Genetic Gains in Grain Yield Through Genomic Selection in Eight Bi-parental Maize Populations under Drought Stress

Yoseph Beyene,* Kassa Semagn, Stephen Mugo, Amsal Tarekegne, Raman Babu, Barbara Meisel, Pierre Sehabiague, Dan Makumbi, Cosmos Magorokosho, Sylvester Oikeh, John Gakunga, Mateo Vargas, Michael Olsen, Boddupalli M. Prasanna, Marianne Banziger, and Jose Crossa

ABSTRACT

Genomic selection incorporates all the available marker information into a model to predict genetic values of breeding progenies for selection. The objective of this study was to estimate genetic gains in grain yield from genomic selection (GS) in eight bi-parental maize populations under managed drought stress environments. In each population, 148 to 300 $F_{2:3}$ (C_0) progenies were derived and crossed to a single-cross tester from a complementary heterotic group. The resulting testcrosses of each population were evaluated under two to four managed drought stress and three to four well-watered conditions in different locations and genotyped with 191 to 286 single nucleotide polymorphism (SNP) markers. The top 10% families were selected from C_0 using a phenotypic selection index and were intermated to form C_1. Selections both at C_1 and C_2 were based on genomic estimated breeding values (GEBVs). The best lines from C_0 were also advanced using a pedigree selection scheme. For genetic gain studies, a total of 55 entries representing the eight populations were crossed to a single-cross tester, and evaluated in four managed drought stress environments. Each population was represented by bulk seed containing equal amounts of seed of C_0, C_1, C_2, C_3, parents, F_1s, and lines developed via pedigree selection. Five commercial checks were included for comparison. The average gain from genomic selection per cycle across eight populations was 0.086 Mg ha^{-1}. The average grain yield of C_3–derived hybrids was significantly higher than that of hybrids derived from C_0. Hybrids derived from C_3 produced 7.3% (0.176 Mg ha^{-1}) higher grain yield than those developed through the conventional pedigree breeding method. The study demonstrated that genomic selection is more effective than pedigree-based conventional phenotypic selection for increasing genetic gains in grain yield under drought stress in tropical maize.

Y. Beyene, K. Semagn, S. Mugo, D. Makumbi, J. Gakunga, M. Olsen, and B.M. Prasanna, International Maize and Wheat Improvement Center (CIMMYT), P. O. Box 1041, Village Market 00621, Nairobi, Kenya; A. Tarekegne and C. Magorokosho, CIMMYT, 12.5 Km peg Mazowe Road, P.O. Box MP163, Mount Pleasant, Harare, Zimbabwe; R. Babu, CIMMYT-India, c/o ICRISAT, Patancheru 502324, Andhra Pradesh, India; B. Meisel, Monsanto South Africa (Pty) Ltd, P.O. Box 69933, Bryanston, 2021, South Africa; P. Sehabiague, Monsanto SAS, Croix de Pardies, BP21, 40305 Peyrehorade, France; S. Oikeh, African Agricultural Technology Foundation (AATF), P.O. Box 30709-00100, Nairobi, Kenya, and M. Vargas, M. Banziger, and J. Crossa, International Maize and Wheat Improvement Center (CIMMYT), Apdo. Postal 6-641, 06600, Mexico DF, Mexico. Received 1 July 2014. *Corresponding author: (y.beyene@cgiar.org).

Abbreviations: ASI, anthesis-silking interval; DTMA, Drought Tolerant Maize for Africa project; GEBVs, genomic estimated breeding values; GS, genomic selection; MABC, marker-assisted backcrosssiing; MARS, marker-assisted recurrent selection; QTL, quantitative trait loci; SSA, sub-Saharan Africa; WEMA, Water Efficient Maize for Africa project.

M AIZE is a staple food in many countries of sub-Saharan Africa (SSA) and is commonly grown by millions of resource-poor smallholder farmers. In SSA, maize covers more than 25 million hectares that produce 38 million metric tons of grain (Shiferaw et al., 2011). The average maize yield in SSA is 1.8 t per hectare (Smale et al., 2011), which is very low compared to that of other maize–growing regions in the developing world. Several factors, including high incidence of abiotic and biotic stresses, high irrigation costs, and inability of farmers to access and purchase good

quality seeds and fertilizers, contribute to the low maize productivity. Development and delivery of improved maize cultivars that can withstand drought stress without significant yield penalty under optimal rainfall conditions (Campos et al., 2004; Ribaut and Ragot 2007; Sambatti and Caylor 2007) are critical for attaining food security in the region. To develop drought-tolerant maize, selection can be performed directly under drought stress, indirectly under well-watered conditions, or simultaneously under both optimal and drought stress conditions (Byrne et al., 1995). However, conventional breeding through pheno-typing for drought tolerance requires multiple locations and years of field testing. Also, direct selection for grain yield under drought is often more difficult due to the low heritability of grain yield under stress (Edmeades et al., 1999; Venuprasad et al., 2007; Ziyomo and Bernardo 2013). Indirect selection for secondary traits that are easy to measure, highly heritable, and highly correlated with grain yield under drought, such as anthesis-silking interval (ASI), leaf senescence, leaf chlorophyll content and several other morpho-physiological traits (Bolanos and Edmeades 1996; Ribaut et al., 1996), were suggested and successfully implemented in CIMMYT's maize breeding program, leading to the development of elite drought-tolerant tropical maize germplasm (Bänziger et al., 2006).

Using marker-assisted backcrossing (MABC), Ribaut and Ragot (2007) introgressed five quantitative trait loci (QTL) associated with yield components and flowering time in maize from a drought-tolerant donor into a drought susceptible recurrent parent (CML287) which resulted in increased grain yield and reduced ASI under water-limited conditions. More recently, using tropical germplasm, a number of meta-QTL regions were identified for grain yield and ASI (Almeida et al., 2013; Semagn et al., 2013) as well as secondary traits associated with stress tolerance (Almeida et al., 2014) under managed drought stress conditions. However, grain yield under drought is a complex trait influenced by genetic background and other environmental factors. Thus, relying on a few QTL from a specific donor line from MABC is unlikely to work in diverse recipient backgrounds to generate drought-tolerant maize germplasm. The reasons for this includes (i) individual QTL associated with drought tolerance often explain a very small proportion of the phenotypic variance for grain yield and ASI, (ii) QTLs for drought-related traits are often genetic-background specific and may not have significant effects in different genetic backgrounds, and (iii) many QTLs associated with grain yield and ASI are detected under either drought stress or well-watered conditions but not both (Semagn et al., 2013). Marker-assisted recurrent selection (MARS) is a marker-based breeding method that seeks to accumulate favorable alleles from several genomic regions within a single population (Edwards and Johnson 1994). In contrast to MABC that targets major effect QTLs that have been mapped and validated across different genetic backgrounds, MARS aims at accumulating a large number of QTLs in a given population using a subset of markers that are significantly associated with target traits (Bernardo 2008).

