Published online May 25, 2017

ORIGINAL RESEARCH

Genome-Wide Analysis of Tar Spot Complex Resistance in Maize Using Genotyping-by-Sequencing SNPs and Whole-Genome Prediction

Shiliang Cao, Alexander Loladze, Yibing Yuan, Yongsheng Wu, Ao Zhang, Jiafa Chen, Gordon Huestis, Jingsheng Cao, Vijay Chaikam, Michael Olsen, Boddupalli M. Prasanna, Felix San Vicente*, and Xuecai Zhang*

Abstract

Tar spot complex (TSC) is one of the most destructive foliar diseases of maize (*Zea mays* L.) in tropical and subtropical areas of Central and South America, causing significant grain yield losses when weather conditions are conducive. To dissect the genetic architecture of TSC resistance in maize, association mapping, in conjunction with linkage mapping, was conducted on an association-mapping panel and three biparental doubled-haploid (DH) populations using genotyping-by-sequencing (GBS) single-nucleotide polymorphisms (SNPs). Association mapping revealed four quantitative trait loci (QTL) on chromosome 2, 3, 7, and 8. All the QTL, except for the one on chromosome 3, were further validated by linkage mapping in different genetic backgrounds. Additional QTL were identified by linkage mapping alone. A major QTL located on bin 8.03 was consistently detected with the largest phenotypic explained variation: 13% in association-mapping analysis and 13.18 to 43.31% in linkage-mapping analysis. These results indicated that TSC resistance in maize was controlled by a major QTL located on bin 8.03 and several minor QTL with smaller effects on other chromosomes. Genomic prediction results showed moderate-to-high prediction accuracies in different populations using various training population sizes and marker densities. Prediction accuracy of TSC resistance was >0.50 when half of the population was included into the training set and 500 to 1,000 SNPs were used for prediction. Information obtained from this study can be used for developing functional molecular markers for marker-assisted selection (MAS) and for implementing genomic selection (GS) to improve TSC resistance in tropical maize.

Core Ideas

- Association and linkage mapping are effective for dissecting genetic architecture of complex traits in maize.
- TSC resistance in maize is controlled by a major QTL and several minor QTL.
- Major QTL on bin 8.03 confirmed by association and linkage mapping.
- TSC resistance in tropical maize could be improved by MAS and GS individually or stepwise.

TAR SPOT COMPLEX is one of the most important foliar diseases of maize in many Central and South American tropical and subtropical areas with moderately cooler and humid climates. Severe TSC outbreaks occurring

S.L. Cao, and J.S. Cao, Maize Research Institute, Heilongjiang Academy of Agricultural Sciences, Harbin 150086, Heilongjiang, China; S.L. Cao, A. Loladze, J.F. Chen, G. Huestis, F. San Vicente, and X.C. Zhang, International Maize and Wheat Improvement Center (CIMMYT), El Batan, Texcoco 56237, México; Y.B. Yuan, Maize Research Institute, Sichuan Agricultural Univ., Wenjiang 611130, Sichuan, China; Y.S. Wu, Maize Research Institute, Guangxi Academy of Agricultural Sciences, Nanning 530007, Guangxi, China; A. Zhang, Agronomy College, Shenyang Agricultural University, Shenyang 110866, Liaoning, China; V. Chaikam, M. Olsen, and B.M. Prasanna, International Maize and Wheat Improvement Center (CIMMYT), P. O. Box 1041, Village Market, Nairobi 00621, Kenya. S. Cao and A. Loladze contributed equally to this work. Received 3 Oct. 2016. Accepted 24 Feb. 2017. *Corresponding authors (XC.Zhang@cgiar.org; F.SanVicente@cgiar.org). Assigned to Associate Editor Nicholas Tinker.

Abbreviations: BLUP, best linear unbiased prediction; DH, doubled-haploid; DTMA, Drought Tolerant Maize for Africa; FDR, false discovery rate; GBS, genotyping-by-sequencing; GS, genomic selection; LD, linkage disequilibrium; LOD, logarithm of odds; MAF, minor allele frequency; MAS, marker-assisted selection; PCA, principle component analysis; PVE, phenotypic variation explained; QTL, quantitative trait loci; r_{MG}, genomic prediction accuracy; SNP, single-nucleotide polymorphism; TSC, tar spot complex.

Published in Plant Genome
Volume 10. doi: 10.3835/plantgenome2016.10.0099

early in the growing season may cause up to 75% grain yield loss (Hock et al., 1989; Pereyda-Hernández et al., 2009; ProMED-mail, 2009). Tar spot complex is caused by an interaction of at least three fungal species: *Phyllachora maydis* Maubl.; *Monographella maydis* E. Müll. & Samuels; and *Coniothyrium phyllachorae* Maubl. (Hock et al., 1992). Although the interaction mechanism between the host and the three pathogens is still not clear, *P. maydis* and *M. maydis* in the complex seem to be most detrimental (Hock et al., 1995). Although it is likely that the disease was present in Central America for centuries, where it has coevolved with maize, TSC was first reported in Mexico only in the beginning of last century and was later detected and described in other Latin American countries (Castaño, 1969; Liu, 1973; Maublanc, 1904; Shurtleff, 1982). In 2015, *P. maydis* was reported for the first time in Indiana and Illinois, although the presence of TSC has not been confirmed (Ruhl et al., 2016).

The use of resistant varieties is one of the most cost-effective and environmentally friendly approaches for controlling the disease (Ceballos and Deutsch, 1992; Mahuku et al., 2016). Dissecting the genetic architecture of TSC resistance with genome-wide molecular markers will allow breeders to improve their breeding efficiency by facilitating the introgression of the resistance genes into susceptible germplasm using MAS or GS strategies. However, only one study has been conducted to determine the genetic basis of TSC resistance in maize using a joint linkage association-mapping approach, and in this study, a major QTL (*qRtsc8-1*) associated with TSC resistance was identified (Mahuku et al., 2016). To develop functional molecular markers closely linked to TSC resistance genes for use in MAS, the genetic architecture of TSC resistance in maize needs to be characterized more comprehensively and with higher resolution.

Linkage mapping and association mapping are the two most commonly used tools for dissecting the genetics of complex traits in plants. Traditional linkage mapping explores the recombination events and marker–trait associations in biparental segregating populations, such as F_2, DH, and recombinant inbred lines. The method is very powerful in capturing major genes with larger effects and rare alleles (Ding et al., 2008; Raman et al., 2013; Trachsel et al., 2016; Wu et al., 2007). However, the resulting mapping resolution is comparatively low and typically produces a large confidence interval of 10 to 20 cM as a result of the limited number of recombination events occurring in the construction of biparental mapping populations (Li et al., 2010; Manichaikul et al., 2006; Zhu et al., 2008). On the other hand, association mapping explores functional variations within genetically diverse panels through linkage disequilibrium (LD) analysis, which is very efficient and effective for identifying new genes or confirming candidate genes (Aoun et al., 2016; Yu and Edward, 2006). Compared with linkage mapping, association mapping is able to identify historical recombination events in a given population, which improves mapping resolution. At the same time, association mapping may detect a high number of false positive associations and is less efficient at identifying rare alleles (Zhu et al., 2005). The combined use of linkage mapping and association mapping offers the opportunity to reduce false positives while increasing statistical power and improving mapping resolution (Bardol et al., 2013; Lu et al., 2010; Xu et al., 2012). This approach is being used in a wide range of plants to study the genetic basis of complex traits including resistance to several diseases of maize (Li et al., 2016a; Mahuku et al., 2016; Kump et al., 2011).

As an alternative to MAS, GS is now being used increasingly in several crops to improve breeding efficiency and increase genetic gains. Genomic selection uses genome-wide markers to predict the breeding values of individuals by capturing the effects of both major and minor genes (Meuwissen et al., 2001). Genomic selection has been shown to be effective in several crops over a wide range of marker densities, trait complexities, and breeding populations (Poland et al., 2012; Spindel et al., 2015; Zhao et al., 2012) where varying levels of prediction accuracy have been achieved. The main factors affecting genomic prediction accuracy are considered to be the relationship between the training and selection populations, training population sizes, population structure of training and testing sets, marker densities, genetic architecture and heritability of target traits, genotype $\times$ environment interactions, and statistical methods (Crossa et al., 2014; Combs and Bernardo, 2013; Endelman, 2011; Ornella et al., 2014; Zhang et al., 2015; Zhao et al., 2012). Taking genetic architecture information of the target traits into account, it may be possible to improve genomic prediction accuracy while implementing GS (Bernardo, 2014; Spindel et al., 2015). Combined linkage mapping and association mapping in conjunction with GS may be an effective tool to increase breeding efficiency by understanding the genetic architecture of complex traits like TSC in maize.

Increasing maker density has been reported to be an important factor for increasing statistical power and improving mapping resolution during association mapping and joint linkage-association mapping (Li et al., 2013), where millions of markers are required to cover all possible recombination events in a broad-based maize association panel with low LD decay. For example, LD decay in maize landraces is <1 kb (Tenaillon et al., 2001). Genotyping-by-Sequencing has been shown to be an economical genotyping alternative for increasing marker density. Romay et al. (2013) first reported that GBS was able to increase the statistical power and improve mapping resolution of association mapping by detecting rare alleles at high confidence levels in a panel of 2815 maize inbred accessions preserved at the US national maize inbred seed bank genotyped with GBS. Li et al. (2015) investigated the impact of GBS in a joint linkage analysis using multiple maize segregating populations and found this approach dramatically improved resolution and statistical power. If GBS technology becomes more affordable, it may also be promising for GS applications within applied breeding

programs. Good prediction accuracies have been obtained in several crops, when GBS SNPs were used to predict various target traits in a wide range of breeding populations (Gowda et al., 2015; Poland et al., 2012).

