Available on at: www.sciencedirect.com

ScienceDirect



Genome-wide association study and genomic prediction of *Fusarium* ear rot resistance in tropical maize germplasm

Yubo Liu^{a,b,c,1}, Guanghui Hu^{b,d,1}, Ao Zhang^{b,c}, Alexander Loladze^b, Yingxiong Hu^{b,e,f}, Hui Wang^{b,e,f}, Jingtao Qu^g, Xuecai Zhang^b, Michael Olsen^h, Felix San Vicente^b, Jose Crossa^b, Feng Lin^{c,*}, Boddupalli M. Prasanna^{h,*}

^a College of Agronomy, Shenyang Agricultural University, Shenyang 110866, Liaoning, China

^b International Maize and Wheat Improvement Center (CIMMYT), El Batan, Texcoco, Mexico

^c College of Biological Science and Technology, Shenyang Agricultural University, Shenyang 110866, Liaoning, China

^d Maize Research Institute, Heilongjiang Academy of Agricultural Sciences, Harbin 150086, Heilongjiang, China

^e CIMMYT–China Specialty Maize Research Center, Shanghai Academy of Agricultural Sciences, Shanghai 200063, China

^f Crop Breeding and Cultivation Research Institute, Shanghai Academy of Agricultural Sciences, Shanghai 200063, China

^g Maize Research Institute, Sichuan Agricultural University, Wenjiang 611130, Sichuan, China

^h International Maize and Wheat Improvement Center (CIMMYT), P. O. Box 1041, Village Market, Nairobi 00621, Kenya

ARTICLE INFO

Article history:

Received 16 June 2020

Received in revised form

24 July 2020

Accepted 3 August 2020

Available online 14 October 2020

Keywords:

Maize

Fusarium ear rot

Genome-wide association study

Genomic prediction

Genomic selection

ABSTRACT

Fusarium ear rot (FER) is a destructive maize fungal disease worldwide. In this study, three tropical maize populations consisting of 874 inbred lines were used to perform genome-wide association study (GWAS) and genomic prediction (GP) analyses of FER resistance. Broad phenotypic variation and high heritability for FER were observed, although it was highly influenced by large genotype-by-environment interactions. In the 874 inbred lines, GWAS with general linear model (GLM) identified 3034 single-nucleotide polymorphisms (SNPs) significantly associated with FER resistance at the P -value threshold of 1×10^{-5} , the average phenotypic variation explained (PVE) by these associations was 3% with a range from 2.33% to 6.92%, and 49 of these associations had PVE values greater than 5%. The GWAS analysis with mixed linear model (MLM) identified 19 significantly associated SNPs at the P -value threshold of 1×10^{-4} , the average PVE of these associations was 1.60% with a range from 1.39% to 2.04%. Within each of the three populations, the number of significantly associated SNPs identified by GLM and MLM ranged from 25 to 41, and from 5 to 22, respectively. Overlapping SNP associations across populations were rare. A few stable genomic regions conferring FER resistance were identified, which located in bins 3.04/05, 7.02/04, 9.00/01, 9.04, 9.06/07, and 10.03/04. The genomic regions in bins 9.00/01 and 9.04 are new. GP produced moderate accuracies with genome-wide markers, and relatively high accuracies with SNP associations detected from GWAS. Moderate prediction accuracies were observed when the training and validation sets were closely related. These results implied that FER resistance in maize is controlled by minor QTL with small effects,

* Corresponding authors.

E-mail addresses: fenglinsn@126.com (F. Lin), b.m.prasanna@cgiar.org (B.M. Prasanna).

¹ These authors contributed equally to this work.

<https://doi.org/10.1016/j.cj.2020.08.008>

0168-8227/© 2020 Crop Science Society of China and Institute of Crop Science, CAAS. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co., Ltd.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

and highly influenced by the genetic background of the populations studied. Genomic selection (GS) by incorporating SNP associations detected from GWAS is a promising tool for improving FER resistance in maize.