Genomic selection (Meuwissen et al., 2001) is another marker-based strategy that incorporates all the available marker information simultaneously into a model to predict the genetic value of progenies for selection (Lorenz 2013). Each marker is considered a putative QTL, reducing the risk of missing small-effect QTLs (Guo et al., 2012). A computer simulation study showed that using all markers gave better prediction accuracy of breeding values than using subsets of markers significantly associated with QTLs (Bernardo and Yu 2007) which was recently verified using empirical data (Massman et al., 2013). Much of the research on GS in crops over the past 5 yr has focused on developing and testing different statistical prediction models that could be used on real data for predicting diverse breeding panels of different crops for different traits and in different environments (de los Campos et al., 2009; Crossa et al., 2010). Many studies have focused on genotyping individuals of different types of breeding populations (i.e., bi-parental, synthetic, doubled haploid lines) that were developed in previous years and evaluated in different environments (Crossa et al., 2013). Most studies have used different cross-validation designs and schemes that attempt to predict performance of untested individuals and environments (VanRaden, 2008; Hayes et al., 2009; Crossa et al., 2010, 2013, 2014; Pérez-Rodríguez et al., 2012; Poland et al., 2012; Dawson et al., 2013; Heslot et al., 2013; Jarquín et al., 2014). De los Campos et al. (2009) and Crossa et al. (2010) examined several statistical models for genomic selection in diverse panels of maize and wheat germplasm from the International Maize and Wheat Improvement Center (CIMMYT) using a random cross-validation scheme that mimics the prediction of unobserved phenotypes based on markers and pedigrees. Massman et al. (2013) published results from a breeding experiment to demonstrate the genetic gains achieved through GS in a bi-parental maize population derived from a cross between B73 and Mo17. This study involved genotyping 233 recombinant inbred lines (RILs) with 284 markers, evaluating the testcrosses under well-watered conditions and advancing the population using GS and MARS and reported superior response to GS for stover yield, as well as stover and grain yield indices by 14 to 50% over MARS. To our knowledge, there are no publications that reported the use of GS to improve tropical maize populations for grain yield under drought stress conditions.

As part of the Drought Tolerant Maize for Africa (DTMA) and Water Efficient Maize for Africa (WEMA) projects, CIMMYT has developed more than 34

Table 1. Summary of the eight bi-parental C$_0$ populations used in the present study.

Population code	Project[†]	Initial cross	Managed drought evaluation sites[‡]	Well-watered evaluation sites[‡]	Population size	Number of SNPs[§] used for genotyping C$_0$	Heritability for grain yield under managed drought sites	Heritability for grain yield under well-watered sites
6x1008	WEMA	CML540/ CML505	Chisumaban, Isinya, Kiboko and Nanga	Embu, Kakamega, Kiboko and Mtwapa	165	201	0.27	0.40
6x1016	WEMA	CML540/ CZL99017	Isinya, Kibokooko and Nanga	Embu, Kakamega, Kiboko and Mtwapa	148	191	0.13	0.36
6x1017	WEMA	CML540/ CML539	Isinya, Kiboko and Nanga	Embu, Kakamega, Kiboko and Mtwapa	184	210	0.19	0.41
6x1020	WEMA	CZL0723/ CZL0724	Kiboko and Nanga	Embu, Kakamega, Kiboko and Mtwapa	181	218	0.26	0.33
6x1028	WEMA	CZL074/ VL062645	Chisumabans and Kiboko	Embu, Kakamega, Kiboko and Mtwapa	174	205	0.10	0.51
JMpop1	DTMA	CML440/ CML504	Tlatizapa, Kiboko, Chiredzi	Tlatizapa, Harare	300	197	0.41	0.39
JMpop2	DTMA	CML444/ CML441	Tlatizapa, Kiboko, Chiredzi	Tlatizapa, Harare	298	286	0.30	0.46
JMpop3	DTMA	CML444/ Malawi	Tlatizapa, Kiboko, Chiredzi	Tlatizapa, Harare	236	197	0.10	0.31
Minimum					148	191	0.10	0.31
Maximum					300	286	0.41	0.51
Mean					211	213	0.22	0.40

[†] WEMA, Water Efficient Maize for Africa; DTMA, Drought Tolerant Maize for Africa.

[‡] Nanga and Tlatizapa are in Zambia and Meixo; Chisumaban, Chiredzi and Harare are in Zimbabwe. All other sites are in Kenya.

[§] SNP, single nucleotide polymorphism.

bi-parental populations since 2009. Testcrosses of progenies derived from these populations were evaluated in two to four managed drought and three to four well-watered environments, followed by advancing the populations using either MARS or GS approaches. Eight of the 34 bi-parental populations were improved using GS that were used for this study. The objective of this study was to estimate genetic gains in grain yield in eight bi–parental maize populations evaluated in four managed drought–stress environments.

MATERIALS AND METHODS
Population Development, Phenotyping of F$_{2:3}$ Testcrosses and Genotyping

A summary of the eight breeding populations used in this study is provided in Table 1. The breeding scheme used for developing the materials is illustrated in Fig. 1. The eight breeding populations used in this study (Table 1) were among the first set of 34 bi-parental populations developed from the DTMA and WEMA projects. Initial phenotypic evaluations were performed on testcrosses derived by crossing 148 to 300 F$_{2:3}$ (cycle 0, abbreviated as C$_0$) with a single-cross tester from the complementary heterotic group. Testcrosses of each population, along with five selected commercial checks (WH504, WH505, H513, CZH0616, and DK8033), were planted using an α lattice design with two replications per location. The testcrosses were phenotyped in two to four managed drought stress and three to four well-watered locations as described by Semagn et al. (2013). The managed drought stress trials were conducted during the dry (rain-free) season by withdrawing irrigation starting from 2 wk before flowering through harvest, whereas the well-watered trials were conducted during the rainy season with supplemental irrigation as needed. Data were analyzed within the managed drought and well-watered trials using a linear mixed model that considered both genotypes and trials as random effects. The five WEMA C$_0$ populations were genotyped with 191 to 218 SNPs by the Monsanto Company using a TaqMan assay (www.appliedbiosystems.com), while the three DTMA C$_0$ populations were genotyped with 197 to 286 SNPs using a KASP assay (Semagn et al., 2014) at LGC genomics (www.lgcgenomics.com/genotyping/kasp-genotyping-chemistry).