In the current study, we combined linkage mapping in three biparental DH populations with association mapping in conjunction with GS using ~1 million GBS SNPs distributed across the genome. The main objectives of this study were to (i) elucidate the genetic architecture of TSC resistance in maize with genome-wide GBS SNPs, (ii) identify major QTL and candidate genes conferring resistance to TSC using the association-mapping approach validated with linkage mapping, (iii) explore the potential of GS for improving TSC resistance in maize, and (iv) investigate the effects of training population size and marker density on genomic prediction accuracy in broad-based maize association panel and biparental DH populations.

Materials and Methods

Plant Materials

An association-mapping panel, designated Drought Tolerant Maize for Africa (DTMA), was used to perform genome-wide association analysis and genomic prediction in the present study. The DTMA panel consists of 282 tropical and subtropical inbred lines derived by the CIMMYT. These lines originate from different breeding programs of CIMMYT and comprise several lines with tolerance or resistance to an array of abiotic and biotic stresses affecting maize in the tropics, improved N-use efficiency, and grain nutritional quality (Wen et al., 2011).

Three biparental DH populations were used to perform linkage-mapping analysis and genomic prediction. The first population (Pop1) was derived from an F_1 cross between inbred lines CML495 and La Posta Sequia C7 F64-2-6-2-2-B-B-B. The second population (Pop2) was developed from an F_1 cross between CML451 and DTPYC9-F46-1-2-1-2-B-B-B. The third population (Pop3) shared the parental line CML451 with Pop2 and was crossed to DTPYC9-F74-1-1-1-1-B-B-B-B, a sister line of the second parental line of Pop2. In these three populations, CML495 and CML451 are widely used CIMMYT maize lines showing good resistance to TSC. The other parental lines are drought-tolerant or drought- and heat-stress tolerant lines (Cairns et al., 2013), which are susceptible to TSC. In total, 174, 100, and 111 DH lines were developed for Pop1, Pop2 and Pop3, respectively, using the protocols described by Prasanna et al. (2012).

Experimental Design and Phenotypic Evaluation

The DTMA panel was evaluated for responses to TSC under consistently high natural disease pressure at five locations in CIMMYT experimental stations in Mexico, namely, in Agua Fria (20°28′ N, 97°38′ W; mega-environment is lowland tropical) in 2009, 2011, and 2012; in Guerrero (17°02′ N, 99°38′ W; mega-environment is lowland tropical) in 2012; and in Veracruz (19°15′ N, 96°12′ W; mega-environment is lowland tropical) in 2012. Pop1 was evaluated for response to TSC at three locations, that is, in Agua Fria in 2011 and 2014 and in Guerrero in 2013. Pop2 was evaluated for response to TSC four times, that is, in Agua Fria in 2012 and 2014; each year had two planting dates. Pop3 was evaluated for response to TSC for three times, that is, in Agua Fria in 2012 with two planting dates and in Agua Fria in 2014.

A randomized complete block design was used for all experiments with three replications per location. Each plot consisted of a single 2-m row with 10 plants per row. Disease evaluation was performed as described by Mahuku et al. (2016). Briefly, disease scores were first taken at 2 wk after flowering and repeated two times at 7-d intervals. Disease severity was recorded using a 1-to-5 scale with a 0.5 increment, where 1 indicates highly resistant (HR), no visible disease symptoms or lesions identifiable on any of the leaves; 2 indicates resistant (R), moderate lesion development below the leaf subtending the ear or disease symptoms covering ~30% of the leaf area; 3 indicates moderately susceptible (MS), heavy lesion development on and below the leaf subtending the ear and a few lesions above it or 50% of the leaf surface have disease symptoms; 4 indicates susceptible (S), many or severe lesions on all but the uppermost leaves, which may have a few lesions, lesions have coalesced and blighted, or 70% of leaf surface has disease symptoms; and 5 indicates highly susceptible (HS), all leaves are dead, no green leaf tissue remaining or disease symptoms on >80% of the leaf surface. Figures showing the symptoms and development of TSC on maize leaves and cobs was also provided in the reference of Mahuku et al. (2016). For each plot, the average score across the three evaluations was used for further analyses.

Phenotypic Data Analysis

META-R software (http://hdl.handle.net/11529/10201) was used to analyze multilocation trials using a mixed linear model where all factors, including genotype refer to samples, environment refers to a combination of years and locations, replications in environments, interaction between genotype and environment, and interaction between replication and environment, were considered as random effects. Best linear unbiased prediction (BLUP) value of genotypes, variance components, and broad-sense heritability were obtained. A previous study (Mahuku et al., 2016) indicated that response to TSC was negatively correlated with maturity. Thus, anthesis date (data not shown) was used as a covariate to correct maturity effect and calculate the final TSC disease response BLUP. Broad-sense heritability of the target trait was calculated as the ratio of total genetic to total phenotypic variance. In multilocation trial analysis, broad-sense heritability was calculated as $h^2 = \sigma_g^2 / \left(\sigma_g^2 + \sigma_{ge}^2 / e + \sigma_e^2 / er\right)$, where σ_g^2, σ_{ge}^2, and σ_e^2 are the genotypic, genotype × environment interaction, and error variance components, respectively, and e and r are the number of environments and of replicates within

each environment included in the corresponding analysis, respectively. Descriptive statistics were conducted in Excel 2010 (Microsoft). Phenotypic data distribution of the four tested populations and the subgroups of association-mapping panel was showed in a violin plot, which was generated using violin plot package in R software (R Development Core Team, 2013).

Genotyping, Single-Nucleotide Polymorphism Calling, Imputation, and Filter

Total genomic DNA was extracted from bulked young leaves for all inbred lines using a CTAB procedure (CIMMYT Applied Molecular Genetics Laboratory, 2003). Genotyping was performed at Cornell University Biotechnology Resource Center (Ithaca, NY). Genomic DNA was digested with the restriction enzyme *ApeK1*. Genotyping-by-sequencing libraries were constructed in 96-plex and sequenced on Illumina HiSeq2000 (Elshire et al., 2011). Single-nucleotide polymorphism calling was performed using the TASSEL GBS Pipeline, where the GBS 2.7 TOPM (tags on physical map) file downloaded from Panzea (www.panzea.org) was used to anchor reads to the Maize B73 RefGen_v2 reference genome (Glaubitz et al., 2014). Imputation was performed with FILLIN method in TASSEL 5.0 (Swarts et al., 2014), and the donor haplotype files required by FILLIN were downloaded from Panzea (www.panzea.org) and consisted of the anonymized GBS 2.7 haplotypes made from 8000-site windows. The parameters for running FILLIN to do imputation were set as the default values, which have been descripted in detail by Swarts et al. (2014). An imputed GBS dataset was used to conduct genome-wide association analysis and genomic prediction in the DTMA association-mapping panel, while unimputed datasets were used for further analyses in the biparental populations. For each inbred line, 955,690 SNPs that are evenly distributed on the maize chromosomes were called; 955,120 of these were assigned to a maize chromosome, and 570 of them could not be anchored to any of the 10 maize chromosomes. TASSEL V5.0 (Bradbury et al., 2007) was used to filter raw GBS datasets for SNPs with minor allele frequency (MAF) >0.05 and missing data rates <20% in the DTMA panel and in each biparental population, respectively. Samples with heterozygosity >0.05 were excluded from the SNP characteristic analyses. Basic genotypic information, including number of SNPs, MAF, missing rate, and heterozygosity rate, was calculated at the population level.

Association Mapping

After filter, 261,055 high-quality SNPs were obtained in the DTMA panel for genome-wide association analysis. Linkage disequilibrium analysis was conducted in TASSEL V5.0 (Bradbury et al., 2007) and the average LD decay distance estimated was 3.5 kb ($r^2 = 0.1$) across the 10 maize chromosomes. Population structure was estimated using an admixture model-based clustering method implemented through the software Structure V2.3.3 (Hubisz et al., 2009), where a subset of 10,000 SNPs with no missing values were randomly selected across the 10 maize chromosomes. Hypotheses were tested for subpopulation number K ranging from 1 to 10, and each K was run seven times with burn-in time and replications both to 100,000. A mixed linear model (principle component analysis [PCA] + K) was applied for association-mapping analysis in GAPIT (Genome Association and Prediction Integrated Tool-R package), where a kinship matrix along with the three principle components were incorporated to avoid spurious associations. Using GAPIT, the kinship matrix was generated automatically with the default Vanraden algorithm (VanRaden, 2008), and PCA was also calculated. The P-value of each SNP was calculated and a false discovery rate (FDR) corrected threshold was used to declare significance with a uniform value of $P = 2.97 \times 10^{-5}$, that is, $-\log(P) = 4.53$. Manhattan and quantile–quantile plots were created in R package qqman using the association-mapping results (Turner 2014).

Candidate Gene Analysis

Based on the association-mapping results, the sequence information of each significantly associated SNP was used to perform BLAST against the B73 RefGen_v2 genome sequence through the MaizeGDB database (http://www.maizegdb.org/), and genes containing the significantly associated SNPs were considered as possible candidate genes for TSC resistance. Candidate gene annotation was performed on PlantGDB (http://www.plantgdb.org/ZmGDB/) and Gramene (http://www.gramene.org/).

