944. Four parents (CML312, CML395, CML442, and CML444) were each used in 15 to 30 hybrids (29–58%) out of the 52 hybrids. The rest (51) of the parents were each used in one to three (2–6%) of the 52 hybrids. This evaluation focused on a selected subset of all the CIMMYT ESA hybrids, as it evaluated the best-performing genotypes from each year. The best-performing hybrids are usually passed onto the farmers through collaborators. The sample of parental lines analyzed does not represent the entire germplasm of CIMMYT ESA program, and therefore these findings should not be generalized for the whole program. However, because these best-performing hybrids are disseminated to farmers in the region, implications of disease or pest outbreaks can be devastating if the hybrids have common susceptible parents, like the outbreak of MLN in Eastern Africa in 2012. If the same lines are commonly used or a few testers are used in hybrid formation, most or all of the hybrids will succumb to problem outbreaks. Therefore, it is better for breeders to create a wide range of single-cross testers used in hybrid formation and consciously incorporate a wide range of parental lines. The results give an indication of the danger of eventually narrowing down the germplasm base and the need to maintain a large or wide genetic base for the whole program.

## Conflict of Interest

The authors declare that there is no conflict of interest.

## Acknowledgments

This work was supported in part by the Drought Tolerance Maize for Africa (DTMA) and Improved Maize for African Soils (IMAS) projects funded by the Bill and Melinda Gates Foundation and USAID, and the MAIZE CGIAR research program. We thank the technical assistants at each site (CIMMYT and NARS) for conducting the trials.

## References

- Bänziger, M., and J. Auriol. 2007. Recent advances in breeding for drought and salinity stress tolerance. In: M.A. Jenks, P.M. Hasegawa, and S.M. Jain, editors, *Advances in molecular breeding toward drought and salt tolerant crops*. Springer, Dordrecht, the Netherlands. p. 587–601.
- Beyene, Y., S. Mugo, K. Semagn, G. Asea, W. Trevisan, A. Tarekegne et al. 2013. Genetic distance among doubled haploid maize lines and their testcross performance under drought stress and non-stress conditions. *Euphytica* 192:379–392. doi:10.1007/s10681-013-0867-5
- 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. doi:10.1093/bioinformatics/btm308
- Cairns, J.E., K. Sonder, P.H. Zaidi, N. Verhulst, G. Mahuku, R. Babu et al. 2012. Maize production in a changing climate. Impacts, adaptation, and mitigation strategies. *Adv. Agron.* 114:1–58. doi:10.1016/B978-0-12-394275-3.00006-7
- CIMMYT. 2005. Maize inbred lines released by CIMMYT. CIMMYT, El Batán, Texcoco, Mexico.
- CIMMYT. 2012a. CGIAR research programme maize 2012 annual report. CIMMYT, Mexico, D.F.
- CIMMYT. 2012b. Maize lethal necrosis (MLN) disease in Kenya and Tanzania: Facts and actions. CIMMYT, Mexico, D.F.
- CIMMYT. 2015. DT maize: A quarterly bulletin of the Drought Tolerant Maize for Africa Project. CIMMYT, Mexico, D.F.
- Duvick, D.N. 1990. The romance of plant breeding and other myths. In: J.P. Gustafson, editor, *Gene manipulation in plant improvement*. Plenum Press, New York. p. 39–54. doi:10.1007/978-1-4684-7047-5\_3
- Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi:10.1371/journal.pone.0019379
- FAO. 2013. FAOSTAT gateway. FAO, Rome. <http://faostat3.fao.org/faostat-gateway/go/to/browse/FB/FB/E> (accessed 4 Sept. 2013).
- Fisher, M., T. Abate, R.W. Lunduka, W. Asnake, Y. Alemayehu, and R.B. Madulu. 2015. Drought tolerant maize for farmer adaptation to drought in sub-Saharan Africa: Determinants of adoption in eastern and southern Africa. *Clim. Change* 133:283–299. doi:10.1007/s10584-015-1459-2
- Legesse, B.W., A.A. Myburg, K.V. Pixley, and A.M. Botha. 2007. Genetic diversity of African maize inbred lines revealed by SSR markers. *Hereditas* 144:10–17. doi:10.1111/j.2006.0018-0661.01921.x
- Li, Y., D. Lv, Y. Wang, S. Chen, and J. Tang. 2004. Study on the genetic diversity of popcorn inbreds and their germplasm relationship with normal corn inbreds using SSR markers. *Maydica* 49:327–333.
- Masuka, B., G.N. Atlin, M. Olsen, C. Magorokosho, M. Labuschagne, J. Crossa et al. 2017a. Gains in maize genetic improvement in Eastern and Southern Africa: I. CIMMYT hybrid breeding pipeline. *Crop Sci.* 57:168–179. doi:10.2135/cropsci2016.05.0343
- Masuka, B., C. Magorokosho, M. Olsen, G.N. Atlin, M. Bänziger, K.V. Pixley et al. 2017b. Gains in maize genetic improvement in Eastern and Southern Africa: II. CIMMYT open-pollinated variety breeding pipeline. *Crop Sci.* 57:180–191. doi:10.2135/cropsci2016.05.0408
- M'mboyi, F., S. Mugo, M. Mwimali, and L. Ambani. 2010. Maize production and improvement in Sub-Saharan Africa. African Biotechnol. Stakeholders Forum (ABSF), Nairobi, Kenya. [http://www.absfafrica.org/downloads/maize\\_book\\_series.pdf](http://www.absfafrica.org/downloads/maize_book_series.pdf) (accessed 5 Sept. 2013).
- Perrier, X., and J.P. Jacquemoud-Collet. 2006. Darwin software. Centre de coopération internationale en recherche agronomique pour le développement. <http://darwin.cirad.fr/> (accessed 2 Aug. 2017).