Selection of C$_0$ to Form C$_1$

Testcrosses of each C$_0$ population were evaluated for 12 to 17 different traits commonly associated with drought tolerance, including grain yield at 12.5% moisture content, anthesis date, number of ears per plant, and leaf senescence. Only grain yield and ASI under drought and grain yield under well-watered conditions were used to develop a multi-trait phenotypic selection index. Selection at C$_0$ was based on Eigen Selection Index Method (Cerón-Rojas et al., 2008) computed from phenotypic data (Bernardo and Yu, 2007). The C$_0$ families were ranked according to the value of the phenotypic selection index and the 10% of the entries with the highest selection index values were selected. The selected families were planted ear-to-row and intermated to form C$_1$. Within each population, selected C$_0$ families were separated into two equal groups, and bulk pollen from the first half of the population was used to pollinate all plants of the other half. The best 18 to 25 ears were selected, individually shelled, and 100 seeds were retained from each ear. Equal amounts of seed from selected ears were bulked to form the subsequent cycle for planting.
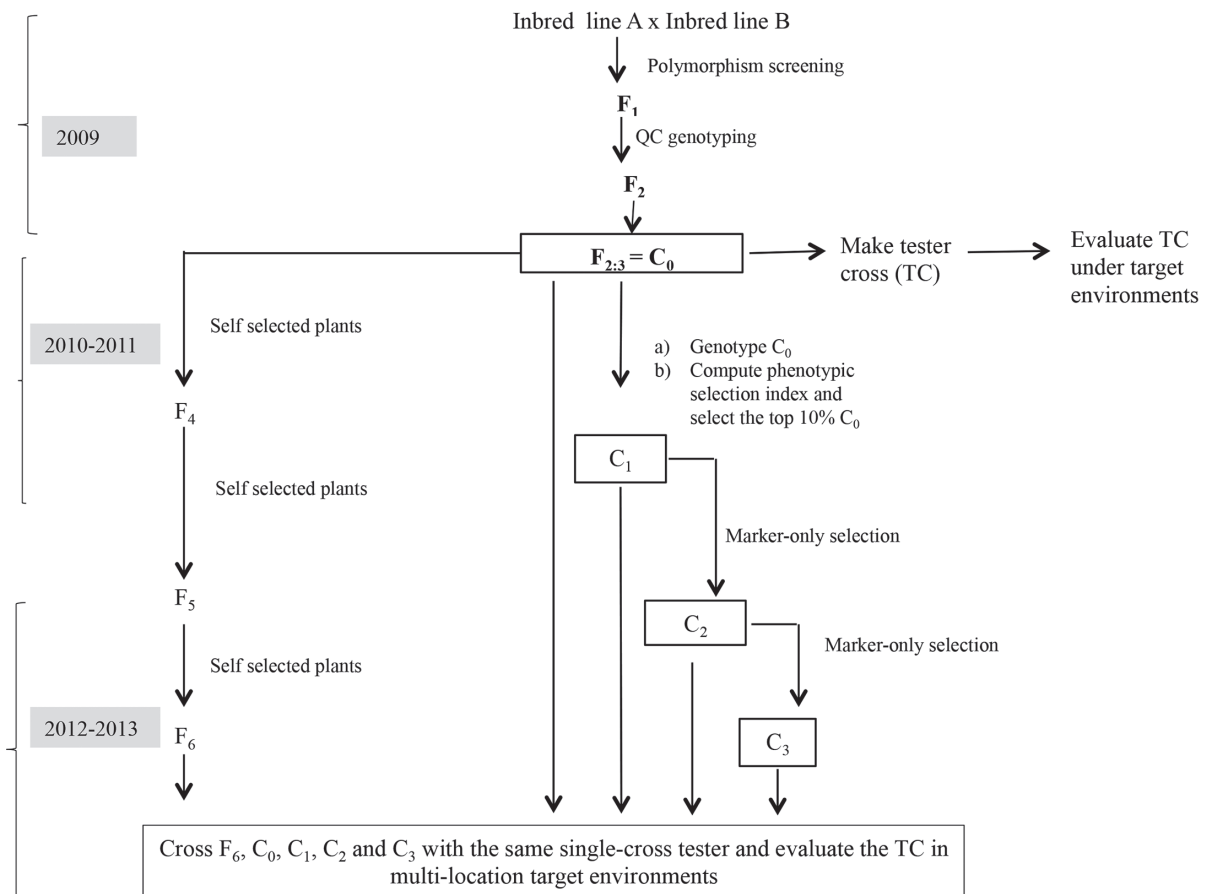
Figure 1. Schematic illustration of the various steps followed in the present study.

Marker-based Selection in C_1 and C_2

For each of the eight bi-parental populations, a balanced bulk of 18 to 25 ears, each represented by 17 to 23 seeds, was planted. Two weeks after planting, a total of 276 plants per population were labeled; leaf samples were collected from every plant and transported to the Biosciences for eastern and central Africa (BecA) laboratory in Nairobi, Kenya (Semagn 2014). DNA was extracted from leaf discs and shipped to the Monsanto Company (five WEMA populations) in the United States or to LGC Genomics (three DTMA populations) in the United Kingdom for SNP genotyping. Genotyping of C_1 plants was done using the full complement of SNPs originally used for genotyping the C_0 populations (Table 1).

The predictive model was the Genomic Best Linear Unbiased Predictor (GBLUP) (Hayes et al., 2009) with a genomic relationship matrix computed according to VanRaden (2008) and Habier et al. (2007). A brief description of the model used for prediction is given in the Appendix. Genomic estimated breeding values were calculated for all C_1 individuals and the top 10% of the C_1 individuals were selected and intermated to form C_2 as described above. The C_1 recombination protocol was repeated in C_2 to generate C_3. All recombination experiments were conducted under well-watered environments at the maize experimental station in Kiboko, Kenya.