Linkage Map Construction and Quantitative Trait Loci Mapping

For each of the three DH populations, a bin map was constructed with high-quality unimputed SNPs using customized R scripts (unpublished data, 2017). To reduce genotyping error and eliminate the low-quality SNPs from the bin map, the following steps were performed: (i) unimputed SNP datasets were filtered with the parameters of MAF > 0.05 and missing rate <20%; (ii) DH lines with heterozygosity rate >5% or missing rate >20% were eliminated from the further analysis; (iii) unlinked SNPs were removed from further analysis, where the window size was eight, similarity rates of all the SNPs within each window were calculated to remove the unlinked SNPs, threshold of similarity rate was 95%; (iv) the consecutive SNPs with high similarity rate (95%) were merged into one bin; and (v) bins were treated as genetic markers to construct a genetic map. In Pop1, Pop2, and Pop3, 437, 494, and 493 bins were constructed with 20,473; 27,818; and 32,607 SNPs, respectively. The genetic maps of Pop1, Pop2, and Pop3 were 987.35, 1150.16, and 1153.99 cM, respectively, resulting in an average marker (bin) density of 2.26, 2.35, and 2.97 cM, respectively. In software QTL IciMapping Version 4.1 (www.isbreeding.net), the MAP function was implemented to construct a linkage map using nnTwoOpt algorithm and the criterion of logarithm of odds (LOD) threshold >3.0. For each individual DH population, QTL

Table 1. Descriptive statistics of phenotypic response to tar spot complex infection in the Drought Tolerant Maize for Africa panel and three biparental doubled-haploid populations.

Population	Mean	Min.	Max.	Median	SD	Skewness	Kurtosis	Heritability
DTMA	2.04	1.31	3.23	2.00	0.45	0.55	0.06	0.80
Pop1	2.33	1.81	3.00	2.30	0.23	0.49	−0.07	0.54
Pop2	2.14	1.18	3.95	2.04	0.54	0.98	1.02	0.88
Pop3	2.50	1.37	4.39	2.29	0.68	0.68	−0.36	0.93

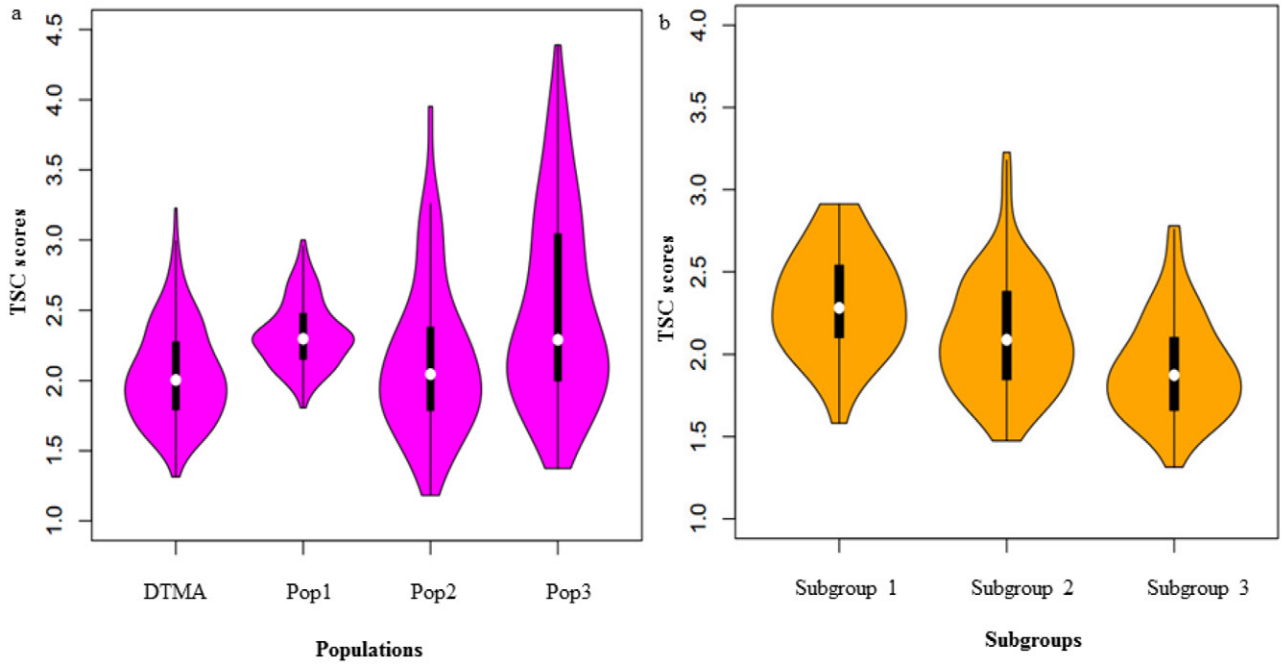

Fig. 1. Violin plots of the tar spot complex (TSC) scores in the (a) Drought Tolerant Maize for Africa (DTMA) association-mapping panel, Pop1, Pop2, and Pop3 and in (b) subgroups of the DTMA association-mapping panel. The black bars inside the plot represent the first and third quartiles, and white dots represent the median. The width of the plot represents probability density of the data at different values.

mapping was implemented in QTL IciMapping Version 4.1 using the BIP function. The scanning step was 1 cM, the largest *P*-value for entering variables in stepwise regression of phenotype on marker variables (PIN) was 0.001, and the largest *P*-value for removing variables was 2× PIN. One-dimensional QTL scanning was conducted, that is, only additive effects for each QTL was estimated. The LOD threshold used for declaring putative QTL was determined as 3.0, and the phenotypic variation explained (PVE) by each QTL was estimated.

Genome Prediction

Genomic prediction was implemented in rrBLUP package (Endelman, 2011) in the association-mapping panel and in each of the three DH populations. A subset of 10,000 SNPs distributed uniformly across the genome, with no missing values and MAF > 0.05, was used for genomic prediction in the DTMA panel. In each of the three DH populations, SNPs in the genetic map were used for genomic prediction. Details of the implementation of rrBLUP were described by Zhao et al. (2012). A five-fold cross-validation scheme with 100 replications was used to generate the training and validation sets and assess the prediction accuracy within each population. The average value of the correlations between the phenotype and the genomic estimated breeding values was defined as genomic prediction accuracy (r_{MG}). To test the effects of training population size and marker density on genomic prediction accuracy, training populations ranging from 10 to 90% of the total population size and number of SNPs varying from 200 to 10,000 were used in the DTMA panel to assess the prediction accuracy. In the three DH populations, the training population size ranged from 10 to 90% of the total population size, with an interval of 10%, and the number of SNPs varied from 200 to 20,000. For all the genomic prediction analyses, the training dataset and validation dataset were independent.

Results

Phenotypic Variation

Phenotypic variation for TSC resistance in the DTMA panel and the three DH populations are described in Table 1 and Fig. 1a. Disease scores varied across populations and the mean values in the four populations were similar, ranging from 2.04 in DTMA panel to 2.50 in Pop3. Sufficient differences were observed between the minimum and maximum scores in all the populations;

the difference between the minimum and maximum scores ranged from 1.19 in Pop1 to 3.02 in Pop3. In the DTMA association-mapping panel, individual location analysis showed that the minimum disease score was 1.25, 1.12, 1.24, 1.32, and 1.28, respectively, in Agua Fria in 2009, 2011, and 2012; in Guerrero in 2012; and in Veracruz in 2012. The maximum disease scores were 3.94, 3.03, 4.32, 2.59, and 3.81, respectively. There was adequate natural expression of the diseases to differentiate resistant and susceptible lines in all the phenotyping locations except for Guerrero in 2012, where the lowest degree of phenotypic variation was observed. For the three DH populations, the parental lines were used as controls to check for adequate levels of disease infection, and the combined analysis across all locations showed that the disease scores of the resistant parental lines were 2.11, 1.48, and 1.42 in Pop1, Pop2, and Pop3, respectively, and the disease scores of the susceptible parental line were 2.95, 3.92, and 3.20, respectively. In the individual location analysis, the disease scores of the susceptible parental line reached 4.89, 4.93, and 4.44 in Pop1, Pop2, and Pop3, respectively. This indicated that adequate infection was obtained under natural inoculation conditions. Broad-sense heritability estimated from the combined analysis was 0.80, 0.54, 0.88, and 0.93 in the DTMA panel, Pop1, Pop2, and Pop3, respectively. High heritability observed in all the populations indicating that the phenotypic data was reliable for further genetic analyses.

Single-Nucleotide Polymorphism Characteristics in Different Genotyping-by-Sequencing Datasets

To compare the differences between imputed and unimputed GBS datasets, basic genotypic information of each population, including number of SNPs, minor allele frequency, missing rate, and heterozygosity rate, is presented in Table 2 and Fig. 2. The total number of SNPs in the imputed dataset in the DTMA panel decreased from 955,120 to 261,055 after filter, while that in the unimputed dataset decreased from 955,120 to 44,146; 47,203; and 45,775 in Pop1, Pop2, and Pop3, respectively. After filtering, different number of SNPs detected between DTMA panel and DH populations reflected the greater genetic diversity in the DTMA panel, and the higher missing rate in the unimputed dataset in the DH populations. The missing rate before filtering in the imputed dataset of the DTMA panel was much lower than that in the unimputed dataset in DH populations, and the missing rate was 15.12% in the DTMA panel, while it ranged from 42.32 to 43.57% in DH populations. However, the missing rate after filtering in all populations was similar, ranging from 6.36% in DTMA panel to 7.90% in Pop2. In both the imputed and unimputed datasets, the heterozygosity rate increased after filtering.

Before filtering, 58.75% of the SNPs had an MAF < 0.05, and the average MAF in the imputed dataset of the DTMA panel was 0.09. In the unimputed dataset of Pop1, 85.06% of the SNPs had an MAF lower than 0.05, and the average MAF was 0.04. After filtering, the average MAF was 0.23, with continuous distribution classes from 0.05 to 0.50 at intervals of 0.05 in the imputed dataset of DTMA panel. In the unimputed dataset of Pop1, the average MAF was 0.42 and 79.74% of the SNPs concentrated to the MAF ranging from 0.40 to 0.50, similar to the theoretical allele frequency of 0.50 expected in each DH population. The distribution of MAFs in Pop2 and Pop3 had a similar tendency to that recorded in Pop1 (data not shown).