- Pingali, P.L., and S. Pandey. 2001. Part 1. Meeting world maize needs: Technological opportunities and priorities for public sector. In: P.L. Pingali, editor, CIMMYT 1999–2000 world maize facts and trends. CIMMYT, Mexico, D.F.
- Poland, J.A., and T.W. Rife. 2012. Genotyping-by-sequencing for plant breeding and genetics. *The Plant Genome* 5:92–102. doi:10.3835/plantgenome2012.05.0005
- Ray, D.K., N.D. Mueller, P.C. West, and J.A. Foley. 2013. Yield trends are insufficient to double global crop production by 2050. *PLoS ONE* 8:e66428. doi:10.1371/journal.pone.0066428
- Romay, M., M. Millard, J. Glaubitz, J. Peiffer, K. Swarts, T. Casstevens et al. 2013. Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol.* 14:R55. doi:10.1186/gb-2013-14-6-r55
- Rovere, R.L., G. Kostandini, A. Tahirou, J. Dixon, W. Mwangi, Z. Guo, and M. Banziger. 2010. Potential impact of investments in drought tolerant maize in Africa. CIMMYT, Addis Ababa, Ethiopia.
- Semagn, K. 2014. Leaf tissue sampling and DNA extraction protocols. In: P. Besse, editor, *Molecular plant taxonomy: Methods and protocols*. Human Press, New York. p. 53–67. doi:10.1007/978-1-62703-767-9\_3
- Smale, M., D. Byerlee, and T. Jayne. 2011. *Maize revolutions in Sub-Saharan Africa*. Policy Res. Working Paper 5659. World Bank. Springer Netherlands, Dordrecht.
- Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30:2725–2729. doi:10.1093/molbev/mst197
- Tefera, T., S. Mugo, Y. Beyene, H. Karaya, and R. Tende. 2011. Grain yield, stem borer and disease resistance of new maize hybrids in Kenya. *Afr. J. Biotechnol.* 10:4777–4783.