Published July 28, 2016

RESEARCH

Genome-wide Association for Plant Height and Flowering Time across 15 Tropical Maize Populations under Managed Drought Stress and Well-Watered Conditions in Sub-Saharan Africa

Jason G. Wallace,* Xuecai Zhang, Yoseph Beyene, Kassa Semagn, Michael Olsen, Boddupalli M. Prasanna,* and Edward S. Buckler

ABSTRACT

Genotyping breeding materials is now relatively inexpensive but phenotyping costs have remained the same. One method to increase gene mapping power is to use genome-wide genetic markers to combine existing phenotype data for multiple populations into a unified analysis. We combined data from 15 biparental populations of maize (*Zea mays* L.) (>2500 individual lines) developed under the Water-Efficient Maize for Africa project to perform genome-wide association analysis. Each population was phenotyped in multilocation trials under water-stressed and well-watered environments and genotyped via genotyping-by-sequencing. We focused on flowering time and plant height and identified clear associations between known genomic regions and the traits of interest. Out of ~380,000 single-nucleotide polymorphisms (SNPs), we found 115 and 108 that were robustly associated with flowering time under well-watered and drought stress conditions, respectively, and 143 and 120 SNPs, respectively, associated with plant height. These SNPs explained 36 to 80% of the genetic variance, with higher accuracy under well-watered conditions. The same set of SNPs had phenotypic prediction accuracies equivalent to genome-wide SNPs and were significantly better than an equivalent number of random SNPs, indicating that they captured most of the genetic variation for these phenotypes. These methods could potentially aid breeding efforts for maize in Sub-Saharan Africa and elsewhere. The methods will also help in mapping drought tolerance and related traits in this germplasm. We expect that analyses combining data across multiple populations will become more common and we call for the development of algorithms and software to enable routine analyses of this nature.

J.G. Wallace, Dep. of Crop and Soil Sciences, The Univ. of Georgia, Athens, GA 30602-6810; J.G. Wallace and E.S. Buckler, Inst. for Genomic Diversity, Cornell Univ., Ithaca, NY 14853; E.S. Buckler, USDA – Agricultural Research Service, Ithaca, NY 14853; X. Zhang, International Maize and Wheat Improvement Center (CIMMYT), Apdo. Postal 6-641, 06600, Mexico, DF, Mexico; Y. Beyene, M. Olsen, and B.M. Prasanna, CIMMYT, P.O. Box 1041, Village Market 00621, Nairobi, Kenya; K. Semagn, Dep. of Agricultural, Food and Nutritional Science, Univ. of Alberta, Edmonton, Canada. Received 16 Oct. 2015. Accepted 06 June 2016. *Corresponding authors (b.m.prasanna@cgiar.org; jason.wallace@uga.edu). Assigned to Associate Editor Seth Murray.

Abbreviations: BLUP, best linear unbiased predictors; *FT*, *FLOWERING LOCUS T*; G × E, genotype × environment; GBS, genotyping-by-sequencing; GWAS, genome-wide association; H^2, broad-sense heritability; h^2, narrow-sense heritability; MITE, miniature inverted-repeat transposable element; NAM, nested association mapping; QTL, quantitative trait locus; RMIP, resample model inclusion probability; SNP, single-nucleotide polymorphism; WEMA, Water-Efficient Maize for Africa

Modern genotyping technologies have made genetic data for plant breeding relatively inexpensive; however, the generation of multilocation phenotypic data remains expensive and laborious. There is an increasing interest in reducing phenotyping costs, either by increasing the efficiency of phenotyping or by extracting maximum information from the available phenotypic data. This paper focuses on the latter by combining modern genotyping with phenotypic data from separate datasets to map important traits under well-watered and drought stress conditions in maize.

The materials for this analysis were generated by CIMMYT under the Water-Efficient Maize for Africa (WEMA) project. This project is a public-private partnership that focuses on breeding drought-tolerant and water-efficient maize for Sub-Saharan

Published in Crop Sci. 56:2365–2378 (2016).
doi: 10.2135/cropsci2015.10.0632

Africa [http://wema.aatf-africa.org/ (accessed 27 June 2016); Beyene et al. 2015; Zhang et al. 2015]. Water-Efficient Maize for Africa has generated 35 biparental maize populations from tropical-adapted germplasm for marker-assisted selection and has mapped quantitative trait loci (QTLs) for drought-related traits. Several of these populations were included in an earlier meta-analysis (Semagn et al. 2013) that combined data from low-density SNP markers to gain power across populations. Our objective in this study was to use high-density genotyping to combine the populations for the identification of QTLs with improved power and resolution.

Unlike structured populations such as nested association mapping (NAM; McMullen et al. 2009) or Multiparent Advanced Generation Inter-Cross (MAGIC; Kover et al. 2009) populations, the WEMA populations were not designed to be analyzed together. Each population comprises a separate dataset and many of the controls used to help unify NAM or MAGIC populations—planting the same fields, using common reference lines, having shared parents, etc.—are lacking. Determining the appropriate statistical framework to unify these populations was thus a major aim of this study. We focus our analysis on two traits: plant height and flowering time (days to anthesis). These traits have been extensively studied in temperate maize material, providing a basis for comparison with our results. Also, since the WEMA populations are largely tropical in origin, they may contain novel QTLs that previous studies (performed mostly with temperate material) have not detected.

In maize, flowering time is a highly polygenic trait (Buckler et al. 2009). Among landraces it can vary from 2 to 11 mo (Kuleshov 1933). Proper flowering time is a crucial component of adapting maize to local conditions, as improper flowering can cause large yield losses, especially in conjunction with drought (Bolaños & Edmeades 1996). Most maize flowering time QTLs have small individual effects (Buckler et al. 2009). Despite the large number of QTLs identified in mapping studies, relatively few have been cloned to individual genes. The best studied of these is *Vgt1*, a natural variant linked to the insertion of a miniature inverted-repeat transposable element (MITE) in an enhancer ~70 kb upstream of the gene *ZmRap2.7* (Vlăduțu et al. 1999, Salvi et al. 2007). Another is *ZmCCT*, caused by a CACTA-like transposon insertion in the promoter region that affects both flowering time and photoperiod sensitivity, especially in tropical maize (Yang et al. 2013).

Maize plant height is also a highly polygenic trait (Wang et al. 2006, Peiffer et al. 2014) that appears to closely follow the infinitesimal model, meaning it is controlled by a large number of loci of small effect. Large-effect mutations for height have been found at over 40 maize loci (Wang et al. 2006) and many of these have been traced to individual genes [MaizeGDB (Andorf et al. 2015)]. Most of these are involved in plant hormone signaling, transportation, or synthesis, especially of gibberellins and brassinosteroids (Salas Fernandez et al. 2009). Some of the better characterized proteins include Dwarf3, which is involved in gibberellin synthesis (Winkler and Helentjaris 1995); nana plant1, which affects brassinosteroid synthesis (Hartwig et al. 2011); Brachytic2, which influences polar auxin transport in the stalk (Multani et al. 2003) and Dwarf8 and Dwarf9, which cause insensitivity to gibberellins by mutations in DELLA proteins (Salas Fernandez et al. 2009).

Here, we present here the results of combining datasets across 15 of the WEMA populations to map QTLs for plant height and flowering time. We use multiple mapping approaches, including a standard joint genome-wide association (GWAS), a Bayesian analysis that treats each population independently, and a recently developed model selection algorithm, FarmCPU (Liu et al. 2016). From these, we identify a set of SNPs associated with each phenotype under either drought or well-watered conditions and we use genomic prediction models to confirm that they identify the genetic regions that are important for phenotypic control.

MATERIALS AND METHODS

Plant Materials and Phenotyping

We initially selected 19 (out of 35) WEMA populations with extensive genetic and phenotypic data. However, four populations were later dropped from the study because of issues related to genotypic data (see the Results section for details). Brief descriptions of the 15 populations have been given in Table 1. Details on population development and phenotyping were described previously (Semagn et al. 2013, Beyene et al. 2015). In brief, the populations were derived from crosses between CIMMYT drought-tolerant donors and CIMMYT inbred lines currently in commercial use in eastern and southern Africa. The F_1 generation was advanced to the F_2, BC_1, or BC_1F_2 generation (Table 1). These plants were selfed and their progeny ($F_{2:3}$ or BC_1F_2 families) were crossed with a single-cross tester from a complementary heterotic group. Test-crosses of each population, together with five commercial checks, were planted in α-lattice designs with two replications and evaluated for multiple traits under four to five well-watered environments and two to four managed drought stress environments at various field stations in Kenya, Zimbabwe and Zambia during 2010 and 2011 (Supplemental Fig. S1). The relationships among populations are complex, with most sharing at least 1 parent with another population (Supplemental Fig. S2).

The test-cross hybrids were planted in α-lattice design trials with two replications per location. At all locations, entries were planted in two-row plots (5 m long), with a spacing of 0.75 m between rows and 0.25 m between hills. Two seeds per hill were planted and thinned to one plant per hill 3 wk after emergence to achieve a final population density of 53,333 plants ha^{-1}. Fertilizers were applied at the rate of 60 kg N ha^{-1} and 60 kg P_2O_5 ha^{-1} as recommended for the area. Nitrogen

Table 1. Details of the 15 WEMA biparental maize populations used in the study.