Table 2. Single-nucleotide polymorphism (SNP) characteristic information of the Drought Tolerant Maize for Africa panel and three biparental doubled-haploid populations before and after filter of genotyping-by-sequencing datasets.

Population	No. of SNPs		Missing rate		Heterozygosity rate	
	Before filter	After filter	Before filter	After filter	Before filter	After filter
			%			
DTMA	955120	261055	15.12	6.36	1.53	1.86
Pop1	955120	44146	42.62	7.68	0.51	2.32
Pop2	955120	47203	42.32	7.90	0.47	2.55
Pop3	955120	45775	43.57	7.60	0.33	2.13

Population Structure Analysis of the Association Mapping Panel

The DTMA panel was divided into three subgroups according to the ad hoc statistic delta K (ΔK) value changes. The numbers of lines were 40, 111, and 131 in Subgroup 1, 2, and 3, respectively. Most of the lines in Subgroup 1 were from the Mexico physiology research group, while lines in Subgroup 2 were mainly from the subtropical breeding program, and lines in Subgroup 3 were mainly from the lowland tropical breeding program. Violin plots of the three subpopulations were provided to show the different responses of the three subgroups to the disease. The mean disease score of the three subpopulations was 2.31, 2.12, and 1.91 in Subgroup 1, 2, and 3, respectively (Fig. 1b).

Genetic Architecture Revealed by Association Mapping

Association mapping results of the DTMA panel are shown in Fig. 3 and Table 3. In total, 155 SNP markers with significant marker–trait associations were detected, with the multiple testing correction threshold of $-\log10(P) > 4.53$ (FDR level of 0.05). These significantly associated SNPs were distributed in four genomic regions in bins 2.05, 3.04 to 3.09, 7.02, and 8.03. Each of these SNPs explained 4.76 to 13% of the total phenotypic variance individually and together explained 52.60% of the total phenotypic variation. The number of significantly associated SNPs detected on chromosomes 2, 3, and 7 was 1, 3, and 1, respectively. The rest of the 150 significantly associated SNPs were located on chromosome 8 (bin 8.03) in a ~33.6 million bp region based on the B73 reference genome (Fig. 3c, 3d) flanked by left and right significant SNPs of S8_57248413 and S8_90835374, respectively. Significantly associated SNPs on chromosome 2, 3, and

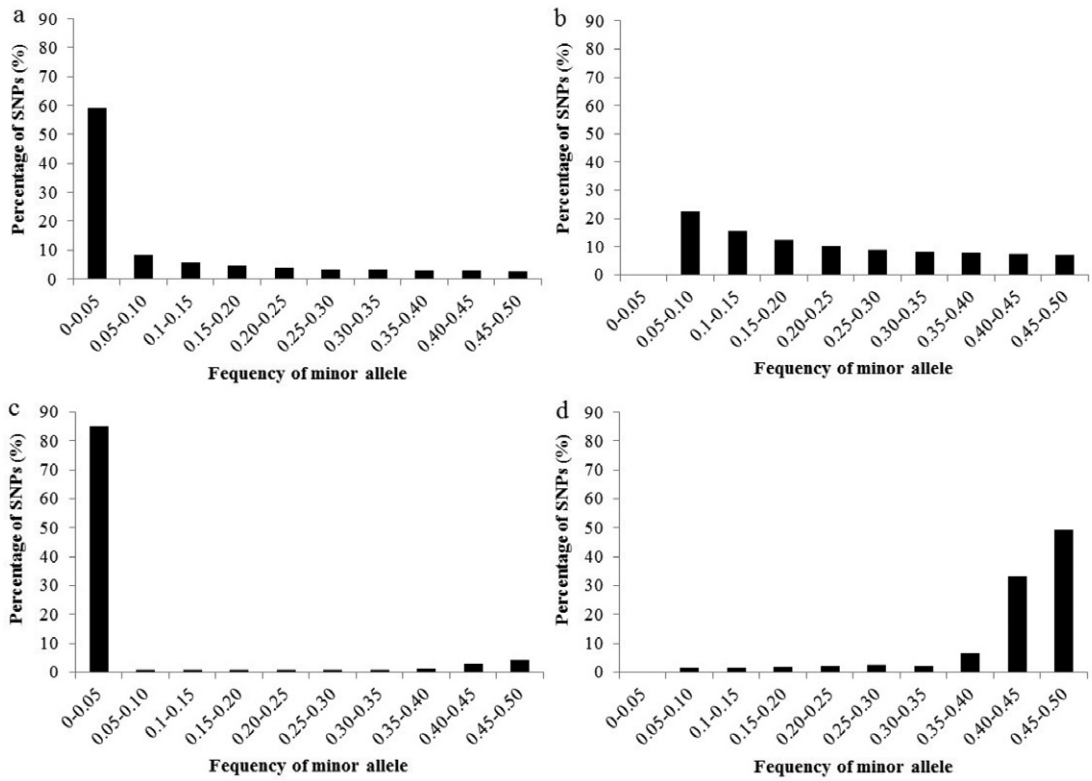

Fig. 2. Frequency distribution of minor alleles in the Drought Tolerant Maize for Africa (DTMA) panel and Pop1 for (a) DTMA panel before filter, (b) DTMA panel after filter, (c) Pop1 before filter, and (d) Pop1 after filter.

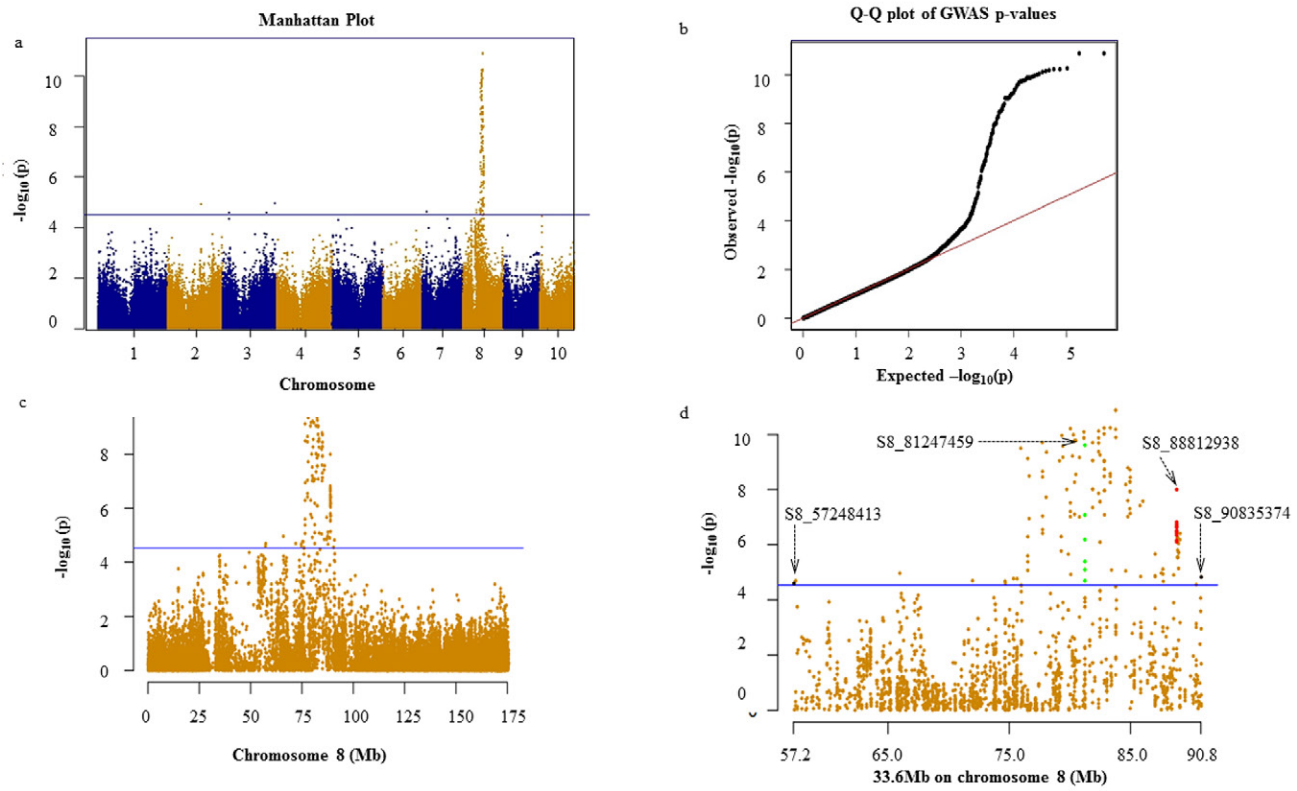

Fig. 3. Association mapping for tar spot complex (TSC) resistance in the Drought Tolerant Maize for Africa (DTMA) panel: (a) Manhattan plot, the *y*-axis represents the *P*-value of the marker–trait association on a $-\log_{10}$ scale. The horizontal line indicates the genome-wide significance threshold false discovery rate (FDR = 0.05); (b) quantile–quantile plot, the *x*-axis represents the expected $-\log_{10}$ (*p*), *y*-axis represents the observed $-\log_{10}$ (*p*); (c) association-mapping results of chromosome 8; (d) association-mapping results of the significant associated region on chromosome 8.