Development of Lines via Pedigree Selection

To compare the effectiveness of GS with that of conventional pedigree selection, the top 10% of the selected C_0 families for GS were also subjected to inbreeding under well-watered environments with visual selection to develop $F_{5:6}$ lines (Fig. 1). The selected lines of each population were planted at Kibos, Kenya. Phenotypic selections were made within and among families based on per se visual evaluation (germination and good stand establishment, plant type, low ear placement, and well-filled ears) and reaction to naturally-occurring major leaf diseases (gray leaf spot caused by *Cercospora zeae maydis*; leaf blight caused by *Exerohilium turcicum*; common rust caused by *Puccinia sorghi*; and maize streak virus caused by maize streak *geminivirus*), and selected plants were selfed to form $F_{3:4}$ lines. Selected $F_{3:4}$ plants were planted at Kiboko, Kenya, at a high plant population density (80,000 plants/ha), and plants with less root and stalk lodging and low ear placement were selfed to form $F_{4:5}$ lines. This procedure was repeated to form $F_{5:6}$ lines. The best five $F_{5:6}$ lines were selected from each population, top-crossed to a single-cross tester and included in the genetic gain study.

Phenotypic Evaluation for Assessing Genetic Gain

A total of 55 entries from the eight populations plus five commercial checks widely used in sub-Saharan Africa (CZH0616, H513, WH505, DK8053, and Pioneer 3253) were used for the genetic gain studies. The DTMA populations (JMpop1,

Table 2. Mean grain yield (GY), anthesis date (AD), and plant height (PH) of testcrosses at cycle 0 (C_0), cycle 1 (C_1), cycle 2 (C_2), cycle 3 (C_3), F_1, pedigree-selected lines, and commercial checks for the eight bi-parental populations evaluated in four managed drought-stress environments. The average gain per selection cycle with and without C_0 is provided. For each population, the highest value is indicated in bold faces.

Entries	Across eight populations			Three DTMA populations[†]			Five WEMA populations[†]		
	GY	AD	PH	GY	AD	PH	GY	AD	PH
	Mg ha⁻¹	days	cm	Mg ha⁻¹	days	cm	Mg ha⁻¹	days	cm
C_0	2.286	63.910	179.900	2.410	63.670	189.160	2.212	64.060	174.300
C_1	2.420	64.080	181.900	2.319	63.940	188.320	2.482	64.160	178.100
C_2	2.438	**64.410**	**184.600**	2.378	**64.830**	**193.340**	2.474	64.160	**179.400**
C_3	**2.593**	64.100	182.200	**2.613**	64.720	193.220	**2.581**	63.730	175.600
Pedigree	2.417	64.400	181.700	–	–	–	2.417	**64.400**	181.700
F_1	2.394	63.930	175.600	–	–	–	2.394	63.930	175.600
Parents	2.361	64.000	178.400	2.186	64.130	183.030	2.431	63.950	176.600
Checks	2.180	64.310	180.100	2.191	63.930	181.030	2.169	64.700	179.300
LSD 0.05	0.219	0.700	5.260	0.342	1.130	6.900	0.257	0.850	6.320
Average gain per cycle (C_0, C_1, C_2, C_3)	0.094	0.090	0.960	0.068	0.404	1.720	0.110	–0.105	0.520
Average gain per cycle (C_1, C_2, C_3)	0.086	0.010	0.150	0.151	0.390	2.450	0.050	–0.215	–1.250

† DTMA, Drought Tolerant Maize for Africa; WEMA, Water Efficient Maize for Africa.

JMpop2, and JMpop3) were represented by 17 entries consisting of bulks of seed from C_0, C_1, C_2, and C_3, plus five parental lines with CML444 being a common parent in the two populations. The WEMA populations were represented by 38 entries consisting of bulks of seed from C_0, C_1, C_2, and C_3 of five populations (6x1008, 6x1016, 6x1017, 6x1020, and 6x1028), eight parental lines with CML540 being a common parent for three populations, five F_1s and five $F_{5:6}$ lines developed through pedigree selection scheme. For each population, C_0 was represented by a single entry consisting of a bulk of an equal amounts of remnant C_0 ($F_{2:3}$) seed from all families per population; C_1, C_2, and C_3 were represented by a single entry that consisted of an equal bulk of ears of selected families harvested at each cycle, while the conventional pedigree method was represented by five $F_{5:6}$ lines advanced though pedigree selection. Each entry was testcrossed to the same single-cross tester. The testcrosses along with the five commercial checks were evaluated in managed drought stress experiments at four locations (Kiboko, Mtwapa, Mbeere-1, and Mbeere-2) in Kenya.

The experimental design was α lattice with three replications per location. The managed drought stress trials were conducted during the dry season by withdrawing irrigation starting from 2 wk before flowering through harvest. At all locations, entries were planted in two-row plots, 5 m in length, with 0.75 m spacing between rows and 0.25 m spacing between hills. Two seeds per hill were initially planted and thinned to one plant per hill 3 wk after seedling emergence to obtain a final population density of 53,333 plants per hectare. Fertilizers were applied as recommended for the area (60 kg N and 60 kg P_2O_5 per ha) with N fertilizer application at planting and 6 wk after emergence. The fields were kept free of weeds by hand weeding. Although several agronomic traits and major diseases were recorded, only grain yield corrected to 12.5% moisture content, anthesis date recorded when 50% of plants within a plot had shed pollen, and plant height between the base of the plant and the first tassel branch were included here. For each trait, standard

α-lattice analyses of phenotypic traits were performed across all populations, across the DTMA and WEMA populations, and for each population using SAS (SAS Institute, 2009).

RESULTS AND DISCUSSION
Genetic Gain Across Eight Bi-parental Populations

C_3–derived hybrids produced significantly higher grain yield than the hybrids derived from C_0, parents, and commercial checks (Table 2). Gains in grain yield were 5.9% from C_0 to C_1, 0.7% from C_1 to C_2, 6.4% from C_2 to C_3, and 13.4% from C_0 to C_3. Although the gain in grain yield at the different selection cycles was not consistent, our results showed relatively higher gains in grain yield both at C_1 and C_3. The higher gain observed in C_1 over C_2 is consistent with the findings of previous empirical and simulation studies (Bernardo and Yu, 2007; Massman et al., 2013). Massman et al. (2013) showed that most of the selection gains in maize grain yield and stover quality were obtained in C_1. Simulation studies have also shown that gains after multiple cycles of GS were slightly larger if selection in C_0 was based on phenotypic values instead of a combination of phenotype plus a marker index (Bernardo and Yu, 2007).

Hybrids derived from C_3 showed 7.3% higher gain in grain yield than those developed using conventional pedigree breeding methods, but the difference was not statistically significant. However, direct comparison between GS and pedigree selection method may be confounded by differences in population size and level of inbreeding between the two methods. Hybrids involving C_3–produced significantly higher grain yield than testcrosses of the respective initial parents used for developing the C_0 populations (9.8%) and the commercial checks (18.9%).