Population code	Population type	Initial cross	Tester	Population size (after filtering)
6x1008	F_2	CML505 × CZL00009	CML395 × CML444	187
6x1015	F_2	CZL04003 × CZL00009	CML395 × CML444	187
6x1016	F_2	CZL99017 × CZL00009	CML395 × CML444	179
6x1017	F_2	CML539 × CZL00009	CML395 × CML444	169
6x1018	F_2	CML505 × CZL99017	CML395 × CML444	188
6x1019	F_2	CZL04008 × CZL0719	CML395 × CML444	186
6x1020	F_2	CZL0723 × CZL0724	CML395 × CML444	180
6x1021	F_2	CZL0723 × CZL0719	CML395 × CML444	182
6x1023	F_2	CZL0618 × VL062655	CML395 × CML444	169
6x1028	F_2	CZL074 × VL062645	CML395 × CML444	170
6X1114	BC_1–F_2	CZL00003 × CKL09001	CML312 × CML442	182
6x1115	BC_1–F_2	VL062655 × CZL00003	CML312 × CML442	175
6x1117	BC_1	CML444 × CKL09007	CML312 × CML442	113
6x1118	BC_1	CML444 × CKL09001	CML312 × CML442	141
6x1120	F_2	CML395 × CKL09008	CML312 × CML442	169

was split-applied at planting and 6 wk after emergence. Fields were kept free of weeds by hand weeding. Managed drought stress trials were conducted during the dry season by withdrawing irrigation 2 wk before flowering and withholding irrigation throughout harvest. In well-watered experiments, supplemental irrigation was provided as needed to avoid moisture stress. Commercial checks were included in each experiment, although only two checks (WH504 and WH505) were included across all experiments. These checks are commercial hybrids from Western Seed Company (Kitale, Kenya)

Anthesis date was measured as the number of days from planting to when 50% of the plants had shed pollen. Plant height was measured as the distance from the base of the plant to the height of the first tassel branch. Raw phenotype data were first converted to field-wise best linear unbiased predictors (BLUPs) to correct for within-field spatial effects using CIMMYT's FieldBook software (http://dtma.cimmyt.org/index.php/information-tools/software-download, accessed 27 June 2016). These data were then cleaned by removing outliers and other spurious data through automatic scripts and manual inspection of the data. Both the raw and cleaned data are available in Supplemental File S1; cleaning scripts are included in Supplemental File S2.

For Bayesian GWAS analysis, the within-field BLUPs were used. For joint GWAS, experiment-wide BLUP values were calculated using the lme4 package in R (Bates et al. 2015) to fit a linear model with environment (station–year combination) as a fixed effect and line and specific field as random effects. The resulting BLUPs are in Supplemental File S1 and the R script used to generate these is included in Supplemental File S2.

Genotype Generation and Imputation

We first selected 19 WEMA biparental populations for which extensive genetic and phenotypic information was available. For the lines in each family, equal amounts of leaf tissue from 15 $F_{2:3}$ plants were bulked and extracted with a DNeasy 96 Plant kit (Qiagen Inc., Valencia, CA). Purified DNA was sent to the Genomic Diversity Facility at Cornell University for genotyping-by-sequencing (GBS; Elshire et al. 2011). Raw flowcell output was processed to genotype calls using the TASSEL-GBS pipeline (Glaubitz et al. 2014). The GBS version 2.7 "tags on physical map" file downloaded from Panzea (www.panzea.org, accessed 27 June 2016) was used to anchor reads to the Maize AGPv2 genome for SNP calling at 955,690 total sites.

Each population was genotyped in two sets, with the first set consisting of 95 lines and the second set containing all remaining lines. The raw GBS data consisted of 955,690 sites but had large amounts of missing data. Three methods to impute missing data were empirically compared: BEAGLE4 (Browning & Browning 2007), FSFHap (Swarts et al. 2014), and FILLIN (Swarts et al. 2014). Imputation accuracy was assessed on the first set (95 lines per population) while waiting for the second to be processed. The donor haplotype files required by FILLIN were downloaded from Panzea (http://www.panzea.org, accessed 27 June 2016) and consisted of the anonymized GBS 2.7 haplotypes made from 8000-site windows.

To gauge imputation accuracy, all genotype calls in the WEMA dataset with exactly seven sequencing reads were identified and a random 1% of these were masked using custom scripts (included in Supplemental File S2). The three algorithms were then run on the masked data and the resulting genotype calls for the masked sites were compared with their original calls. Accuracy was assessed as described by Swarts et al. (2014) and was defined as the Pearson correlation coefficient between the original and imputed datasets when major alleles are scored as 0, minor alleles as 1, and heterozygous calls as 0.5.

Heritability Estimation

Broad-sense heritability (H^2) was estimated by fitting a mixed linear model in R with the lme4 package (Bates et al. 2015). Station–year combination was fitted as a fixed effect and population and line as random effects. Heritability (H^2) was determined as the proportion of total phenotypic variance explained by the combined line and population terms:

$$H^2 = \frac{V_l + V_p}{V_t}, \quad [1]$$

where V_l is the variance explained by the individual lines, V_p is the variance explained by the population term, and V_t is the total phenotypic variance.

Narrow-sense heritability (h^2) was determined within each environment (experiment) by calculating a kinship matrix for each population in TASSEL (Bradbury et al. 2007) and applying it to a mixed linear model with a single dummy SNP. Narrow-sense heritability was estimated as the proportion of total variance explained by the kinship matrix:

$$h^2 = \frac{V_g}{V_g + V_r}, \quad [2]$$

where V_g is the genetic variance and V_r is the residual variance. This same calculation was attempted across the entire dataset but proved infeasible because of the excessive computing time required.

Joint GWAS

Only SNPs with an allele frequency of ≥ 0.05 across the entire dataset (380,530 SNPs) were included in the GWAS analysis. Joint GWAS was performed in R 3.1.2 (R Core Team 2015) across all populations by running a single-SNP scan across the genome with the following model:

$$\begin{aligned} Y = {} & \text{Population} + PC_{1\text{–}13} \\ & + \text{Additive:Population} \\ & + \text{Dominance:Population}, \end{aligned} \quad [3]$$

where Y is the matrix of genotypes, Population is a factor identifying the 15 WEMA populations, $PC_{1\text{–}13}$ is the first 13 principal coordinates, and Additive:Population and Dominance:Population are the additive and dominance genetic terms, respectively, nested within the Population term. Both population and principal coordinates were included to account for the strong population structure in WEMA (Fig. 1, Supplemental File S3). Principal coordinates were calculated with the cmdscale() (classic multidimensional scaling) function in R on a distance matrix created in TASSEL across all of the families. We chose to use 13 principal coordinates in the association model because additional ones explained negligible amounts of variance (Supplemental Fig. S3). Additive and dominance terms were nested within population to allow associations to have a different effect in each population; this was done on the basis of the results in the maize NAM population, where QTLs at the same locus can have different sizes or even directions of effect in different populations (e.g., Buckler et al. 2009). *P*-values were determined by performing a log-likelihood test between the full model (above) and a model with just the population and principal coordinate terms.

Bayesian GWAS

Bayesian GWAS was performed in a similar manner to joint GWAS, except that each field was first analyzed separately, so that the linear models included only additive and dominant genetic terms (no population or principal coordinates). The R package BayesFactor (Version 0.9.11–1, Morey & Rouder 2015) was used to calculate a Bayes factor for each SNP. The $\log_{10}$ of each Bayes factor was then taken; those with negative values were set to 0. Functionally, this means that data from a given field can only increase the probability that a SNP is associated with a trait. We set this minimum because keeping negative values would allow fields where a causal SNP is not segregating to degrade the signal from those where it is (Supplemental Fig. S4).

We then took advantage of the ability to use the posterior probability from one field as the prior probability for the next to simply add the log-Bayes factors together across fields to generate a unified Bayes factor for each population and for the entire WEMA population. The advantage of this approach is that each field could be run separately and the resulting Bayes factors combined at the end without the need to unify phenotypes across different environments.

FarmCPU Model Selection and Bootstrapping

To identify high-confidence SNPs through model selection, 100 bootstraps on each phenotype were performed, each time randomly setting 10% of the phenotypic values within each population as missing. Model selection was subsequently performed using FarmCPU (Liu et al. 2016) on each bootstrapped dataset. The resample model inclusion probability (RMIP; Valdar et al. 2009) value for each SNP is the fraction of bootstraps in which it was among the top 30 most significant SNPs identified by the model, based on initial tests that identified this as a good compromise between including signal and reducing noise. Only FarmCPU SNPs with RMIP ≥ 0.05 (5 out of 100 iterations) are discussed in the text because of their greater resolution.

Variance Explained

To determine the amount of variance explained by the identified SNPs, a model was fitted that included a population term, the first 13 genetic principal coordinates, and interactions between the population term and all additive and dominance terms derived from identified SNPs. Since missing genotypes would disrupt this model, missing genotype calls were set to the mean within each population; if an entire population was missing genotypes at a site, all such calls were set to the major allele (defined across the entire dataset). The variance explained by the resulting model (Table 2) is the adjusted R^2 value; the proportion of genetic variance explained is the adjusted R^2 divided by H^2.