Table 3. Significantly associated single-nucleotide polymorphisms (SNPs) revealed by association-mapping analysis in the Drought Tolerant Maize for Africa panel on chromosomes 2, 3, and 7, with the top five significantly associated SNPs on chromosome 8 listed.

SNP name†	Allele	*P*-value	PVE‡	MAF§	Candidate gene	Annotation
			%			
S2_143450705	A,T	1.28×10^{-5}	5.14	0.40	–	–
S3_187460281	A,T	2.63×10^{-5}	4.76	0.30	GRMZM2G072853	–
S3_224167844	A,C	1.17×10^{-5}	5.19	0.15	GRMZM2G339540	Leucine-rich repeat receptor-like protein kinase family protein
S3_28621298	A,G	2.57×10^{-5}	4.77	0.44	GRMZM2G180622	Sarcoplasmic reticulum histidine-rich calcium-binding protein
S7_21279210	A,G	2.41×10^{-5}	4.80	0.18	GRMZM2G063340	Rps12 ribosomal proteins12
S8_83777799	A,G	1.29×10^{-11}	13.00	0.25	–	–
S8_83777830	C,G	1.33×10^{-11}	12.98	0.27	–	–
S8_83335753	A,T	5.73×10^{-11}	12.11	0.23	GRMZM2G104025	60S ribosomal protein L18–3-like
S8_82826985	A,G	5.83×10^{-11}	12.1	0.24	–	–
S8_80045991	G,T	6.11×10^{-11}	12.07	0.23	GRMZM2G106526	ATP10 protein

† SNP name, chromosome_position, for example, S2_14345070 means the SNP is on chromosome 2, the physical position is 143450705 bp.

‡ PVE, phenotypic variation explained.

§ MAF, minor allele frequency.

7, and the top five SNPs on chromosome 8 are listed in Table 3. The most significant SNP S8_83777799 (*P*-value 1.29×10^{-11}) was detected on chromosome 8 at position 83,777,799 in the B73 reference genome; this SNP alone explained 13% of the total phenotypic variation. Thus, association-mapping results revealed that TSC resistance is controlled by a major QTL on chromosome 8 with several minor QTL located on other chromosomes.

In total, 46 putative candidate genes were identified among the 155 significantly associated SNPs, 28 of which were hypothetical genes with known predicted functions referring to 71 SNPs in total, and the rest of them were putative uncharacterized proteins. For the 28 putative candidate genes with known predicted functions, the number of SNPs detected for each candidate gene ranged from one to 12, and 2.54 SNPs per gene were identified on average (Supplemental Table S1). Some candidate genes contain more than one significant SNP, and multiple SNPs identified in candidate genes is valuable for developing haplotypes for implementing MAS. Functional annotation of the 28 hypothetical genes showed that several of them are associated with disease resistance or defense responses in plants. For instance, GRMZM2G062974 was annotated as an endochitinase that plays a significant role in resistance to plant diseases. Twelve SNPs linked with endochitinase were significantly associated with TSC resistance in association-mapping analysis (Fig. 3d) including the most significant associated SNP S8_88812938 (chromosome_position). Several hypothetical genes with protein kinase function, such as leucine-rich repeat receptor-like protein kinase (GRMZM2G073884, GRMZM2G339540), protein NSP-INTERACTING KINASE 3-like (GRMZM5G835516), MAP kinase family protein (GRMZM2G063144), and protein kinase superfamily protein (GRMZM2G117465) were also detected.

Quantitative Trait Loci Detected by Linkage Mapping

Quantitative trait loci detected by linkage mapping in all three DH populations are summarized in Table 4. In Pop1, three QTL were detected in bins 4.08 to 4.09, 7.03, 8.03, explaining 8.63, 9.19, and 13.18% of the total phenotypic variation, respectively. In Pop2, four QTL were detected in bins 1.05, 2.04, 5.04 to 5.05, and 8.03 accounting for 17.82, 6.19, 5.12, and 43.31% of the total phenotypic variation, respectively. In Pop3, four QTL were detected in bins 1.06, 7.03, 8.03, and 10.07 explaining 32.29, 5.96, 16.78, and 6.01% of the phenotypic variation, respectively.

A major QTL located in bin 8.03 was detected in all the three DH populations and had the largest LOD score and PVE relative to the other detected QTL, with the only exception in Pop3. This was consistent with the association-mapping results, where a major QTL located on bin 8.03 was identified as well. In addition to the major QTL on bin 8.03, other QTL detected by association mapping were also confirmed by linkage mapping in different genetic backgrounds, with the only exception of the QTL on chromosome 3 (bins 3.04, 3.06, and 3.09). The QTL located in bin 2.05 was validated by linkage mapping in Pop2 in a neighboring bin 2.04, and the QTL located in bin 7.02 was validated by linkage mapping both in Pop1 and Pop3 in the neighboring bin 7.03. Additional unique QTL were detected by linkage mapping, which were not identified via association mapping, that is, QTL in bins 1.05 to 1.06, 4.08 to 4.09, 5.04 to 5.05, and 10.07. Thus, integration of linkage mapping and association mapping proved to be a powerful tool for increasing statistical power, and improving the mapping resolution.

Genomic Prediction Accuracy

Genomic prediction accuracies (r_{MG}) obtained from five-fold cross-validations, and 100 replications for all the four populations are shown in Fig. 4, where the prediction accuracies were moderate to high and varied across

Table 4. Linkage-mapping analysis results of tar spot complex resistance in three biparental doubled-haploid populations.

Population	Chromosome	Left marker (position)†	Right marker (Position)	LOD‡	PVE§	Additive effect
					%	
Pop1	4	C4M190 (197840452)	C4M191 (209059831)	5.82	8.63	0.07
	7	C7M315 (144585946)	C7M316 (149528489)	9.19	14.41	0.09
	8	C8M347 (70713777)	C8M348 (82228021)	13.18	21.85	0.12
Pop2	1	C1M30 (92194874)	C1M31 (100468505)	9.07	17.82	0.19
	2	C2M95 (36770127)	C2M96 (45491073)	3.59	6.19	0.11
	5	C5M266 (171692383)	C5M267 (174071529)	3.06	5.12	0.10
	8	C8M390 (46943754)	C8M391 (83381560)	17.88	43.31	0.30
Pop3	1	C1M38 (184932563)	C1M39 (188828594)	14.07	32.29	0.38
	7	C7M366 (129063233)	C7M367 (131806297)	3.25	5.96	0.17
	8	C8M399 (23160646)	C8M401 (38729344)	8.24	16.78	0.28
	10	C10M509 (147329967)	C10M511 (147901114)	3.24	6.01	0.17

† Marker name comprises information on specific chromosome and bin, for example, C4M190 means the 190th bin on chromosome 4 and the physical position is 197,840,452 bp.

‡ LOD, logarithm of odd.

§ PVE, phenotypic variation explained.

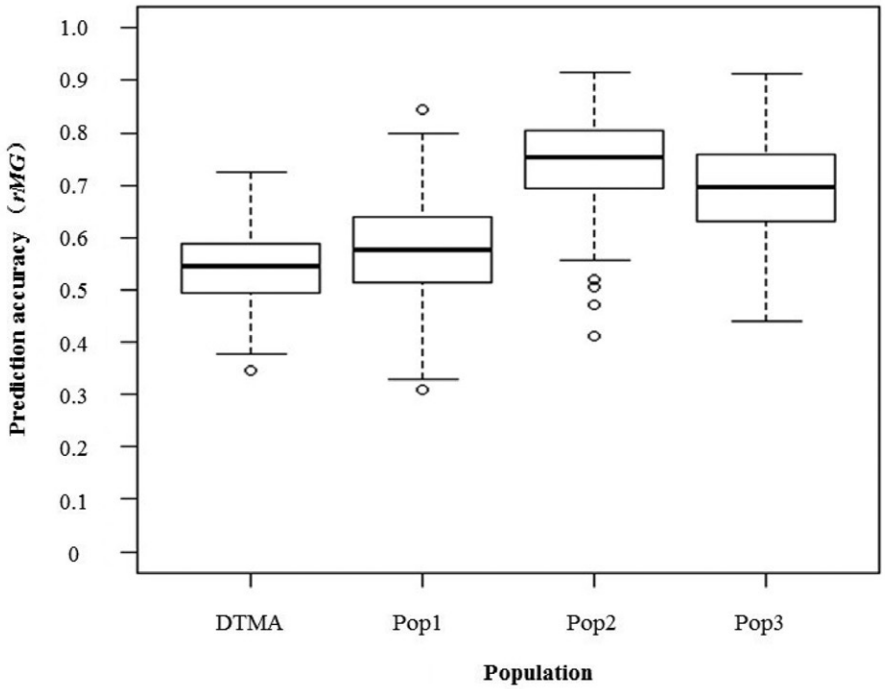

Fig. 4. Genomic prediction accuracy (r_{MG}) of tar spot complex (TSC) resistance in the Drought Tolerant Maize for Africa (DTMA) panel, Pop1, Pop2, and Pop3. Ten thousand randomly selected SNPs without missing data and minor allele frequencies (MAFs) >0.05 were used for prediction in the DTMA panel. Single-nucleotide polymorphisms in the genetic map were used for genomic prediction in the three biparental doubled-haploid populations.

populations. The lowest r_{MG} was observed in the DTMA panel with a mean value of 0.55, where 10,000 high-quality SNPs were used for prediction. The r_{MG} observed in Pop1, Pop2, and Pop3 were 0.58, 0.74, and 0.69, respectively. In Pop1, smaller phenotypic differences (1.19) and lower heritability (0.54) caused a moderate r_{MG} compared with the other two DH populations.