© 2020 Crop Science Society of China and Institute of Crop Science, CAAS. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Fusarium ear rot (FER), mainly caused by *Fusarium verticillioides* (Sacc.) Nirenberg, is a destructive maize fungal disease worldwide, as the disease can significantly reduce grain yields and results in poor grain quality [1,2]. Moreover, *Fusarium verticillioides* produce several types of mycotoxins, and consumption of the maize grains contaminated by mycotoxins is fatal to humans and animals [3,4]. The most economical and effective strategy for controlling FER is the development and deployment of maize varieties with genetic resistance [5]. This requires the identification of source germplasm with stable FER resistance and good agronomic performance. Understanding the genetic basis and detection of favorable alleles associated with FER resistance will enable the implementation of an effective breeding strategy for improvement of FER resistance in the appropriate breeding backgrounds.

Broad genetic variation for FER resistance exists in maize. A few maize inbred lines showing stable resistance to FER have been identified in several previous studies, although fully immune materials have not been reported in maize [6,7]. A total of 1687 diverse maize inbred lines were evaluated for FER resistance across years by Zila et al. [6], which showed that most novel resistance sources to FER are mainly from tropical maize. Furthermore, response to FER in 940 tropical maize inbred lines was evaluated by Chen et al. [7] in three environments, and 63 lines with FER resistance were identified. These inbred lines could serve as valuable donors for improving FER resistance through breeding. Moreover, genetic studies involved with these novel resistance sources could provide a better understanding of the genetic architecture of FER resistance in tropical maize germplasm.

Linkage mapping in different bi-parental populations in maize undertaken by different teams earlier has revealed several genomic regions conferring FER resistance [8–10]. The quantitative trait loci (QTL) identified in these studies were distributed on all the ten maize chromosomes. Also, the QTL detected in different studies were inconsistent and varied in different environments and populations studied. The major QTL detected in these studies need to be furtherly validated before implementing marker-assisted-selection (MAS) in breeding programs. Genome-wide association study (GWAS) is a powerful alternative strategy to detect QTL in a collection of genetically diverse germplasm. The strategy of GWAS has higher mapping resolution than linkage mapping in maize, because maize has abundant genetic diversity and faster linkage disequilibrium decay [11]. However, GWAS may also detect a high number of false-positive associations, and it is also less efficient at detecting alleles with small effects from many loci, as well as the rare alleles. Bigger population sizes and higher density markers are required to

increase the statistical mapping power of GWAS. Several GWAS analyses have been performed for genetic mapping of FER resistance in maize [12–14]. The results of these studies showed that resistance to FER in maize is controlled by several minor QTL with small effects, and it is highly influenced by the genetic background of the populations studied, the phenotyping environments, and the genotype-by-environment interaction. These results limit the effectiveness of MAS in pyramiding small-effect favorable alleles for improving FER resistance.

As an alternative to MAS especially for improving complex traits, genomic selection (GS), also known as genomic prediction (GP), offers significant promise in crops like maize. In GS/GP, the genome-wide markers are used to predict the genomic-estimated breeding values (GEBVs) of individuals by capturing the effects of both major and minor genes [15–17]. Therefore, GS captures a greater proportion of the genetic variation of the selected trait as compared to MAS, where only the selected markers that are significantly associated with the target trait are used. Several studies have been conducted to investigate the effectiveness of implementing GS for improving resistance to some major maize diseases, e.g., maize lethal necrosis (MLN) [18,19], tar spot complex [20], and gibberella ear rot [21]. Medium to high prediction accuracies were reported in these studies, which suggested that GS is a potential genomic tool for improving complex traits under polygenic control. Larger training population size across locations and years are required to obtain high prediction accuracy. Incorporating known trait-marker associations into the prediction model may also improve prediction accuracy [22–24].

In the present study, GWAS and GP analyses were performed on 874 maize inbred lines, where all the inbred lines were screened in four to six environments in Mexico to evaluate their response to FER, and genotyped with genotyping-by-sequencing. The main objectives of the present study are to: (1) identify the genomic regions, single-nucleotide polymorphisms (SNPs), and putative candidate genes conferring FER resistance in tropical maize germplasm through GWAS; (2) explore the potential of GS for improving FER resistance in tropical maize, and (3) estimate the accuracies of GP by incorporating the significantly associated SNPs into the prediction model.