Candidate Genes

Candidate genes for flowering time were identified from the literature (Danilevskaya et al. 2008, Dong et al. 2012, Hirsch et al. 2014). Plant height candidate genes were compiled by mining the MaizeGDB database (http://maizegdb.org/, accessed 28 June 2016) for known height mutants and searching the maize genome annotations on Phytozome Version 9 (http://www.phytozome.net, accessed 28 June 2016) for genes annotated with the terms "auxin," "brassinosteroid," and/or "gibberellin."

Genome-wide Prediction

Genome-wide prediction was performed using the rrBLUP package for R (Endelman 2011; R Core Team 2015). Genotypes used for prediction consisted of three sets: (i) 10,000 randomly selected SNPs from across the genome to sample the entire genome but also to incur a much smaller computational burden than using all ~380,000 segregating SNPs; (ii) the SNPs identified from FarmCPU model selection for each trait; and (iii) an

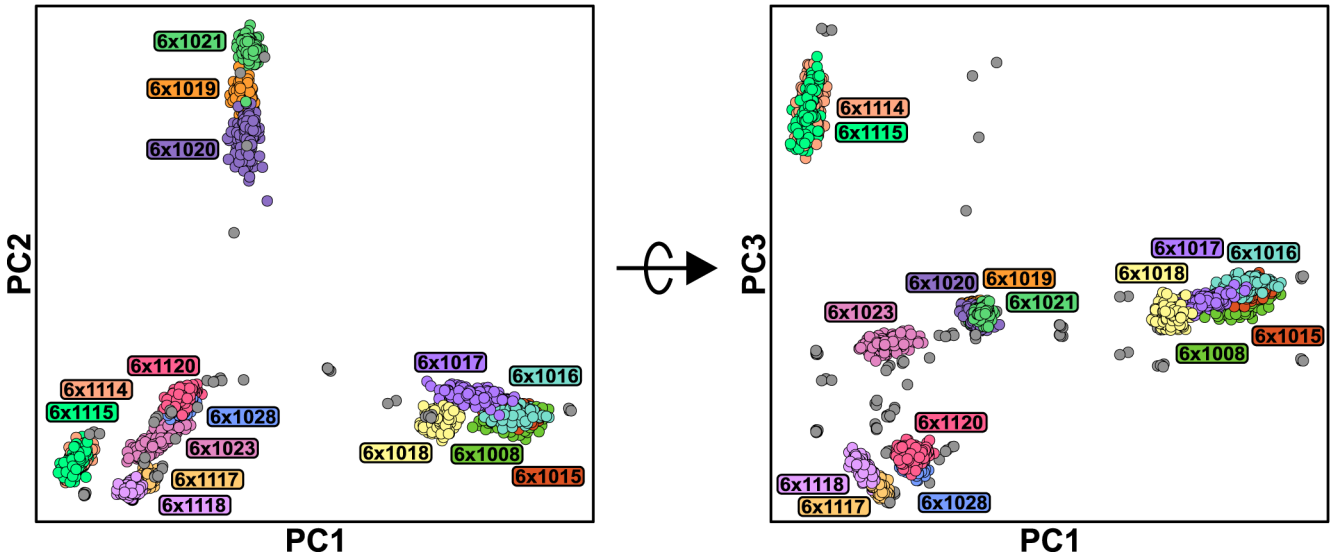

Fig. 1. Multidimensional scaling of the 15 Water-Efficient Maize for Africa (WEMA) biparental maize populations. The 15 WEMA populations that passed quality filtering are shown in a multidimensional scaling plot (principal coordinate analysis). The F_2 lines are colored according to population; gray dots show the parental lines. The populations separate into roughly four main clusters, of 6x1008–6x1015, 6x1019–6x1021, 6x1114 and 6x1115, and all others. An interactive three-dimensional version of this plot is included in Supplemental File S3.

Table 2. Genome-wide association summaries of FarmCPU hits.

Trait	Treatment	FarmCPU SNPs†	Lines (*n*)	Population–environment combinations (*n*)	Heritability‡	Variance explained§	Proportion of genetic variance explained
Anthesis date	Well-watered	115	2795	61	0.47	0.37	0.80
Anthesis date	Drought	108	2789	49	0.83	0.30	0.36
Plant height	Well-watered	143	2795	59	0.72	0.47	0.65
Plant height	Drought	120	2789	48	0.56	0.27	0.49

† SNP, single-nucleotide polymorphism.

‡ Broad-sense heritability (H^2) based on combined data from all environments.

§ Adjusted R^2 based on fitting a linear model with 13 principal coordinates, a population term, additive and dominance effects, and interactions between the population and additive and dominance effects.

equal number of random SNPs that have the same coverage and allele frequencies as the model-selected SNPs but that are at least 5 Mbp away from any hit. All models were run with 100 cross-validations, in which 10% of phenotypes were randomly masked. The resulting predictions were compared to the masked (true) values to generate a prediction accuracy estimate, measured using a Pearson correlation coefficient. *P*-values for the differences in distributions of correlations were determined via a two-sided *t*-test.

RESULTS AND DISCUSSION

Imputation Accuracy

The proportion of missing sites within population in the raw GBS data varied from 65.6 to 77.1% (Table 3). Of the three methods tested to impute this missing data, BEAGLE4 (Browning & Browning 2007) imputed the largest number of polymorphisms but provided the lowest accuracy (r^2 = 0.917), probably because it forces all sites to be imputed (Supplemental Fig. S5). FSFhap (Swarts et al. 2014) showed the highest accuracy (0.949) but the lowest

Table 3. Genotype statistics before and after imputation

	Raw data			Imputed and cleaned data		
Population code	MAF†	Missing	Hetero-zygosity	MAF	Missing	Hetero-zygosity
6x1008	0.37	0.71	0.03	0.37	0.04	0.29
6x1015	0.41	0.70	0.04	0.46	0.02	0.49
6x1016	0.19	0.70	0.01	0.45	0.02	0.47
6x1017	0.33	0.71	0.02	0.43	0.01	0.48
6x1018	0.40	0.66	0.06	0.35	0.04	0.25
6x1019	0.41	0.70	0.04	0.45	0.02	0.47
6x1020	0.40	0.71	0.04	0.42	0.03	0.40
6x1021	0.41	0.71	0.05	0.45	0.02	0.45
6x1023	0.38	0.71	0.03	0.46	0.01	0.47
6x1028	0.42	0.67	0.05	0.45	0.01	0.45
6x1114	0.23	0.69	0.03	0.23	0.02	0.23
6x1115	0.25	0.66	0.03	0.26	0.02	0.25
6x1117	0.18	0.67	0.03	0.26	0.02	0.50
6x1118	0.20	0.68	0.03	0.26	0.02	0.48
6x1120	0.40	0.71	0.04	0.47	0.04	0.45
Combined	0.18	0.77	0.01	0.19	0.08	0.12

† MAF, minor allele frequency

percent imputed, largely because of its inability to handle backcross populations at the time. FILLIN (Swarts et al. 2014) appears to be a good compromise for the WEMA data, as its imputation accuracy was nearly as good as FSFHap (0.942) while still imputing over 80% of the masked sites; for this reason, FILLIN was chosen to impute the full genotype dataset. After imputation, the proportion of missing sites within each population varied from 1.2 to 8.3%. Summary statistics for each population's genotype data before and after imputation are given in Table 3.

Multiple filtering steps were performed to remove low-quality genotypic data. Four of the original 19 populations selected for our study showed high levels of errant genotypes, probably because of contamination, mistaken parental identity, or an incorrect pollination method during their creation; these were excluded from all analyses (Supplemental Fig. S6). Some other populations (e.g., 6x1018) showed several genome segments that distorted the overall allele frequencies or heterozygosity levels, to the point that the genome-wide averages no longer matched their expected population type (F_2, BC_1, or BC_1F_2; see Table 1), even though most of the genome segregated as expected (Supplemental Fig. S7). These probably resulted from incomplete inbreeding of the founder lines, leading to segregation of more than two haplotypes at some genomic locations. All genotype data (raw, imputed, and cleaned) are available from Panzea (http://www.panzea.org, accessed 28 June 2016).

Phenotypes and Heritability

Phenotype data were collected by CIMMYT at multiple field stations (locations) in eastern and southern Africa, under both well-watered and managed drought stress conditions (Supplemental Fig. S1). No single location contained data for all of the 15 populations; additionally, the relative positions of each population within a location were not available, rendering each population–location combination effectively a separate experiment. Data from each location were checked and outlier phenotypes were discarded, together with one experiment that had an obvious entry error (Supplemental Fig. S8). The phenotypic variance differed greatly among environments (Fig. 2) and correlations across environments were generally low (Supplemental Fig. S9). Both raw and cleaned phenotype data are available in Supplementary File S1.

The H^2 for each trait across all populations within each treatment (well-watered or managed drought) varied from 0.47 to 0.83 (Table 2). Plant height was more heritable under well-watered conditions ($H^2 = 0.72$) than under drought conditions ($H^2 = 0.56$), the same pattern that was seen in these materials for grain yield and anthesis-silking interval (Semagn et al. 2013; Beyene et al. 2015). Days to anthesis, however, showed higher H^2 under drought conditions (0.83) than under well-watered conditions (0.47), although the GWAS analyses still explained more genetic and total variation under well-watered conditions despite the lower H^2 (Table 2). It may be that the drought stress exposed more differences among genotypes but most of the genes underlying these differences had too small of an individual effect to be detected.