The commercial checks were among the best hybrids available to farmers in the region and such high genetic gain over the mean of the commercial checks was indeed highly remarkable. Moreover, there was no significant difference among hybrids for anthesis date and plant height, except hybrids involving F_1s (Table 2), clearly demonstrating the advantages of GS for increasing grain yield without affecting maturity and plant and ear height.

Altogether, the three cycles of selection increased grain yield under managed drought by 13.4% of which 7.5% of the gain was due to genomic selection at C_1 and C_2 and the remaining 5.9% due to phenotypic selection at C_0. The overall average genetic gain across the three cycles of selection was 0.094 Mg ha^{-1} per cycle (Table 2). A total of 10 seasons over 4 yr were needed from development of F_1s to harvesting of C_3, with two seasons per year needed to go from F_1 to C_0 testcross evaluation and three seasons per year needed for rapid cycling recombination (Fig. 1). The average 0.094 Mg ha^{-1} gain in grain yield per selection cycle is therefore equivalent to 0.070 Mg ha^{-1} yr^{-1} (i.e., three cycles × 0.094 Mg ha^{-1} per cycle divided by 4 yr).

A review of genetic gain studies from conventional pedigree selection conducted both in temperate and tropical maize germplasm reported highly variable results (Edmeades 2013). In SSA, preliminary estimates of yield gains from conventional selection revealed 0.039 to 0.080 Mg ha^{-1} yr^{-1} under optimal conditions, but only 0.018 Mg ha^{-1} yr^{-1} under drought stress (Edmeades 2013). A recent study using 67 hybrids developed at CIMMYT and released between 2000 and 2011 showed genetic gains of 0.032 and 0.109 kg ha^{-1} yr^{-1} for grain yield under managed drought and well-watered conditions, respectively (B. Masuka, personal communication, 2014). Therefore, the average gain observed under drought in our study using GS was two- to fourfolds higher than what has been reported from conventional phenotypic selection under drought stress in SSA. In the the United States, yield gains during the past 70 yr have been positive and linear, varying from 0.065 to 0.075 Mg ha^{-1} yr^{-1} (Duvick 2005), which is comparable to the overall genetic gain in our studies, and reinforces the usefulness of genomic selection for achieving high genetic gains in grain yield under drought stress conditions. GS also provided significant time savings over conventional breeding methods because up to three cycles of markers based selection were completed per year (Fig. 1) compared to the one breeding cycle per year for the conventional pedigree selection (for phenotyping, selection, and recombination). GS could, therefore, be a method of choice for developing improved germplasm because (i) a large number of genome-wide molecular markers are available at a cost that is comparable to or lower than the cost of phenotyping; (ii) there is difficulty in establishing low-cost, high-throughput, and accurate phenotyping under drought stress to generate reliable phenotypic data in multi-location experiments, and (iii)

identification of few markers that are significantly associated with the trait of interest under drought stress often misses a substantial portion of the genetic variance contributed by loci having small effects (Poland et al., 2012).

Genetic Gain Across the Three DTMA and Five WEMA Bi-parental Populations

For the DTMA populations, the response to selection for GY from C_1 to C_2 and from C_2 to C_3 was 2.5 and 9.9%, respectively (Table 2). Hybrids developed from C_3 produced 8.4% more GY than hybrids developed from C_0. They also produced 19.5% higher grain yields than hybrids developed from the parents and 19.3% more yield than the commercial checks. The regression analysis showed that the average yield gain per selection cycle across the three DTMA populations ranged from 0.068 to 0.151 Mg ha^{-1}. For WEMA populations, the response to selection for grain yield showed a 12.2% increase from C_0 to C_1 and a 4.3% increase from C_2 to C_3, but was about zero from C_1 to C_2 (Table 2). The increase in grain yield from C_0 to C_3 was 16.7% and from C_1 to C_3 was 4.0%. Hybrids developed from C_3 produced higher grain yields than hybrids developed using pedigree breeding method (6.8%), and the hybrids involving parents (6.2%), and the commercial checks (19.0%). The regression analysis showed an average yield gain per selection cycle across the five WEMA populations ranging from 0.050 to 0.110 Mg ha^{-1}. In the DTMA and WEMA populations, anthesis date and plant height did not show statistically significant differences, except that plant height of hybrids derived from C_3 was significantly higher than those hybrids developed from the parents used in the initial crosses. Therefore, the increase in grain yield at the different selection cycles was not due to an increase in days to maturity and plant height.

The DTMA and WEMA populations showed similar trends with little or no response to genomic selection from C_1 to C_2, although gain was observed from C_2 to C_3 (Fig. 2). However, there is also some level of discrepancies in the response to selection from C_0 to C_1. The DTMA populations showed a 3.8% reduction in grain yield between C_0 and C_1, while the WEMA populations showed the highest gain in grain yield (12.2%) between C_0 and C_1. A number of factors, including differences in the C_0 population size and number of markers used in selection (Table 1) might have contributed for such differences.

Genetic Gains for Individual Bi-parental Populations

Genetic gains in GY for each individual bi-parental population are shown in Table 3 and Fig. 3. We compared grain yield at C_3 with the base population at C_0 to determine whether genomic selection had increased GY in each population. Our results indicated that most of the populations had higher grain yield at C_3 than at C_0. Hybrids formed
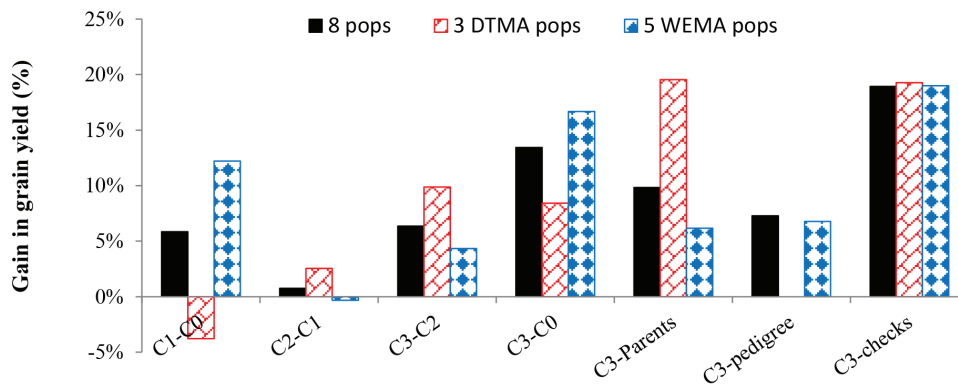
Figure 2. Comparison of the overall genetic gain in grain yield across the eight populations, the three Drought Tolerant Maize for Africa (DTMA) and five Water Efficient Maize for Africa (WEMA) populations evaluated in four managed drought-stress environments.