2. Materials and methods

2.1. Plant materials

Three maize inbred line panels were used to perform the GWAS and GP analyses in the present study. The first population, designated as CIMMYT maize inbred lines (CMLs), consists of 254 tropical and subtropical maize inbred lines

developed by the International Maize and Wheat Improvement Center (CIMMYT). The CMLs represent elite lines that were selected based on their per se performance, good general combining ability, and a significant number of key traits (e.g., abiotic and biotic stress tolerance, nitrogen use efficiency, etc.). From 1984 to 2019, CIMMYT developed and released a total of 603 CMLs, which represents a significant portion of the genetic diversity of CIMMYT-derived improved tropical and subtropical maize germplasm [25]. In the present study, the most frequently used lines and testers before CML300, as well as all the lowland tropical, mid-altitude/subtropical lines between CML300 and CML577, were selected for screening their response to FER.

The second population, designated as Drought Tolerant Maize for Africa (DTMA) association mapping (AM) panel, consisted of 296 tropical and subtropical inbred lines developed by CIMMYT. These lines originated from different breeding programs of CIMMYT and consisted of several lines with tolerance or resistance to an array of abiotic and biotic stresses affecting maize in the tropics [26].

The third population, designated as SYN_DH, consisted of 324 maize double haploid (DH) lines derived from four tropical maize synthetic populations. Parental lines of each of these synthetic populations were selected based on the information of the kernel color (white or yellow) and heterotic groups. The CIMMYT maize heterotic groups were classified as group “A” (dent type) and group “B” (flint type). Besides, the parental lines were also selected based on the information of their general combining ability for grain yield, tolerance/resistance to major abiotic and biotic stresses, and adaptation to lowland tropical breeding environments. In total, 63 parental lines were used to form the four synthetic populations; each synthetic population was formed through twice intermated pollinations, once self-pollination, and once DH process to advance all the materials as inbred lines. The number of DH lines in each synthetic population ranged from 39 to 97.

2.2. Experimental design and phenotypic evaluation

A total of 874 tropical maize inbred lines were used in the present study for screening their response to FER. Detailed information about these inbred lines is provided in Table S1. The CMLs population of 254 inbred lines was screened in Mexico at two experimental stations, Agua Fria (AF) located in the state of Puebla (Latitude: 20°28'N; Longitude: 97°38'W), and Tlaltizapan (TL) located in the state of Morelos (18°41'N, 99°07'W). The trials were carried out at both AF and TL experimental stations during the winter season (November–May) in 2017–2018 and 2018–2019, and the summer season (May–November) in 2018. The winter season was defined as Cycle “A”, and the summer season was defined as Cycle “B”. The environment was defined as a combination of location and year-season. In total, the CMLs population was screened in six environments, designated as AF_18A, AF_18B, AF_19A, TL_18A, TL_18B, and TL_19A, respectively. The DTMA AM panel of 282 inbred lines was screened in four environments at the AF experimental station during the summer season in 2008 (AF_08B), 2010 (AF_10B), and 2011 (AF_11B); and at TL experimental station during the winter season in 2013–

2014 (TL_14A). The SYN_DH population of 325 DH lines was screened in four environments at the AF experimental station during the summer season of 2015 (AF_15B) and 2016 (AF_16B), and at the TL experimental station during the summer season of 2015 (TL_15B), and the winter season of 2015–2016 (TL_16A). The 63 parental lines used to form the SYN_DH population, were screened at four environments at the AF experimental station during the winter season in 2016–2017 (AF_17A) and 2018–2019 (AF_19A), and the summer season in 2019 (AF_19B); and at the TL experimental station during the winter season in 2018–2019 (TL_19A).

A randomized complete block design was utilized for all the trials with three replications per environment: each plot is a single 2.5 m row, with 20 cm spacing between plants in a row and 75 cm spacing between rows. Two tropical maize inbred lines, CML155 and Colombia-35, were used in all the trials as the FER resistant and susceptible checks, respectively. These two lines were selected based on data from multiple years of artificial inoculation trials [7]. Days to 50% anthesis and 50% silking were also recorded.

Artificial inoculation with *F. verticillioides* was applied on all the trials with a nail punch/sponge method [27]. A single-spore of aggressive *Fusarium verticillioides* was previously isolated and inoculated on sterilized maize kernels, followed by incubation for 14 days at 25 °C for the increase. After incubation, the spores were harvested by washing the inoculated grains with sterile water, and the final concentration of the spore suspension was adjusted to 5×10^6 spores mL⁻¹ with a hemocytometer. Tween-20 was added to the suspension (0.2 mL L⁻¹) as a surfactant and mixed well before the field inoculation. The primary ear of each plant was artificially inoculated 10–15 days after pollination. Since the maize inbred lines had different flowering times, the inoculations were conducted several times to ensure the uniformity of the disease development and lower the chances of disease escape.