To estimate mapping power more directly, h^2 values for each field location were also calculated (Fig. 3; see Supplemental Fig. S10 for distributions within individual populations). Calculation of mapping power across all field locations simultaneously would have been computationally prohibitive. Average h^2 ranged from 0.27 to 0.49, with higher values under well-watered conditions for both traits.

Genome-wide Association

Two algorithms for GWAS were performed for each phenotype and condition: a "normal" joint analysis and a Bayesian analysis. The joint GWAS used BLUPs from across all environments, whereas Bayesian analysis treated every population–environment combination independently and unified them only at the end.

The results of these analyses are shown in Fig. 4 and Fig. 5 for days to anthesis and plant height, respectively. Despite the different statistical frameworks, both methods yielded highly similar results. Several major genomic regions associated with each phenotype were detected; these were largely consistent across both treatment conditions. For days to anthesis, the largest associations were located on chromosomes 2, 5, 7, and 8, with treatment-specific peaks appearing on chromosome 6 (well-watered) and chromosome 10 (managed drought). For plant height, the strongest associations were found on chromosomes 2, 5, and 8, with some potential treatment-specific signals detected on chromosomes 3 and 7. Although linkage disequilibrium overall decays rapidly in WEMA (Supplemental Fig. S11), large linkage blocks result from variants present in only a few families. These blocks spread the association signal out, such that it was often not possible to pinpoint a specific genomic region at the core of each association.

Because of the relatively low resolution of the initial analyses, an additional analysis using bootstrapped phenotypes and model selection was performed using FarmCPU (Liu et al. 2016) to refine the location of associated SNPs (Fig. 4 and Fig. 5). This method of model selection is conceptually similar to the NAM GWAS method used previously for mapping within the US NAM population (Wallace et al. 2014), another large collection of biparental populations. Model selection improves GWAS results by identifying groups of individual SNPs that best explain the variance in the model (e.g., Wang et al. 2011; Segura et al. 2012). In practice, this results in higher resolution because it reduces association peaks to one or only a few specific sites. For this reason, we focus on the model-selected SNPs for the remainder of the paper.

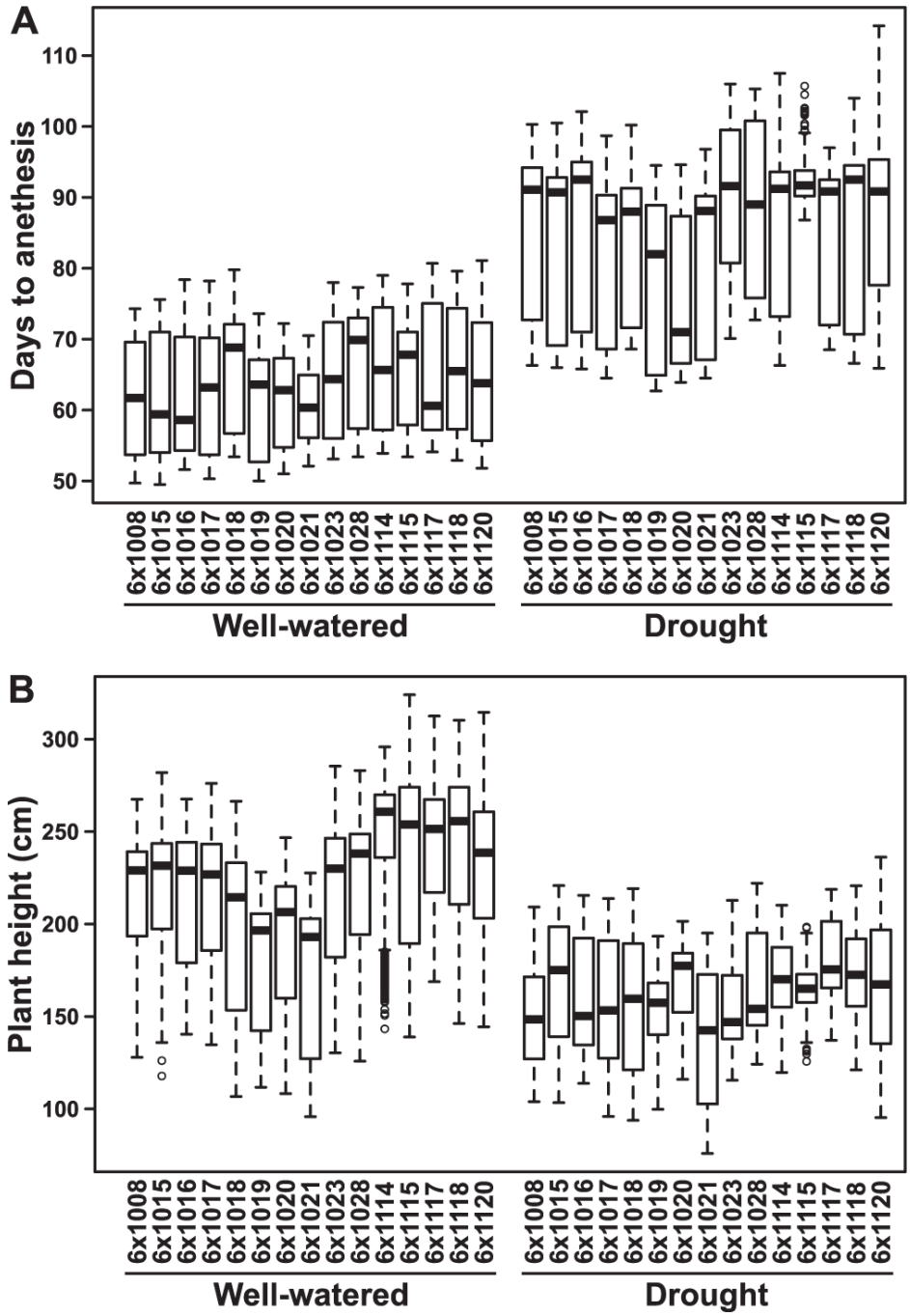

Fig. 2. Phenotypic variance across the 15 Water-Efficient Maize for Africa (WEMA) biparental maize populations and conditions. The distribution of phenotypes is shown for both days to anthesis (A) and plant height (B), divided by treatment condition and further by population. Note that since all well-watered trials were performed during the wet season and all drought trials during the dry season, the differences between conditions probably have additional factors (light quality, cumulative heat units, etc.) contributing to them beyond just water availability.

Flowering Time Associations

The results of analyses for flowering time are shown in Fig. 4. In total, we found 115 SNPs with significant associations with flowering time under well-watered conditions and 108 SNPs under managed drought conditions, explaining 80 and 36% of the genetic variance, respectively. A full list of associated SNPs and candidate genes is given in Supplemental File S1 but some of the strongest associations are highlighted below.

One of the strongest associations under both well-watered and drought conditions is near *Vgt1/ZmRap2.7* (chromosome 8). *Vgt1* is among the largest flowering time loci in maize (Buckler et al. 2009) and results from a MITE insertion that decreases expression of *ZmRap2.7*, a flowering time-related transcription factor 70 kb downstream (Salvi et al. 2007). Whether the same MITE insertion is segregating in the tropical maize populations is unknown but can be determined once whole-genome sequencing of the

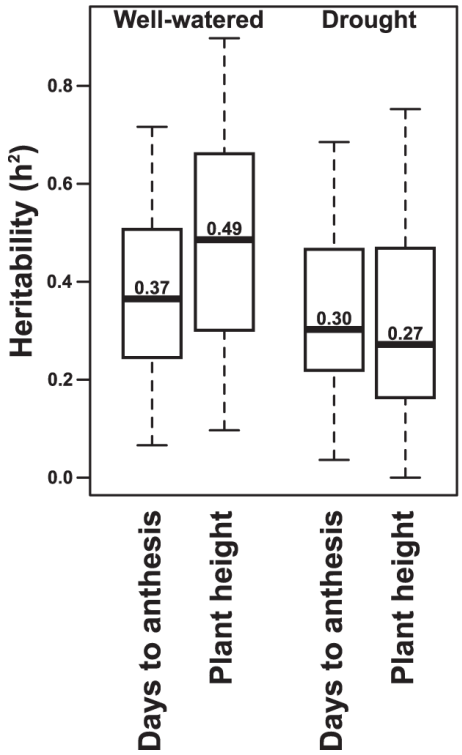

Fig. 3. Distribution of narrow-sense heritability of maize within treatment. Boxplots show the distribution of heritabilities within each experimental field for each trait under well-watered and drought conditions, with the median value indicated within each box.

founder parents of the 15 populations (currently underway) is completed.