The effects of marker density and training population size on r_{MG} are presented in Fig. 5. The r_{MG} continuously increased as the number of SNPs increased in all the populations. In the DTMA panel, only a marginal increase was observed on r_{MG}, where the number of SNPs increased from 1000 to 10,000, and r_{MG} obtained with 261,948 high-quality SNPs was similar to that obtained from 10,000 randomly selected high-quality SNPs (data not shown). The maximum r_{MG} with minimum standard error was observed when 3000 SNPs were used for genomic prediction. In Pop1, a very slight increase in r_{MG} was observed when the number of SNPs increased from 500 to 20,000, indicating that 500 SNPs were sufficient to obtain good prediction accuracy of TSC resistance in biparental DH populations. The training population

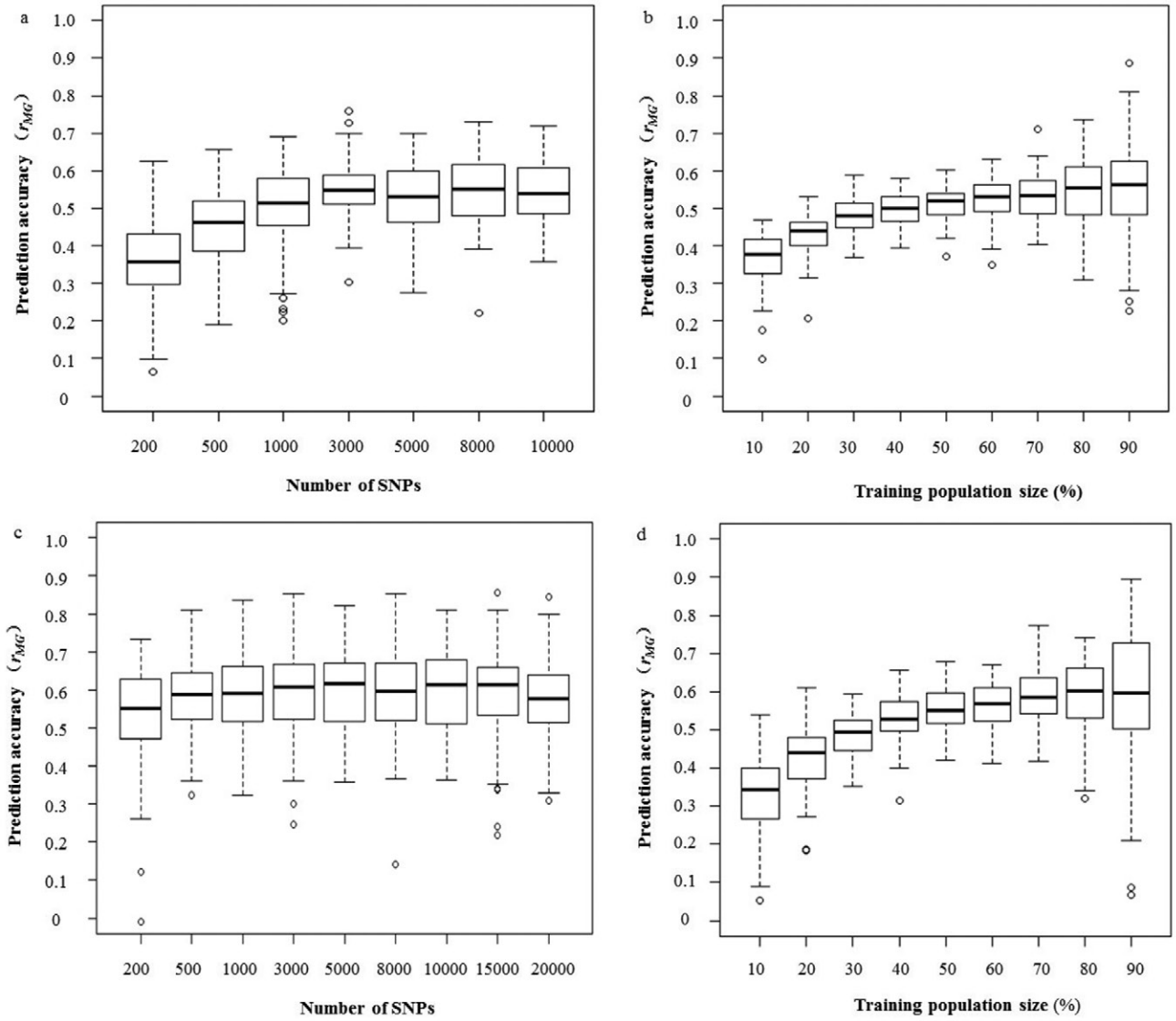

Fig. 5. Genomic prediction accuracy (r_{MG}) of tar spot complex (TSC) resistance in the Drought Tolerant Maize for Africa (DTMA) panel and Pop1 with different training population size and marker density: (a) DTMA panel, the number of single-nucleotide polymorphisms (SNPs) varied from 200 to 10000; (b) DTMA panel, training population size ranged from 10 to 90% of the total population size with an interval of 10%; (c) Pop1, the number of SNPs varied from 200 to 20000; (d) Pop1, training population size ranged from 10 to 90% of the total population size with an interval of 10%.

size affected the prediction accuracy considerably in all the populations, and the r_{MG} continuously increased as training population size increased in all the populations. In the DTMA panel, a slight increase was observed in r_{MG} when the training population size increased from 40 to 90%. A relatively high r_{MG} with the smallest standard error was observed when 60% of the total genotypes were used as a training population. A similar trend was also observed in Pop1, where relative high prediction accuracy with the smallest standard error was observed when 50% of the total genotypes were used as a training population. This indicated that the optimum size of training population was 50 to 60% of the total number of genotypes used. However, prediction accuracy can vary depending on the population size of total genotypes used and the heritability of the trait of interest.

Discussion

Combined linkage and association mapping can complement the strengths and weaknesses of each approach, and has been successfully used in several crops to dissect the genetic architecture of complex traits (Chen et al., 2014; Li et al., 2016b; Lu et al., 2010). In this study, combined linkage mapping and association mapping was implemented for detecting and validating genomic regions conferring resistance to TSC in maize. Genome-wide association analysis revealed four QTL on chromosomes 2, 3, 7, and 8. All the QTL revealed by association mapping were validated by linkage mapping in different genetic backgrounds, with the exception of a QTL on chromosome 3. A major QTL located on bin 8.03 was consistently detected in the association-mapping panel and all the three DH populations, and it exhibited the

highest PVE in Pop1 and Pop2. A QTL in bin 8.03 of Pop3 had the second largest PVE. The physical position of the major QTL detected by association mapping was between 57.2 and 90.8 Mb, while the position of the major QTL in bin 8.03 in Pop1, Pop2, and Pop3 was 70.7 to 82.2, 46.9 to 83.4, and 23.2 to 38.7 Mb, respectively. In addition to the QTL detected by both association mapping and linkage mapping, several QTL were identified through linkage mapping in different genetic backgrounds. Three QTL located on bin 4.08 to 4.09, 5.04 to 5.05, and 10.07 were identified in Pop1, Pop2, and Pop3, respectively. The QTL located in bin 1.05 to 1.06 was identified in two genetic backgrounds, that is, Pop2 and Pop3. These results reveal that combined linkage mapping and association mapping is a powerful approach to detect QTL associated with TSC resistance across the maize genome in different genetic backgrounds. Furthermore, it was also determined that TSC resistance is controlled by a major QTL located on bin 8.03 in addition to several minor QTL with smaller effect distributed on other chromosomes. The QTL information generated in this study will be valuable for fine mapping of the major gene in bin 8.03 and for developing functional molecular markers for implementing MAS for TSC resistance.

Compared with a previous study on TSC resistance (Mahuku et al., 2016), the populations and phenotypic data used in the current study were slightly different in both association and linkage-mapping analysis. In the association analysis, an unbalance dataset was used in the previous study, and association-mapping analysis was performed separately for seven environments. In the current study, a balance dataset from five environments was used, eliminating two environments with low heritability and small numbers of genotypes. Additional QTL were identified in the current study through the use of this improved phenotypic data as well as through the use of higher-density GBS SNPs. Mahuku et al. (2016) identified three QTL located in bins 2.07, 7.02, and 8.03 with an Illumina MaizeSNP50 BeadChip containing 56,110 evenly spaced SNPs. In the present study, we verified all the three previously reported QTL and also identified a new QTL on chromosome 3. Furthermore, the genetic region of the major QTL located on bin 8.03 was narrowed from 81 million to 33.6 million bp based on significantly associated SNPs. These results demonstrate that improved phenotypic data and higher density GBS SNPs can increase statistical power and improve mapping resolution of association mapping as compared with the lower-density chip-based SNP arrays.

In regard to biparental linkage analysis, the current linkage-mapping analysis used ultrahigh-density genetic maps constructed with GBS SNPs, which increased statistical power and detected more QTL with improved mapping resolution relative to the previous study of Mahuku et al. (2016). Linkage-mapping analysis in the current study validated all the QTL regions previously reported by Mahuku et al. (2016) with the exception of the QTL located on chromosome 6, which was not identified in this study. Three new QTL located on chromosomes 2, 4, and 5 were detected only in the current study. One reason for this difference is that the populations and phenotypic data used in these two studies were slightly different. Two of the DH populations used in the current study were identical to those used in the previous study, but the third DH population in the current study replaced the $F_{2:3}$ population from the previous work. In addition, the number of locations used for phenotyping was also different in these two studies. Populations were evaluated only in two, one, and one location in the previous study, while in the current study, the three DH populations were evaluated in three, four, and three locations, respectively. Improved phenotypic data with higher heritability (0.54 in Pop1, 0.88 in Pop2, and 0.93 in Pop3) resulting in the detection of more robust QTLs in the current study. Mapping resolution of the current study was also improved by using the ultrahigh-density genetic maps constructed with GBS SNPs, and the interval of the detected QTL (Table 4) ranged from 0.57 to 36.44 million bp with an average of 9.66 million bp (physical position difference between the right marker and left marker), which is smaller than the normal interval of the genetic map constructed with few hundred markers.