from C_3 of populations JMPop1, JMPop3, 6x1008, and 6x1020 showed the highest gains, ranging from 11.5 to 20.8% in comparison with their corresponding hybrids formed from C_0. The response to selection of population JMPop2 was unique in that the best cycle was C_0; however, C_3 produced 6.7% more GY than C_1 and 2.3% more than C_2, indicating an increase in genetic gains from GS. Population JMPop3 also showed an initial decrease in GY from C_0 to C_1 but then a consistent increase from C_1 to C_2 (12.5%) and from C_2 to C_3 (13.5%). Among the five WEMA populations, four populations (6x1008, 6x1016, 6x1020, and 6x1028) had 15.3 to 29.1% higher grain yields for C_3–derived hybrids than their corresponding C_0–derived hybrids, while population 6x1017 showing a 0.4% reduction in grain yield from C_3 to C_0. We also compared grain yield across three selection cycles (C_1, C_2, and C_3) to determine whether there is a consistent trend across all populations (Fig. 3). While most populations showed the highest grain yield at C_3, populations 6x1028 and 6x1017 showed the highest gain at C_1 and C_2, respectively (Fig. 3), clearly demonstrating that GS tended to increase grain yield in the majority of the populations but not always.

CONCLUSIONS

Our study involved a large dataset obtained from the evaluation of eight bi-parental populations under managed drought stress environments. The results demonstrated the efficiency of GS in maize, with an average gain per cycle of 0.086 Mg ha^{-1} under drought stress without significant changes in maturity and plant height. Our results also showed differential response to GS in different populations, with the majority of the populations showing a consistent increase across the different selection cycles. Only one population showed a reduction in grain yield after two cycles of marker-based selection. Higher gains in grain yield under drought stress were achieved using genomic selection compared to conventional pedigree selection. Furthermore, the overall gain in average grain yield using GS was two- to fourfolds higher than the previously reported gain in average GY under drought stress

in SSA using conventional phenotypic selection. Moreover, GS offered the advantage of significant time-savings over conventional breeding methods, since up to three cycles of GS could be conducted within a year. Results of this study would be useful to maize breeders planning to utilize GS for improving stress resilience of maize.

APPENDIX
Genomic Best Linear Unbiased Predictor (GBLUP)

The standard linear genetic model considers that the phenotypic response of the *ith* individual (γ_i) is explained by a genetic factor specific to that individual (g_i), and a residual comprising all other non-genetic factors (ε_i), among others, the environmental effects (temporal or spatial) and the effects described by the experimental design. Then, the linear genetic model for n genotypes ($i = 1,...,n$) is represented as $\gamma_i = g_i + \varepsilon_i$. In this standard linear genetic model, the genetic factor g_i can be described by using a summation of molecular marker effects or by using pedigree.

Meuwissen et al. (2001) were the first to propose doing an explicit regression of phenotypes on the marker genotypes using the simple parametric regression model $g_i = \sum_{j=1}^{p} x_{ij}\beta_j$, such that $\gamma_i = \sum_{j=1}^{p} x_{ij}\beta_j = \varepsilon_i$ ($j = 1,2,...,p$), where β_j is the regression of γ_i on the *jth* marker covariate ($j = 1,2,...,p$), and x_{ij} is the number of copies of bi-allelic markers (0,1,2 or −1,0,1). In matrix notation, this can be represented as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad [A1]$$

In Eq. [A1] genetic values are $g_i = \sum_{j=1}^{p} x_{ij}\beta_j$ and this gives the transformation

$$\gamma_i = g_i + \varepsilon_i$$

with $\mathbf{g} \sim N(0, K\sigma_\beta^2)$, $\varepsilon \sim N(0, \mathbf{D}\sigma_e^2)$ $\left(\mathbf{D} = Diag\left\{\frac{1}{n_i}\right\}\right.$, ($n_i$ is the number

Table 3. Mean grain yield, anthesis date, and plant height of testcrosses at cycle 0 (C_0), cycle 1 (C_1), cycle 2 (C_2), cycle 3 (C_3), F_1, pedigree-selected lines, and commercial checks for each bi-parental population evaluated in four managed drought stress environments. The average gain per selection cycle with and without C_0 is provided. For each population, the highest value is indicated in bold faces.

Entries	JMPop1	JMPop2	JMPop3	6x1008	6x1016	6x1017	6x1020	6x1028
Grain yield, Mg ha^{-1}								
C_0	2.171	**2.517**	2.540	2.283	2.261	2.532	2.141	1.844
C_1	2.592	2.237	2.129	2.697	2.583	2.384	2.317	**2.429**
C_2	2.185	2.333	2.562	2.471	2.573	**2.965**	2.263	2.098
C_3	**2.622**	2.387	**2.883**	**2.754**	2.918	2.375	**2.734**	2.127
F_1	–	–	–	2.28	**2.957**	2.120	2.436	2.177
Parents	2.133	2.248	2.181	2.635	2.464	2.446	2.4	2.214
Pedigree	–	–	–	2.47	2.423	2.456	2.327	2.413
LSD 0.05	0.832	0.831	0.536	0.631	0.516	0.392	0.498	0.430
Average gain per cycle (C_0, C_1, C_2, C_3)	0.095	−0.040	0.146	0.119	0.192	0.011	0.173	0.052
Average gain per cycle (C_1, C_2, C_3)	0.015	0.048	0.377	0.029	0.168	−0.005	0.208	−0.150
Anthesis days, days								
C_0	62.000	63.830	65.170	63.170	63.830	63.670	62.330	67.330
C_1	62.200	64.500	64.830	63.500	64.330	63.670	62.830	66.500
C_2	**64.830**	64.330	65.330	63.170	**64.500**	62.330	63.500	**67.330**
C_3	63.670	**64.670**	65.830	62.170	63.000	63.670	**63.670**	66.170
F1	–	–	–	63.000	63.500	64.170	62.670	66.330
Parents	62.000	66.330	64.080	62.330	63.920	63.420	63.170	66.920
Pedigree	–	–	–	**63.600**	64.330	**63.700**	63.530	66.830
LSD 0.05	2.820	1.840	1.465	1.230	1.480	1.700	1.040	1.280
Average gain per cycle (C_0, C_1, C_2, C_3)	0.764	0.235	0.248	−0.333	−0.232	−0.134	0.469	−0.265
Average gain per cycle (C_1, C_2, C_3)	0.735	0.085	0.500	−0.665	−0.665	0.000	0.420	−165.000
Plant height, cm								
C_0	184.070	190.920	192.480	**186.230**	184.700	179.450	154.830	166.580
C_1	186.150	**193.100**	185.720	174.300	185.550	184.750	**170.330**	175.930
C_2	**192.250**	189.630	198.150	174.570	**197.650**	182.780	164.930	176.700
C_3	188.350	192.300	**199.020**	176.380	189.600	180.550	166.170	165.580
F_1	–	–	–	169.780	193.830	171.450	166.400	**176.870**
Parents	172.000	193.800	183.160	175.290	184.060	181.210	167.630	174.950
Pedigree	–	–	–	185.470	196.400	**186.170**	168.630	172.050
LSD 0.05	20.370	11.730	17.180	11.730	8.680	11.300	10.490	10.950
Average gain per cycle (C_0, C_1, C_2, C_3)	1.894	0.067	2.480	−0.293	2.680	0.133	2.862	−0.223
Average gain per cycle (C_1, C_2, C_3)	1.100	−0.400	6.650	1.040	2.025	−2.100	−2.080	−5.175