We also found several associations near genes related to the circadian clock, including *ZmPRR37* and its paralog, *ZmPRR37*.1 (chromosomes 7 and 2, respectively). Both of these genes are part of the core oscillator in maize and are homologs of the *Arabidopsis thaliana* (L.) Heynh. gene *PRR3* (Hayes et al. 2010; Dong et al. 2012). In *A. thaliana*, PRR3 regulates the circadian clock by stabilizing another core circadian gene, TOC1, though PRR3 expression appears to be limited to the vasculature (Para et al. 2007). A minor association on chromosome 8 may tag *Gigz1a*, an ortholog of *A. thaliana GIGANTEA*, shown to be involved in photoperiod sensing in maize (Miller et al. 2008; Dong et al. 2012). This signal is detected primarily from the raw GWAS data; however, when FarmCPU model selection is applied, the position of the strongest SNP shifts and instead tags *ZCN14* (see below).

Two strong associations were found with the phosphatidylethanolamine-binding protein population kinases, *ZCN14* and *ZCN18* (chromosomes 8 and 2, respectively). Phosphatidylethanolamine-binding protein population kinases are transcription factors involved in the control of flowering time, with the best studied members being the *A. thaliana* genes *FLOWERING LOCUS T* (*FT*) and *TERMINAL FLOWER 1* (Danilevskaya et al. 2008). *ZCN14* is the most highly conserved member of the population relative to *FT*, although its expression pattern is very different: *A. thaliana FT* is limited to the leaf blade, whereas maize *ZCN14* expression occurs in reproductive primordia and thus after the transition to reproductive growth (Danilevskaya et al. 2008). The discovery of a strong association signal near *ZCN14* implies that *ZCN14* still has some function in flowering time control. Less is known about *ZCN18*, which is expressed mostly in stems and (to a lesser extent) in leaves. Although a signal is also found near *ZCN8*, another phosphatidylethanolamine-binding protein kinase population member involved in flowering time, FarmCPU did not identify it as a major association; instead, it identifies the SNPs around *Vgt1* mentioned earlier.

Another strong association occurs near *zfl1* (chromosome 10), a homolog of the *A. thaliana LEAFY* gene. *zfl1* is thought to be an upstream regulator of conserved floral identity genes. Loss-of-function mutants of *zfl1* and its paralog, *zfl2*, display a variety of defects in floral development (Bomblies et al. 2003).

Two MADS-box proteins with known functions in floral development, Zap1b and ZmM5 (chromosomes 7 and 9), are also located near strong association signals. *Zap1b* (also called *ZmMADS1*) is adjacent to a cluster of identified SNPs, one of which shows strong association under drought conditions, whereas *ZmM5* (*ZmMADS3*) is near a consistently strong association under both well-watered and drought conditions. Ectopic expression of *Zap1b* results in severe developmental defects in transgenic maize lines (Heuer et al. 2001), including defects in flower development. Less is known about *ZmM5*, although it has similar expression patterns to *Zap1b* (embryo, developing flowers, and stem nodes) (Heuer et al. 2001) and so is probably involved in similar developmental programs.

Lastly, we found a weak association near *PhyA2* (chromosome 5), a phytochrome gene. Phytochromes sense light quality, with phytochrome A sensing far-red light (Quail 2002) and often being involved in the photoperiod response. However, rice (*Oryza sativa* L.) *PhyA* mutants show a normal photoperiod response (Takano et al. 2001), indicating that maize PhyA2 may influence flowering time through a different mechanism.

Interestingly, we found no signal around the *Dwarf8* locus on chromosome 1. This region is near several flowering time candidate genes and has been implicated in flowering time before (Thornsberry et al. 2001), yet none of our three methods detected any significant signal here (Fig. 4). Although the link between *Dwarf8* itself and flowering time has been called into question (Larsson et al. 2013), it is still surprising that none of the candidate genes in this region associated with flowering-time effects in these tropical maize populations.

We also found several associations that did not appear to be associated with candidate genes, specifically on chromosomes 5, 6, and 8 (Fig. 4). These may represent novel

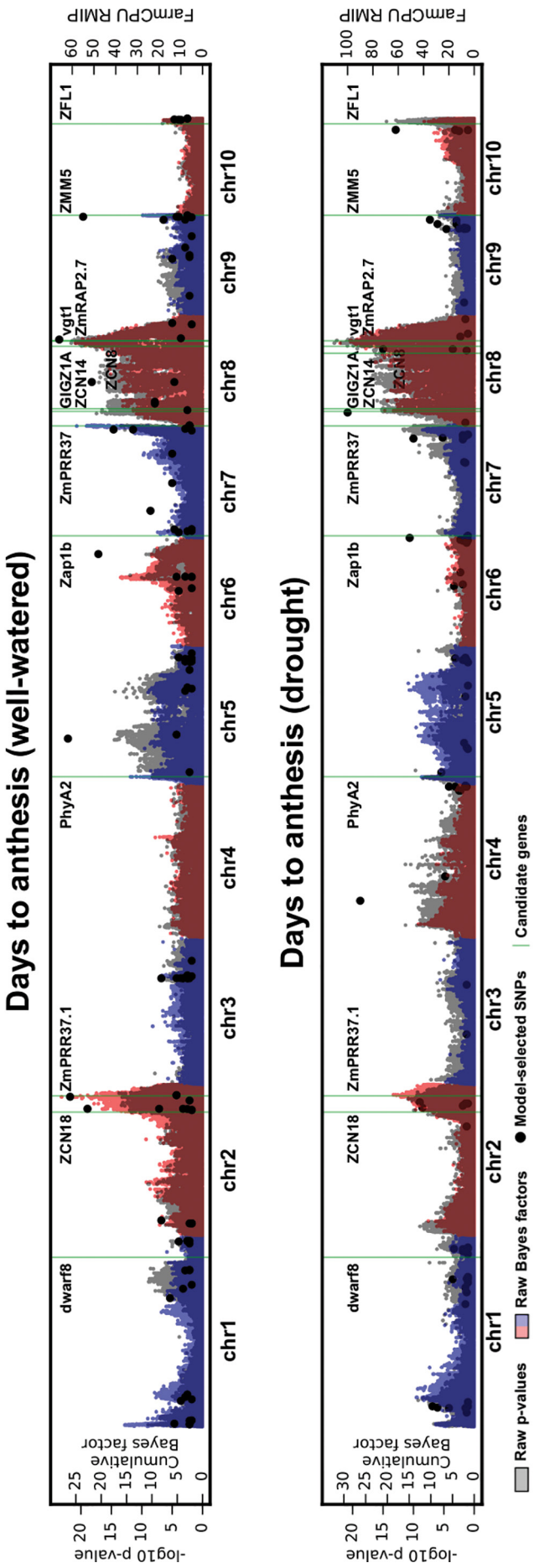

Fig. 4. Genome-wide association (GWAS) analysis of days to anthesis in maize under well-watered and drought stress conditions. The combined association analyses for days to anthesis across all 15 WEMA populations is shown for well-watered (A) and drought (B) conditions. The Bayesian (red or blue) and joint GWAS (gray) results are shown to indicate the raw association score at each point; the two methods show strong concordance. FarmCPU model-selected single-nucleotide polymorphisms (SNPs) are shown as dark black circles and the nearby candidate genes as vertical green lines. All SNPs discussed in the main text come from the model-selected set. RMIP, resample model inclusion probability.

genes that are not known to play a role in flowering time. Alternatively, some might be explained by structural variation in the maize genome (Chia et al. 2012), resulting in annotated genes from the reference genome actually being in different genomic locations within the WEMA tropical maize populations, thus putting the association peaks in new locations. For developing Africa-adapted improved tropical maize germplasm with drought tolerance, identifying the nature or exact location of these genes is perhaps less important than finding the significant marker associations that can be used to introgress favorable alleles through marker-assisted breeding. Both novel loci and well-characterized ones will continue to be mined for useful alleles in tropical-adapted germplasm that might be useful for breeding applications.

Plant Height Associations

We found 143 SNPs associated with plant height under well-watered conditions and 120 SNPs under drought conditions, explaining 65 and 49% of the total genetic variance, respectively. The loci identified for plant height show more variation than for flowering time, in that more hits appear to demonstrate stronger effects (as measured by $-\log_{10}$ *P*-values) under one treatment condition than the other. In addition, since there were many more previously known candidate genes for plant height than for flowering time, many of the hits in this study were relatively close to multiple candidates.