Genotyping-by-sequencing is a low-coverage sequencing technology generating a large number of SNPs at a lower genotyping cost per SNP per sample. However, GBS also results in a very high rate if missing data. Imputation of missing data is generally conducted before downstream analysis is performed. Recently, Wu et al. (2016) reported that the missing rate of an unimputed GBS dataset in a diverse tropical maize panel was 57.99% before filter, which indicates that the imputed GBS dataset is more appropriate for running association-mapping analysis in a diverse maize panel because of the high rate of missing data within unimputed GBS datasets. In the current study, the imputed GBS dataset was used for association mapping, and the missing rate in the diverse DTMA tropical maize panel was 15.17% before filter and 6.66% after filter. Statistical power and mapping resolution of association mapping was improved by including more imputed GBS SNPs. However, within the biparental mapping populations, incorrectly imputed SNPs could result in identification of false crossovers, which would adversely affect the accuracy of the genetic map. For this reason, unimputed GBS data was used for ultrahigh-density bin map construction and linkage-mapping analysis in all the three DH populations. Because an imputed dataset was used in association-mapping analysis and unimputed datasets were used for linkage-mapping analysis, combined linkage mapping and association mapping proceeded as a parallel approach rather than an integrated approach. The integrated approach of employing a common set of SNPs will be assessed in the further studies.

Factors affecting genomic prediction accuracy have been investigated in several crops over a wide range of target traits (Poland et al., 2012; Spindel et al., 2015;

Zhang et al., 2015). In the current study, genomic prediction accuracy for TSC in maize has been estimated in different types of populations with different training population sizes and marker densities. Genomic prediction accuracy observed in DH populations was always higher than that observed in the association-mapping panel, even though heritability of TSC in Pop1 was lower than that of the association-mapping panel. This suggests that the genomic prediction accuracy is relatively lower in a population with a broader genetic base, and novel statistical models are required to be developed to improve prediction accuracies in broad genetic base populations. Similarly, Gowda et al. (2015) found that the genomic prediction accuracy for improving resistance to maize lethal necrosis disease was also relatively lower in a broader genetic population. In actual application, however, maize breeding rarely proceeds using source populations as broad as the DTMA panel since maize breeding programs are generally focused on specific geographies. Within geographies, germplasm is focused on relative maturity targets, and source breeding populations are typically further restricted by organization of lines into discrete heterotic groups, often with important founder lines contributing substantially to breeding progress.

Appropriate training population sizes to maximize prediction accuracy while minimizing standard error were also determined for each population. In all populations, relatively high prediction accuracies with the smallest standard error were observed when 50 to 60% of the total genotypes were used as a training population. Currently, one of the most significant challenges for maize breeders is to select an optimum subset of lines or families using MAS or GS prior to multiple-location testing. Optimal training population size for running genomic prediction within a biparental DH population is also of interest. Results from this study indicate that good prediction accuracy of TSC resistance can be achieved by phenotyping and genotyping as little as half of the breeding material as the training set. This confirms the results of Crossa et al. (2014) showing that a good prediction of maize grain yield in a biparental population can be achieved when only half of the population is included in the training set. However, the prediction accuracy could vary depending on the population sizes used and the target traits.

The relationship between the training population and selection population is an important factor affecting prediction accuracy. In this study, prediction accuracy was only assessed within each population, which is the most favorable situation for running GS since the training and prediction sets are closely related. However, most breeders also would like to know the prediction accuracy across populations or use pooling of multiple populations into the training set to predict the rest. This will be assessed in the further studies.

Conclusion

The majority of the presently grown commercial maize varieties in Central and South America are susceptible to TSC. Developing and deploying improved maize varieties with resistance to TSC is important, as it provides the most cost-effective approach for controlling the spread and impact of TSC in the maize production areas in the Americas. In the current study, the genetic architecture of TSC resistance in maize was dissected through combined linkage and association mapping in conjunction with high density GBS SNPs. Results indicate that TSC resistance in maize is controlled by a major QTL located on bin 8.03 with several minor QTL with smaller effect on other chromosomes. This study is the first report on TSC resistance in maize using sequencing-based genotyping, which increases the statistical power by detecting more QTL and improves the mapping resolution of detected QTL. This study revealed SNPs that are significantly associated with TSC resistance genes. The sequence information of these SNPs can be used to develop assays for MAS, and can be fitted as fixed effects in GS models to improve prediction accuracy (Bernardo, 2014). Tar spot complex resistance in tropical maize could be improved by implementing MAS and GS individually or by implementing them in a stepwise fashion. The decision of a breeding strategy to implement MAS and GS stepwise for multiple traits in a maize improvement program requires further research and development.

Acknowledgments

The authors gratefully acknowledge the financial support from the MasAgro project funded by Mexico's Secretariat of Agriculture, Livestock, Rural Development, Fisheries and Food (SAGARPA), and the CGIAR Research Program (CRP) MAIZE. We thank Ed Buckler's Lab in Cornell University for GBS SNP calling. The State Administration of Foreign Experts Affairs of People's Republic of China and the Postdoctoral Workstation of Heilongjiang Academy of Agricultural Sciences provide a supporting fellowship to Dr. Shiliang Cao.

References

Aoun, M., M. Breiland, M.K. Turner, A. Loladze, S. Chao, S. Xu, K. Ammar, J.A. Anderson, J.A. Kolme, and M. Acevedo. 2016. Genome-wide association mapping of leaf rust response in a durum wheat worldwide germplasm collection. Plant Genome 9. doi:10.3835/plantgenome2016.01.0008

Bardol, N., M. Ventelon, B. Mangin, S. Jasson, V. Loywick, F. Couton, C. Derue, P. Blanchard, A. Charcosset, and L. Moreau. 2013. Combined linkage and linkage disequilibrium QTL mapping in multiple families of maize (*Zea mays* L.) line crosses highlights complementarities between models based on parental haplotype and single locus polymorphism. Theor. Appl. Genet. 126:2717–2736. doi:10.1007/s00122-013-2167-9

Bernardo, R. 2014. Genomewide selection when major genes are known. Crop Sci. 54:68–75. doi:10.2135/cropsci2013.05.0315

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. doi:10.1093/bioinformatics/btm308

Cairns, J.E., J. Crossa, P.H. Zaidi, P. Grudloyma, C. Sanchez, J.L. Araus, T. Suriphat, D. Makumbi, C. Magorokosho, M. Bänziger, A. Menkir, S. Hearne, and G.N. Atlin. 2013. Identification of drought, heat, and combined drought and heat tolerant donors in maize. Crop Sci. 53:1335–1346. doi:10.2135/cropsci2012.09.0545