of observations collected from the ith individual) where $\mathbf{K} = \mathbf{XX'}$. Habier et al. (2007), Meuwissen et al. (2001), and VanRaden (2008) related σ_g^2 and σ_β^2 by defining $\mathbf{G} = \mathbf{K}/C = \mathbf{XX'}/C$ for $C = \sum_j 2p_j(1-p_j)$ (p_j is the allelic frequency of the jth marker), from where $\sigma_g^2 = \sigma_\beta^2 C$. Thus, $\hat{\sigma}_g^2$ is an estimate of $\sigma_g^2 = \left[\sum_{j=1}^p x_{ij}\beta_j\right] = \sum_j 2p_j(1-p_j)\sigma_\beta^2$. This model is named GBLUP, the $\mathbf{G}$ matrix is the genomic derived relationship matrix, and the genetic predicted values are $\hat{\mathbf{g}} = \left[\lambda\mathbf{G}^{-1} + \mathbf{D}^{-1}\right]^{-1}\mathbf{D}^{-1}\mathbf{y}$ (Van Raden 2008) where the ridge parameter λ is the ratio between the residual and the genetic variance.
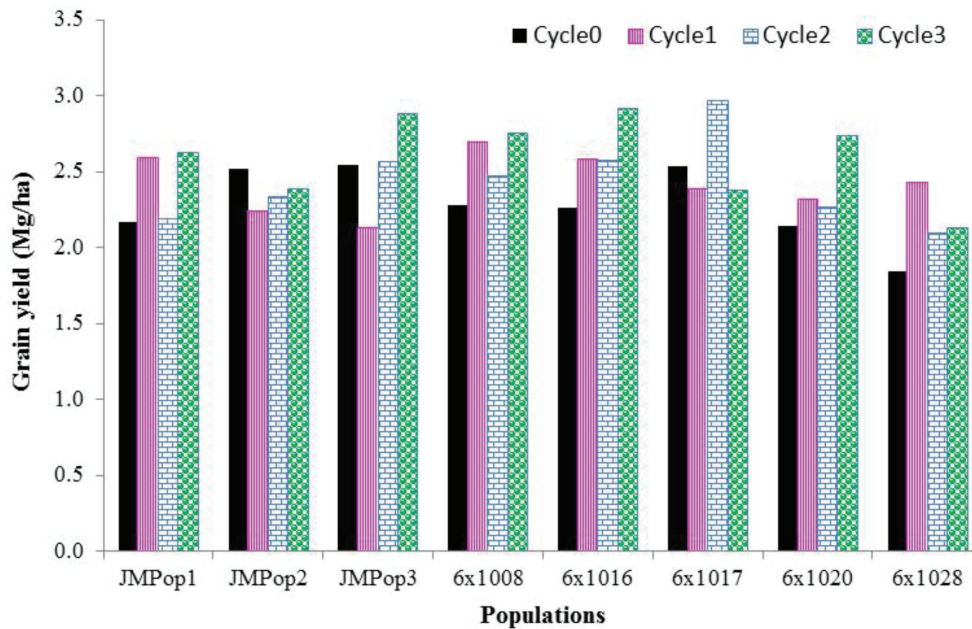
Figure 3. Comparison of mean grain yield for each population evaluated in four managed drought-stressed environments.

References

Almeida, G., D. Makumbi, C. Magorokosho, S. Nair, A. Borem, J.M. Ribaut et al. 2013. QTL mapping in three tropical maize populations reveals a set of constitutive and adaptive genomic regions for drought tolerance. Theor. Appl. Genet. 126(3):583–600. doi:10.1007/s00122-012-2003-7

Almeida, G., S. Nair, A. Borem, J. Cairns, S. Trachsel, J.M. Ribaut et al. 2014. Molecular mapping across three populations reveals a QTL hotspot region on chromosome 3 for secondary traits associated with drought tolerance in tropical maize. Molec. Breeding. In press. doi: 10.1007/s11032-014-0068-5.

Bänziger, M., P.S. Setimela, D. Hodson, and B. Vivek. 2006. Breeding for improved drought tolerance in maize adapted to southern Africa. Agric. Water Manage. 80:212–224. doi:10.1016/j.agwat.2005.07.014

Bernardo, R. 2008. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. Crop Sci. 48:1649–1664. doi:10.2135/cropsci2008.03.0131

Bernardo, R., and J.M. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. Crop Sci. 47:1082–1090. doi:10.2135/cropsci2006.11.0690

Bolanos, J., and G.O. Edmeades. 1996. The importance of the anthesis-silking interval in breeding for drought tolerance in tropical maize. Field Crops Res. 48:65–80. doi:10.1016/0378-4290(96)00036-6

Byrne, P.F., J. Bolanos, G.O. Edmeades, and D.L. Eaton. 1995. Gains from selection under drought versus multilocation testing in related tropical maize populations. Crop Sci. 35:63–69. doi:10.2135/cropsci1995.0011183X003500010011x

Campos, H., M. Cooper, J.E. Habben, G.O. Edmeades, and J.R. Schussler. 2004. Improving drought tolerance in maize: A view from industry. Field Crops Res. 90:19–34. doi:10.1016/j.fcr.2004.07.003

Cerón-Rojas, J.J., J. Sahagún-Castellanos, F. Castillo-González, A. Santacruz-Varela, I. Benítez-Riquelme, and J. Crossa. 2008. A molecular selection index method based on eigen analysis. Genetics 180:547–557. doi:10.1534/genetics.108.087387

Crossa, J, Y. Beyene, S. Kassa, P. Pérez, J.M. Hickey, C. Chen et al. 2013. Genomic Prediction in maize breeding populations with genotyping-by-sequencing. G3: Genes|Genomes|Genetics. 3:1903–1926.