The strongest SNP associations were located near several auxin-related genes. Auxin is one of the major regulators of plant growth and architecture (Gallavotti 2013); therefore, it is not surprising to find strong associations near genes related to it. The most robust signal is *auxin response factor-transcription factor 20* (chromosome 5), which was identified by FarmCPU in 90% of bootstrap iterations under well-watered conditions and 84% under managed drought conditions. Auxin response factor population proteins directly bind to auxin to initiate downstream signaling by binding to conserved auxin response elements that modulate the expression of downstream genes (Guilfoyle et al. 1998). Under well-watered conditions, we also found an association near *GRMZM2G108851* (chromosome 8). This gene has been annotated as both a putative auxin-responsive gene and a cytochrome b561 or ferric reductase transmembrane protein, indicating that it is a membrane-bound electron shuttle (INTERPRO domain *IPR006593*). Since such electron carriers participate in a wide variety of biochemical

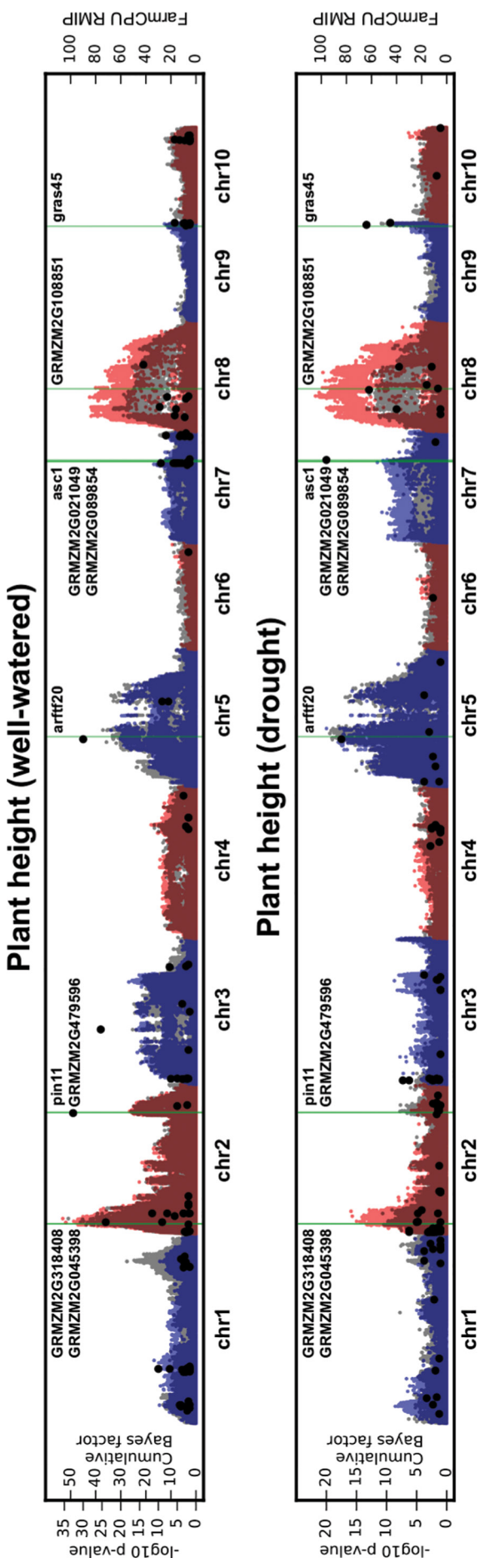

Fig. 5. Genome-wide association analysis (GWAS) of maize plant height under well-watered and drought stress conditions. The combined associations for plant height are shown for well-watered (A) and drought (B) conditions across the Water-Efficient Maize for Africa (WEMA) populations. The Bayesian (red or blue) and joint GWAS (gray) results are shown to indicate the raw association score at each point; the two methods show strong concordance. FarmCPU model-selected single-nucleotide polymorphisms (SNPs) are shown as dark black circles and the nearby candidate genes as vertical green lines. RMIP, resample model inclusion probability.

functions, its potential role in height determination (beyond responding to auxin) is unknown.

Under well-watered conditions, we found an extremely robust association (RMIP = 0.98) on chromosome 2 near two auxin-related genes: *Pin11* and *GRMZM2G479596*. *Pin11* is a member of the PIN-FORMED auxin transporter population, the members of which are involved in shuttling auxin across cellular membranes (Křeček et al. 2009). *GRMZM2G479596* is a putative (and uncharacterized) auxin-response factor. Under drought conditions, however, this locus exhibited reduced signal and the strongest signal (RMIP = 0.96) was on chromosome 7 instead. Several candidate genes could be found near this hit, though the closest was *GRMZM2G089854*, which has homology to a putative auxin-independent growth promoter in rice. Other nearby genes include the cyclin gene *asc1* and another putative auxin-responsive protein, GRMZM2G021049. Although little is known about the latter, mutations of asc1 are known to result in short, infertile plants (Brooks et al. 2009).

In addition to auxin-related genes, an association near the beginning of chromosome 2 is located near two brassinosteroid-related genes, *GRMZM2G045398* and *GRMZM2G318408*. Brassinosteroids are plant hormones related to growth, and mutants defective in brassinosteroid synthesis exhibit strong dwarfing phenotypes across several species (Bishop and Koncz 2002). Both proteins exhibit homology to *A. thaliana* BRH1 (brassinosteroid-responsive RING-H2). Downregulation of *BRH1* in *A. thaliana* does not affect plant height (but does affect stem and leaf morphology), whereas overexpression mutants showed no distinct phenotype (Molnár et al. 2002), so it is unknown what effect these genes may have in maize. An association was also found near *gras45* (*GRMZM2G028039*), a *GIBBERELLIC ACID-INSENSITIVE, REPRESSOR of GA1*, and *SCARECROW* population transcription factor localized to the end of chromosome 9. GRAS-population transcription factors are involved in several aspects of plant signaling, including gibberellic acid signaling and phytochrome A signal transduction (Hirsch & Oldroyd 2009), both of which have potential for involvement in plant height.

A strong association (RMIP = 0.76) on chromosome 3 under well-watered conditions had no nearby candidate gene. As with flowering time, this could represent a genuinely novel association or it could also be an alignment mismatch between the WEMA materials and the B73 reference genome. Again, identifying the exact gene responsible for each association is less important than the ability to use these

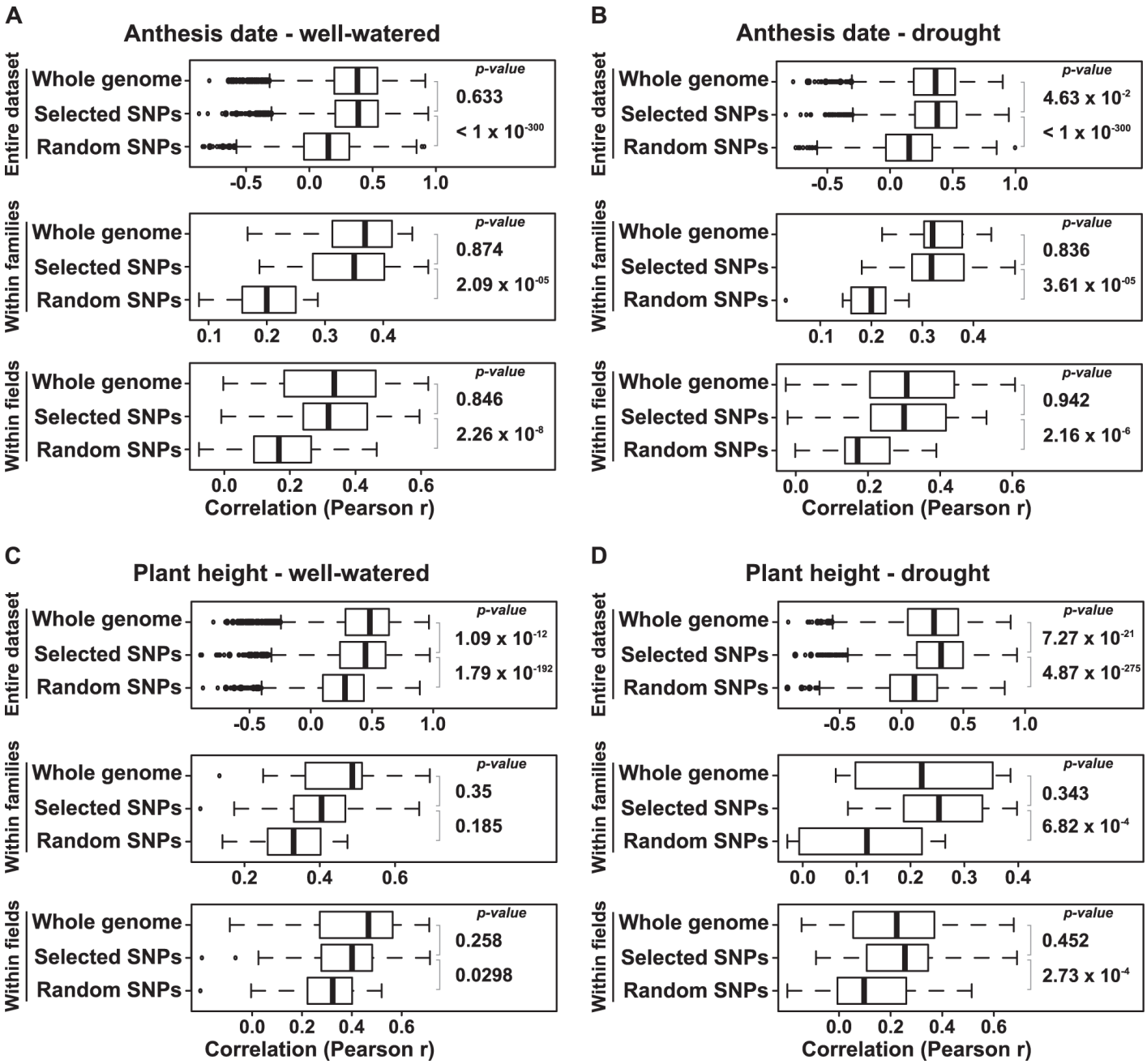

Fig. 6. Genome-wide prediction of maize phenotypes. Phenotypic prediction was performed across the entire Water-Efficient Maize for Africa (WEMA) dataset in 100 cross-validations using either genome-wide single-nucleotide polymorphisms (SNPs), the model-selected SNPs, or an equivalent number of random SNPs matched for allele frequencies and coverage. Each panel (A–D) shows the distribution of raw correlations across the entire WEMA dataset (top), the mean correlations within each family (middle), and the mean correlations within each field (bottom). *P*-values from comparing distributions via a two-sided *t*-test are shown to the right of each plot.