Castaño, A. 1969. 'Tar spot' (mancha de asfalto) of corn leaf. Agricultura Tropical 25:332.
Ceballos, H., and J.A. Deutsch. 1992. Inheritance of resistance to tar spot complex in maize. Phytopathology 82:505–512. doi:10.1094/Phyto-82-505
Chen, Z., B. Wang, X. Dong, H. Liu, L. Ren, J. Chen, A. Hauck, W. Song, and J. Lai. 2014. An ultra-high density bin-map for rapid QTL mapping for tassel and ear architecture in a large F_2 maize population. BMC Genomics 15:433. doi:10.1186/1471-2164-15-433
CIMMYT Applied Molecular Genetics Laboratory. 2003. Laboratory protocols, 3rd ed. CIMMYT, Mexico, D.F. p. 7–11.
Combs, E., and R. Bernardo. 2013. Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. Plant Genome 6. doi:10.3835/plantgenome2012.11.0030
Crossa, J., P. Pérez, J. Hickey, J. Burgueno, L. Ornella, J. Cerón-Rojas, X. Zhang, S. Dreisigacker, R. Babu, Y. Li, D. Bonnett, and K. Mathews. 2014. Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity 112:48–60. doi:10.1038/hdy.2013.16
Ding, J., X. Wang, S. Chander, J. Yan, and J. Li. 2008. QTL mapping of resistance to Fusarium ear rot using a RIL population in maize. Mol. Breed. 22:395–403. doi:10.1007/s11032-008-9184-4
Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:E19379. doi:10.1371/journal.pone.0019379
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
Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One 9:E90346. doi:10.1371/journal.pone.0090346
Gowda, M., B. Das, D. Makumbi, R. Babu, K. Semagn, G. Mahuku, M.S. Olsen, J.M. Bright, Y. Beyene, and B.M. Prasanna. 2015. Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. Theor. Appl. Genet. 128:1957–1968. doi:10.1007/s00122-015-2559-0
Hock, J., U. Dittrich, B.L. Renfro, and J. Kranz. 1992. Sequential development of pathogens in the maize tar spot disease complex. Mycopathologia 117:157. doi:10.1007/BF00442777
Hock, J., J. Kranz, and B. Renfro. 1989. El complejo "mancha de asfalto" de maíz: Su distribucción geográfica, requisitos ambientales e importancia económica en México. Rev. Mex. Fitopatol. 7:129–135.
Hock, J., J. Kranz, and B.L. Renfro. 1995. Studies on the epidemiology of the tar spot disease complex of maize in Mexico. Plant Pathol. 44:490–502. doi:10.1111/j.1365-3059.1995.tb01671.x
Hubisz, M.J., D. Falush, M. Stephens, and J.K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. Mol. Ecol. Resour. 9:1322–1332. doi:10.1111/j.1755-0998.2009.02591.x
Kump, K.L., P.J. Bradbury, R.J. Wisser, E.S. Buckler, A.R. Belcher, M.A. Oropeza-Rosas, J.C. Zwonitzer, S. Kresovich, M.D. McMullen, D. Ware, P.J. Balint-Kurti, and J.B. Holland. 2011. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat. Genet. 43:163–168. doi:10.1038/ng.747
Li, C., Y. Li, P.J. Bradbury, X. Wu, Y. Shi, Y. Song, D. Zhang, E. Rodgers-Melnick, E.S. Buckler, Z. Zhang, Y. Li, and T. Wang. 2015. Construction of high-quality recombination maps with low-coverage genomic sequencing for joint linkage analysis in maize. BMC Biol. 13:78. doi:10.1186/s12915-015-0187-4
Li, H., S. Hearne, M. Banziger, Z. Li, and J. Wang. 2010. Statistical properties of QTL linkage mapping in biparental genetic populations. Heredity 105:257–267. doi:10.1038/hdy.2010.56
Li, H., Z. Peng, X. Yang, W. Wang, J. Fu, J. Wang, Y. Han, Y. Chai, T. Guo, N. Yang, J. Liu, M.L. Warburton, Y. Cheng, X. Hao, P. Zhang, J. Zhao, Y. Liu, G. Wang, J. Li, and J. Yan. 2013. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat. Genet. 45:43–50. doi:10.1038/ng.2484
Li, X., Z. Zhou, J. Ding, Y. Wu, B. Zhou, R. Wang, J. Ma, S. Wang, X. Zhang, Z. Xia, J. Chen, and J. Wu. 2016a. Combined linkage and association mapping reveals QTL and candidate genes for plant and ear height in maize. Front. Plant Sci. 7:833. doi:10.3389/fpls.2016.00833
Li, Y., X. Shi, H. Li, J.C. Reif, J. Wang, Z. Liu, S. He, B. Yu, and L. Qiu. 2016b. Dissecting the genetic basis of resistance to soybean cyst nematode combining linkage and association mapping. Plant Genome 9. doi:10.3835/plantgenome2015.04.0020
Liu, L.J. 1973. Incidence of tar spot disease of corn in Puerto Rico. J. Agric. Univ. P R. 57:211–216.
Lu, Y., S. Zhang, T. Shah, C. Xie, Z. Hao, X. Li, M. Farkhari, J.M. Ribaut, M. Cao, T. Rong, and Y. Xu. 2010. Joint linkage–linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. Proc. Natl. Acad. Sci. USA 107:19585–19590. doi:10.1073/pnas.1006105107
Mahuku, G., J. Chen, R. Shrestha, L.A. Narro, K.V.O. Guerrero, A.L. Arcos, and Y. Xu. 2016. Combined linkage and association mapping identifies a major QTL (*qRtsc8-1*), conferring tar spot complex resistance in maize. Theor. Appl. Genet. 129:1217–1229. doi:10.1007/s00122-016-2698-y
Manichaikul, A., J. Dupuis, S. Sen, and K.W. Broman. 2006. Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus. Genetics 174:481–489. doi:10.1534/genetics.106.061549
Maublanc, A. 1904. Especes nouvelles de champignons inferieurs. Bulletin Societe Mycologie France 20:72.
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.
Ornella, L., P. Pérez, E. Tapia, J.M. González-Camacho, J. Burgueño, X. Zhang, S. Singh, F. San Vicente, D. Bonnett, S. Dreisigacker, R. Singh, N. Long, and J. Crossa. 2014. Genomic-enabled prediction with classification algorithms. Heredity 112:616–626. doi:10.1038/hdy.2013.144
Pereyda-Hernández, J., J. Hernández-Morales, J.S. Sandoval-Islas, S. Aranda-Ocampo, C. de León, and N. Gómez-Montiel. 2009. Etiología y manejo de la mancha de asfalto (*Phyllachora maydis* Maubl.) del maíz en Guerrero, México. Agrociencia 43:511–519.
Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, S. Dreisigackere, J. Crossa, H. Sánchez-Villedae, M. Sorrellsd, and J.L. Jannink. 2012. Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Genome 5:103–113. doi:10.3835/plantgenome2012.06.0006
Prasanna, B.M., V. Chaikam, and G. Mahuku. 2012. Doubled haploid technology in maize breeding: Theory and practice. CIMMYT, Mexico, D.F.
ProMED-mail. 2009. Undiagnosed fungus, maize: Guatemala (02): Tar spot. 20 Apr. 2009. Archive no. 20090420.1491. http://www.promedmail.org/post/20090420.1491 (accessed 27 Apr. 2017)
Raman, H., R. Raman, P. Eckermann, N. Coombes, S. Manoli, X. Zou, D. Edwards, J. Meng, R. Prangnell, J. Stiller, J. Batley, D. Luckett, N. Wratten, and E. Dennis. 2013. Genetic and physical mapping of flowering time loci in canola (*Brassica napus* L.). Theor. Appl. Genet. 126:119–132. doi:10.1007/s00122-012-1966-8
R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/
Romay, M.C., M.J. Millard, J.C. Glaubitz, J.A. Peiffer, K.L. Swarts, T.M. Casstevens, R.J. Elshire, C.B. Acharya, S.E. Mitchell, S.A. Flint-Garcia, M.D. McMullen, J.B. Holland, E.S. Buckler, and C.A. Gardner. 2013. Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. 14:R55. doi:10.1186/gb-2013-14-6-r55
Ruhl, G., M.K. Romberg, S. Bissonnette, D. Plewa, T. Creswell, and K.A. Wise. 2016. First report of tar spot on corn caused by *phyllachora maydis* in the United States. Plant Disease, 100:1496. doi:10.1094/PDIS-12-15-1506-PDN
Shurtleff, M.C. 1982. Compendium of corn diseases, 2nd edition, American Phytopathological Society, St Paul, MN.
Spindel, J., H. Begum, D. Akdemir, P. Virk, B. Collard, E. Redoña, G. Atlin, J.L. Jannink, and S.R. McCouch. 2015. Genomic selection and association mapping in rice (*Oryza sativa*): Effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLoS Genet. 11:E1004982. doi:10.1371/journal.pgen.1004982

Swarts, K., H. Li, J.A.R. Navarro, D. An, M.C. Romay, S. Hearne, C. Acharyae, J.C. Glaubitze, S. Mitchelle, R.J. Elshireg, E.S. Buckler, and P.J. Bradbury. 2014. Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. Plant Genome 7. doi:10.3835/plantgenome2014.05.0023

Tenaillon, M.D., M.C.S. Anthony, D. Long, R.L. Gaut, J.F. Doebley, and B.S. Gaut. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc. Natl. Acad. Sci. USA 98:9161–9166. doi:10.1073/pnas.151244298

Trachsel, S., D. Sun, F. San Vicente, H. Zheng, E. Antonio Suarez, R. Babu, and X. Zhang. 2016. Identification of QTL for early vigor and stay-green conferring tolerance to drought in two connected advanced backcross populations in tropical maize (*Zea mays* L.). PLoS ONE 11:e0149636. doi:10.1371/journal.pone.0149636

Turner, S.D. 2014. qqman: An R package for visualizing GWAS results using Q-Q and Manhattan plots. doi:10.1101/005165

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

Wen, W., J.L. Araus, T. Shah, J. Cairns, G. Mahuku, M. Bänziger, J.L. Torres, C. Sánchez, and J. Yan. 2011. Molecular characterization of a diverse maize inbred line collection and its potential utilization for stress tolerance improvement. Crop Sci. 51:2569–2581. doi:10.2135/cropsci2010.08.0465

Wu, J., J. Ding, Y. Du, Y. Xu, and X. Zhang. 2007. Genetic analysis and molecular mapping of two complementary dominant genes determining resistance to sugarcane mosaic virus in maize. Euphytica 156:355–364. doi:10.1007/s10681-007-9384-8

Wu, Y., F. San Vicente, K. Huang, T. Dhliwayo, D.E. Costich, K. Semagn, N. Sudha, M. Olsen, B.M. Prasanna, X. Zhang, and R. Babu. 2016. Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. Theor. Appl. Genet. 129:753–765. doi:10.1007/s00122-016-2664-8

Xu, Y., Y. Lu, C. Xie, S. Gao, J. Wan, and B.M. Prasanna. 2012. Whole-genome strategies for marker-assisted plant breeding. Mol. Breed. 29:833–854. doi:10.1007/s11032-012-9699-6

Yu, J., and E.S. Edward. 2006. Genetic association mapping and genome organization of maize. Curr. Opin. Biotechnol. 17:155–160. doi:10.1016/j.copbio.2006.02.003

Zhang, X., P. Pérez-Rodrígue, K. Semagn, Y. Beyene, R. Babu, M.A. López-Cruz, F. San-Vicente, M. Olsen, E.S. Buckler, J.L. Jannink, B.M. Prasanna, and J. Crossa. 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

Zhao, Y., M. Gowda, W. Liu, T. Würschum, H. P. Maurer, F. H. Longin, N. Ranc, J. C. Reif. 2012. Accuracy of genomic selection in European maize elite breeding populations. Theor. Appl. Genet. 124:769-776. doi:10.1007/s00122-011-1745-y

Zhu, C., M. Gore, E.S. Buckler, and J. Yu. 2008. Status and prospects of association mapping in plants. Plant Genome 1:5–20. doi:10.3835/plantgenome2008.02.0089

Zhu, J., S.M. Kaeppler, and J.P. Lynch. 2005. Mapping of QTL controlling root hair length in maize (*Zea mays* L.) under phosphorus deficiency. Plant Soil 270:299–310. doi:10.1007/s11104-004-1697-y