Crossa, J., G.L. Campos, P. Pérez, D. Gianola, J. Burgueño, J.L. Araus et al. 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186:713–724. doi:10.1534/genetics.110.118521

Crossa, J., P. Pérez, J. Hickey, J. Burgueño, L. Ornella, J. Cerón-Rojas et al. 2014. Genomic prediction in CIMMYT maize and wheat breeding programs open. Heredity 112:48–60. doi:10.1038/hdy.2013.16

Dawson, J.C., J.B. Endelman, N. Heslot, J. Crossa, J. Poland, S. Dreisigacker et al. 2013. The use of unbalanced historical data for genomic selection in an international wheat breeding program. Field Crops Res. 154:12–22. doi:10.1016/j.fcr.2013.07.020

de los Campos, G., H. Naya, D. Gianola, J. Crossa, A. Legarra, E. Manfredi, K. Weigel, and M. Cotes. 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics 182:375–385. doi:10.1534/genetics.109.101501

Duvick, D.N. 2005. Genetic progress in yield of United States maize (Zea mays L.). Maydica 50:193–202.

Edmeades, G.O. 2013. Progress in achieving and delivering drought tolerance in maize-An update. International Service for the Acquisition of Agri-Biotech Applications, Ithaca, NY.

Edmeades, G.O., J. Bolaños, S.C. Chapman, H.R. Lafitte, and M. Bänziger. 1999. Selection improves drought tolerance in tropical maize populations: I. Gains in biomass, grain yield, and harvest index. Crop Sci. 39:1306–1315. doi:10.2135/cropsci1999.3951306x

Edwards, M., and L. Johnson. 1994. RFLPs for rapid recurrent selection. Proceedings Joint Plant Breeding Symposia Series. Am. Soc. for Hort. Sci., Corvallis, OR; CSSA, Madison, WI. p. 33–40.

Guo, Z.G., D.M. Tucker, J.W. Lu, V. Kishore, and G. Gay. 2012. Evaluation of genome-wide selection efficiency in maize nested association mapping populations. Theoret. Appl. Genet. 124:261–275. doi:10.1007/s00122-011-1702-9

Habier, D., R.L. Fernando, and J.C.M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389–2397.

Hayes, B.J., P.J. Bowman, A.J. Chamberlain, and M.E. Goddard. 2009. Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. 92:433–443. doi:10.3168/jds.2008-1646

Heslot, N., J.L. Jannink, and M.E. Sorrells. 2013. Using genomic prediction to characterize environments and optimize prediction accuracy in applied breeding data. Crop Sci. 53:921–933. doi:10.2135/cropsci2012.07.0420

Jarquín, D., J. Crossa, X. Lacaze, P. Cheyron, J. Daucourt, J. Lorgeou et al. 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor. Appl. Genet. 127:595–607. doi:10.1007/s00122-013-2243-1

Lorenz, A.J. 2013. Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: A simulation experiment. G3: Genes|Genomes|Genetics. 3:481–491.

Massman, J.M., H.-J.G. Jung, and R. Bernardo. 2013. Genome-wide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. Crop Sci. 53:58–66. doi:10.2135/cropsci2012.02.0112

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Pérez-Rodríguez, P., D. Gianola, J.M. González-Camacho, J. Crossa, Y. Manès, and S. Dreisigacker. 2012. Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. G3: Genes|Genomes|Genetics. 2:1595–1605.

Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S.Y. Wu, Y. Manes et al. 2012. Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Gen. 5:103–113. doi:10.3835/plantgenome2012.06.0006

Ribaut, J.M., D.A. Hoisington, J.A. Deutsch, C. Jiang, and D. González de León. 1996. Identification of quantitative trait loci under drought conditions in tropical maize. 1. Flowering parameters and the anthesis-silking interval. Theor. Appl. Genet. 92:905–914. doi:10.1007/BF00221905

Ribaut, J.M., and M. Ragot. 2007. Marker-assisted selection to improve drought adaptation in maize: The backcross approach, perspectives, limitations, and alternatives. J. Exp. Bot. 58:351–360. doi:10.1093/jxb/erl214

Sambatti, J.B.M., and K.K. Caylor. 2007. When is breeding for drought tolerance optimal if drought is random? New Phytol. 175:70–80. doi:10.1111/j.1469-8137.2007.02067.x

SAS Institute. 2009. The SAS System for Windows. SAS Inst., Cary, NC.

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.

Semagn, K., R. Babu, S. Hearne, and M. Olsen. 2014. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement. Mol. Breed. 33:1–14. doi:10.1007/s11032-013-9917-x

Semagn, K., Y. Beyene, M. Warburton, A. Tarekegne, S. Mugo, B. Meisel et al. 2013. Meta-analyses of QTL for grain yield and anthesis silking interval in 18 maize populations evaluated under water-stressed and well-watered environments. BMC Genomics 14:313. doi:10.1186/1471-2164-14-313

Shiferaw, B., B.M. Prasanna, J. Hellin, and M. Bänziger. 2011. Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. Food Security 3:307–327. doi:10.1007/s12571-011-0140-5

Smale, M., D. Byerlee, and T. Jayne. 2011. Maize revolutions in Sub-Saharan Africa: Forthcoming, World Bank, Washington, DC and Tegemeo Inst., Nairobi, Kenya.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. doi:10.3168/jds.2007-0980

Venuprasad, R., H.R. Lafitte, and G.N. Atlin. 2007. Response to direct selection for grain yield under drought stress in rice. Crop Sci. 47:285–293. doi:10.2135/cropsci2006.03.0181

Ziyomo, C., and R. Bernardo. 2013. Drought tolerance in maize: Indirect selection through secondary traits versus genome-wide selection. Crop Sci. 53:1269–1275. doi:10.2135/cropsci2012.11.0651