GWAS hits as tags for loci that could be potentially used to improve tropical maize germplasm.

Genome-wide Prediction

To confirm the utility of the identified SNPs, genome-wide prediction was performed both across and within the WEMA populations. Although the prediction accuracy was modest (average correlations of 0.3–0.4), the SNPs selected from GWAS had the same accuracy as those found by using a genome-wide SNP set and were significantly more accurate than an equivalent number of random SNPs (Fig. 6). The only time this did not occur was for plant height under well-watered conditions, where the difference within populations and within fields followed a similar pattern but was not statistically significant.

CONCLUSIONS

Both plant height and flowering time are important agronomic traits that are significantly impacted by drought stress. Several hits near candidate genes known to

control these traits were identified, along with discovery of a number of novel loci. Some of the alleles at these loci could be unique to tropical maize germplasm, since tropical germplasm generally has higher diversity than temperate germplasm in maize. The next step is to extend this analysis to other important drought-related traits, then to move desirable loci into elite breeding lines through, for example, marker-assisted selection or genomic selection. The regions identified in these populations are likely to be shared with other maize varieties, but given the high diversity of tropical maize, only a subset is likely to be shared with any given inbred or landrace accession.

A major aim of this analysis was to determine methods to unify data across the diverse WEMA populations and, more generally, across similar datasets containing multiple populations, testing locations, and treatment conditions. Toward this end, three methods were compared: a joint GWAS analysis, a Bayesian GWAS approach, and a model selection approach. Both the joint GWAS and FarmCPU model selection approaches relied on BLUPs made by unifying the phenotypes across environments. This sort of unification frequently uses common check lines to account for environment-specific effects; two such checks were present in all environments in this study but this is not always the case in ad hoc breeding datasets. In contrast, the Bayesian approach handles each environment separately and has no such constraint. Additionally, the Bayesian GWAS is modular, such that environments can be added or dropped with minimal computational overhead. Similar modifications of analysis models for the other two methods would require rerunning the analysis, often at great computational expense. Unfortunately, the higher resolution gained by model selection is not available for the Bayesian approach; consequently, no single approach evaluated is completely superior to the others.

Genome-wide prediction results indicated that the loci identified underlie the majority of the genetic variation for traits evaluated in this germplasm. Although our prediction accuracy agreed with previous analyses in maize (reviewed in Lin et al. 2014), higher accuracy is always the goal. One way to potentially improve prediction accuracy across populations would be to include genotype × environment (G × E) interaction terms in the model (Lopez-Cruz et al. 2015). This was not performed in the current analysis, primarily because the large number of separate environments caused by the lack of location data would have led to a massive number of model terms. Including all such terms would not only raise the computational burden exponentially but also overparameterize many of the models. Nonetheless, we can identify locations where G × E interactions occurred by using the Bayesian association analysis, which includes it implicitly by mapping in each field separately (Supplemental Fig. S12). A previous analysis of these materials that treated each field station as a unified environment found that including G × E interactions increased genomic prediction accuracy for whole-genome SNPs (Zhang et al. 2015), implying that the same would probably hold true for these model-selected SNPs. More generally, the different reactions of plants under drought versus well-watered conditions are a type of G × E. This interaction is apparent in the different heritabilities of the traits under well-watered and drought conditions (Table 2). It also probably explains other systematic differences between the two conditions; for example, the raw associations for plant height (Fig. 5) have –log10 *P*-values and cumulative Bayes factors that are roughly twice as high under well-watered conditions as under drought. Some of this is caused by the lower heritability, but some may also be a result of the signal being spread across more loci, thus making each of them appear less significant overall.

Many groups are working to increase the speed and efficiency of phenotypic data collection, especially at the field level (Prasanna et al., 2013; Araus and Cairns, 2014). These efforts take various forms, from entering data directly into databases with tablet computers to mounting sensors on tractors or unmanned aerial vehicles (drones) to scan entire fields rapidly. Nonetheless, for the foreseeable future, phenotyping will remain a major bottleneck for plant breeding and therefore efforts to maximize the utility of information from existing phenotype data will become increasingly valuable. Even with high-throughput phenotyping methods, the ability to unify ever larger and more disparate datasets will still be important, so the development of user-friendly and robust statistical methods and software packages to process and analyze such large and complex datasets efficiently is extremely important. Efforts to improve genomic prediction and to decrease the length of breeding cycles—another key constraint to improving genetic gain—are also underway. The combination of these methods should enable faster, more powerful applied breeding to meet the needs of a global 21st-century population.

Acknowledgments

We thank Sara Miller for editing the manuscript, Alberto Romero and Cinta Romay for their compiled list of flowering time candidate genes, and members of the Buckler lab for general feedback. This work was supported by National Science Foundation grant IOS-1238014, the USDA – Agricultural Research Service (USDA-ARS), the University of Georgia, Consultative Group for International Agricultural Research Program MAIZE, Cornell–CIMMYT Genomic Selection project financed by the Bill and Melinda Gates Foundation, and the WEMA project funded by the Bill & Melinda Gates Foundation, the Howard G. Buffett Foundation, and the United States Agency for International Development.

References

Andorf, C.M., E.K. Cannon, J.L. Portwood, J.M. Gardiner, L.C. Harper, M.L. Schaeffer, et al. (2015) MaizeGDB 2015: New

tools, data, and interface for the maize model organism database. Nucleic Acids Res. 44(D1): D1195–D1201. doi:10.1093/nar/gkv1007

Araus, J.L., and J.E. Cairns. 2014. Field high-throughput phenotyping: The new crop breeding frontier. Trends Plant Sci. 19:52–61. doi:10.1016/j.tplants.2013.09.008

Bates, D., M. Mächler, B.M. Bolker, and S.C. Walker. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Software 67(1). doi:10.18637/jss.v067.i01

Beyene, Y., K. Semagn, S. Mugo, A. Tarekegne, R. Babu, B. Meisel, et al.. 2015. Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. Crop Sci. 55(1): 154–163. doi:10.2135/cropsci2014.07.0460

Bishop, G.J., and C. Koncz. 2002. Brassinosteroids and plant steroid hormone signaling. Plant Cell 14(Suppl):S97–S110.

Bolaños, J., and G.O. Edmeades. 1996. The importance of the anthesis-silking interval in breeding for drought tolerance in tropical maize. F. Crop. Res. 48(1):65–80. doi:10.1016/0378-4290(96)00036-6

Bomblies, K., R.-L. Wang, B. Ambrose, R.J. Schmidt, R.B. Meeley, and J. Doebley. 2003. Duplicate *FLORICAULA/LEAFY* homologs *zfl1* and *zfl2* control inflorescence architecture and flower patterning in maize. Development 130(11):2385–2395. doi:10.1242/dev.00457

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(19):2633–2635. doi:10.1093/bioinformatics/btm308

Brooks, L., J. Strable, X.L. Zhang, K. Ohtsu, R.L. Zhou, A. Sarkar, et al. 2009. Microdissection of shoot meristem functional domains. PLoS Genet. 5(5):E1000476. doi:10.1371/journal.pgen.1000476

Browning, S.R., and B.L. Browning. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81(5):1084–1097. doi:10.1086/521987

Buckler, E.S., J.B. Holland, P.J. Bradbury, C.B. Acharya, P.J. Brown, C. Browne, et al. 2009. The genetic architecture of maize flowering time. Science 325(5941):714–718. doi:10.1126/science.1174276

Chia, J.M., C. Song, P.J. Bradbury, D. Costich, N. de Leon, J. Doebley, et al. 2012. Maize HapMap2 identifies extant variation from a genome in flux. Nat. Genet. 44(7):803–807. doi:10.1038/ng.2313

Danilevskaya, O.N., X. Meng, Z. Hou, E.V. Ananiev, and C.R. Simmons. 2008. A genomic and expression compendium of the expanded *PEBP* gene population from maize. Plant Physiol. 146(1):250–264. doi:10.1104/pp.107.109538

Dong, Z., O. Danilevskaya, T. Abadie, C. Messina, N. Coles, and M. Cooper. 2012. A gene regulatory network model for floral transition of the shoot apex in maize and its dynamic modeling. PLoS One 7(8):E43450. doi:10.1371/journal.pone.0043450

Elshire, R.J., J.C. Glaubitz, Q. Sun, J. Poland, K. Kawamoto, E.S. Buckler, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6(5):E19379. doi:10.1371/journal.pone.0019379

Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4:250–255. doi:10.3835/plantgenome2011.08.0024

Gallavotti, A. 2013. The role of auxin in shaping shoot architecture. J. Exp. Bot. 64(9):2593–2608. doi:10.1093/jxb/ert141

Glaubitz, J., T. Casstevens, and F. Lu. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One 9(2):E90346. doi:10.1371/journal.pone.0090346

Guilfoyle, T.J., T. Ulmasov, and G. Hagen. 1998. The ARF family of transcription factors and their role in plant hormone-responsive transcription. Cell. Mol. Life Sci. 54(7):619–627. doi:10.1007/s000180050190

Hartwig, T., G.S. Chuck, S. Fujioka, A. Klempien, R. Weizbauer, D.P.V. Potluri, et al. 2011. Brassinosteroid control of sex determination in maize. Proc. Natl. Acad. Sci. USA 108(49):19814–19819. doi:10.1073/pnas.1108359108

Hayes, K.R., M. Beatty, X. Meng, C.R. Simmons, J.E. Habben, and O.N. Danilevskaya. 2010. Maize global transcriptomics reveals pervasive leaf diurnal rhythms but rhythms in developing ears are largely limited to the core oscillator. PLoS One 5(9):E12887. doi:10.1371/journal.pone.0012887

Heuer, S., S. Hansen, J. Bantin, R. Brettschneider, E. Kranz, H. Lörz, et al. 2001. The maize MADS box gene *ZmMADS3* affects node number and spikelet development and is co-expressed with *ZmMADS1* during flower development, in egg cells, and early embryogenesis. Plant Physiol. 127(1):33–45. doi:10.1104/pp.127.1.33

Hirsch, C.N., J.M. Foerster, J.M. Johnson, R.S. Sekhon, G. Muttoni, B. Vaillancourt, et al. 2014. Insights into the maize pan-genome and pan-transcriptome. Plant Cell 26(1):121–135. doi:10.1105/tpc.113.119982

Hirsch, S., and G.E.D. Oldroyd. 2009. GRAS-domain transcription factors that regulate plant development. Plant Signal. Behav. 4(8):698–700. doi:10.4161/psb.4.8.9176

Kover, P.X., W. Valdar, J. Trakalo, N. Scarcelli, I.M. Ehrenreich, M.D. Purugganan, et al. 2009. A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. PLoS Genet. 5(7):E1000551. doi:10.1371/journal.pgen.1000551

Křeček, P., P. Skůpa, J. Libus, S. Naramoto, R. Tejos, J. Friml, et al. 2009. The PIN-FORMED (PIN) protein family of auxin transporters. Genome Biol. 10(12):249. doi:10.1186/gb-2009-10-12-249

Kuleshov, N.N. 1933. World's diversity of phenotypes of maize. J. Am. Soc. Agron. 25:688. doi:10.2134/agronj1933.00021962002500100006x

Larsson, S.J., A.E. Lipka, and E.S. Buckler. 2013. Lessons from *Dwarf8* on the strengths and weaknesses of structured association mapping. PLoS Genet. 9(2):E1003246. doi:10.1371/journal.pgen.1003246

Lin, Z., B.J. Hayes, and H.D. Daetwyler. 2014. Genomic selection in crops, trees and forages: A review. Crop Pasture Sci. 65:1177–1191. doi:10.1071/CP13363

Liu, X.L., M. Huang, B. Fan, E.S. Buckler, and Z.Z. Zhang. 2016. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet. 12(2):E1005767. doi:10.1371/journal.pgen.1005767

Lopez-Cruz, M., J. Crossa, D. Bonnett, S. Dreisigacker, J. Poland, J.-L. Jannink, et al. 2015. Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. G3 (Bethesda) 5(4): 569–582.

McMullen, M.D., S. Kresovich, H.S. Villeda, P. Bradbury, H.H. Li, Q. Sun, et al. 2009. Genetic properties of the maize nested association mapping population. Science 325(5941):737–740. doi:10.1126/science.1174320

Miller, T.A., E.H. Muslin, and J.E. Dorweiler. 2008. A maize *CONSTANS*-like gene, *conz1*, exhibits distinct diurnal expression patterns in varied photoperiods. Planta 227(6):1377–1388. doi:10.1007/s00425-008-0709-1

Molnár, G., S. Bancoş, F. Nagy, and M. Szekeres. 2002. Characterisation of *BRH1*, a brassinosteroid-responsive RING-H2 gene from *Arabidopsis thaliana*. Planta 215(1):127–133. doi:10.1007/s00425-001-0723-z

Morey, R. D., Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs. R package version 0.9.12–2. R Foundation for Statistical Computing. https://cran.r-project.org/web/packages/BayesFactor/index.html (accessed 1 July 2016).

Multani, D.S., S.P. Briggs, M.A. Chamberlin, J.J. Blakeslee, A.S. Murphy, and G.S. Johal. 2003. Loss of an MDR transporter in compact stalks of maize *br2* and sorghum *dw3* mutants. Science 302(5642):81–84. doi:10.1126/science.1086072

Para, A., E.M. Farre, T. Imaizumi, J.L. Pruneda-Paz, F.G. Harmon, and S.A. Kay. 2007. PRR3 is a vascular regulator of TOC1 stability in the Arabidopsis circadian clock. Plant Cell 19(11): 3462–3473.

Peiffer, J.A., M.C. Romay, M.A. Gore, S.A. Flint-Garcia, Z.W. Zhang, M.J. Millard, et al. 2014. The genetic architecture of maize height. Genetics 196(4):1337–1356. doi:10.1534/genetics.113.159152

Prasanna, B.M., J.L. Araus, J. Crossa, J.E. Cairns, N. Palacios, B. Das, et al. 2013. High-throughput and precision phenotyping for cereal breeding programs. In: P.K. Gupta and R.K. Varshney, editors, Cereal genomics II. Springer-Verlag, Dordrecht. p. 341–374.

Quail, P.H. 2002. Phytochrome photosensory signalling networks. Nat. Rev. Mol. Cell Biol. 3(2):85–93. doi:10.1038/nrm728

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/ (accessed 28 June 2016).

Salas Fernandez, M.G., P.W. Becraft, Y. Yin, and T. Lübberstedt. 2009. From dwarves to giants? Plant height manipulation for biomass yield. Trends Plant Sci. 14(8):454–461. doi:10.1016/j.tplants.2009.06.005

Salvi, S., G. Sponza, M. Morgante, D. Tomes, X. Niu, K.A. Fengler, et al. 2007. Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proc. Natl. Acad. Sci. USA 104(27):11376–11381. doi:10.1073/pnas.0704145104

Segura, V., B.J. Vilhjálmsson, A. Platt, A. Korte, Ü. Seren, Q. Long, et al. 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat. Genet. 44(7):825–830. doi:10.1038/ng.2314

Semagn, K., Y. Beyene, M.L. 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(1):313. doi:10.1186/1471-2164-14-313

Swarts, K., H.H. Li, J.A.R. Navarro, D. An, M.C. Romay, S. Hearne, et al. 2014. Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. Plant Genome 7(3): doi:10.3835/plantgenome2014.05.0023.

Takano, M., H. Kanegae, T. Shinomura, A. Miyao, H. Hirochika, and M. Furuya. 2001. Isolation and characterization of rice phytochrome A mutants. Plant Cell 13(3):521–534. doi:10.1105/tpc.13.3.521

Thornsberry, J.M., M.M. Goodman, J. Doebley, S. Kresovich, D. Nielsen, and E.S. Buckler. 2001. *Dwarf8* polymorphisms associate with variation in flowering time. Nat. Genet. 28(3):286–289. doi:10.1038/90135

Valdar, W., C.C. Holmes, R. Mott, and J. Flint. 2009. Mapping in structured populations by resample model averaging. Genetics 182(4):1263–1277. doi:10.1534/genetics.109.100727

Vlăduţu, C., J. McLaughlin, and R.L. Phillips. 1999. Fine mapping and characterization of linked quantitative trait loci involved in the transition of the maize apical meristem from vegetative to generative structures. Genetics 153(2):993–1007.

Wallace, J.G., P.J. Bradbury, N. Zhang, Y. Gibon, M. Stitt, and E.S. Buckler. 2014. Association mapping across numerous traits reveals patterns of functional variation in maize. PLoS Genet. 10(12):E1004845. doi:10.1371/journal.pgen.1004845

Wang, D., K.M. Eskridge, and J. Crossa. 2011. Identifying QTLs and epistasis in structured plant populations using adaptive mixed LASSO. J. Agric. Biol. Environ. Stat. 16(2):170–184. doi:10.1007/s13253-010-0046-2

Wang, Y., J. Yao, Z. Zhang, and Y. Zheng. 2006. The comparative analysis based on maize integrated QTL map and meta-analysis of plant height QTLs. Chin. Sci. Bull. 51(18):2219–2230. doi:10.1007/s11434-006-2119-8

Winkler, R.G., and T. Helentjaris. 1995. The maize *Dwarf3* gene encodes a cytocrhome p450-mediated early step in gibberellin biosynthesis. Plant Cell 7:1307–1317. doi:10.1105/tpc.7.8.1307

Yang, Q., Z. Li, W.Q. Li, L.X. Ku, C. Wang, J.R. Ye, et al. 2013. CACTA-like transposable element in *ZmCCT* attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. Proc. Natl. Acad. Sci. USA 110(42):16969–16974. doi:10.1073/pnas.1310949110

Zhang, X., P. Perez-Rodriguez, K. Semagn, Y. Beyene, R. Babu, M.A. Lopez-Cruz, et al. 2015. Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. Heredity 114:291–299. doi:10.1038/hdy.2014.99