At the grain maturity stage, the inoculated ears were harvested by hand and individually rated for disease severity based on a seven-grade infected kernel evaluation scales: where 0 = no visible disease symptoms, 0.05 = 1 to 5%, 0.10 = 6 to 10%, 0.25 = 11 to 25%, 0.50 = 26 to 50%, 0.75 = 51 to 75%, and 1 = 76 to 100%. The grade was then converted into an average percentage of the infected area (Fv%) to represent the FER severity [28] using the below formula:

$$Fv\% = \frac{\sum(\text{grade} \times \text{number of ears in the grade})}{1 \times \text{total number of ears}} \times 100$$

2.3. Phenotypic data analysis

For each population, the best linear unbiased estimates (BLUEs) and broad-sense heritability (H^2) of Fv% were calculated within single environment analysis and combined analysis across environments (CombinedENV) using the META-R software [29] (<http://hdl.handle.net/11529/10201>). For the single environment analysis, a mixed linear model (MLM) was performed including line as a fixed effect, and replication as a random effect. The broad-sense heritability (H^2) was estimated using the following formula:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2/r} \quad (1)$$

For the CombinedENV, a mixed linear model was performed including line as a fixed effect, environment, replication within the environment, and line-by-environment interaction as random effects. The environment was considered as the combination of location and year-season. The broad-sense heritability (H^2) was estimated using the following formula:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{ge}^2/n + \sigma_e^2/nr} \quad (2)$$

where σ_g^2 is the genotypic variance, σ_{ge}^2 is the genotype-by-environment variance, σ_e^2 is the error variance, n is the number of environments, and r is the number of replications. Besides, Pearson's correlation coefficients between environments were calculated by using R function `cor.test` within each population.

2.4. Genotypic data

A genotyping-by-sequencing protocol developed by Elshire et al. [30] was applied, which has been commonly used by the maize research community. Genotyping was conducted at the Genomic Diversity Facility of Cornell University, USA. Genomic DNA was digested with the restriction enzyme *ApeKI*, and DNA libraries were constructed by multiplexing 96 samples per library, and each DNA library was sequenced on one sequencing lane of Illumina HiSeq2000. Details of analyses of SNP calling and imputation have been previously described [20]. For each inbred line, 955,690 SNPs evenly distributed on the maize chromosomes were called. The imputed dataset was used to conduct further GWAS and GP analyses. Genome Association and Prediction Integrated Tool (GAPIT) [31] was used to filter raw marker datasets for SNPs with minor allele frequency (MAF) greater than 0.05 and missing data rate less than 20% within each population and across all the populations.

2.5. GWAS analysis

After filtering, GWAS analyses were performed on the populations of CMLs, DTMA AM panel, SYN_DH, and in all the 874 inbred lines with 212,353, 228,272, 212,558, and 201,970 high-quality SNPs, respectively. In the R package [32] of GAPIT, the linkage disequilibrium analysis was conducted within each of the above four populations. Principal component analysis (PCA) was performed within each population to stratify the population structure, and the relative kinship matrix (K) was generated for each population to assess the relatedness among individuals within each population. The population structure within each population was illustrated using the first two principal components, and a heat map of pairwise genetic distances between the genotyped inbred lines. Both the GLM (general linear model) and MLM (mixed linear model) methods were applied in GAPIT to detect the associations between the SNPs and the target trait of Fv% representing the disease severity. In the GLM method, GWAS was conducted by incorporating BLUEs, SNP markers, and PCA.

To eliminate false-positive associations, GWAS was conducted with the MLM method by incorporating BLUEs, SNP markers, PCA, and K . In both GLM and MLM, the first three principal components, the default parameter in GAPIT, were used to stratify the population structure, and K was generated for assessing the relatedness among individuals. The quantile–quantile and Manhattan plots were created. Based on the suggestions of the previous GWAS researches of FER in maize [7,12,14], the P-value threshold used to declare the significant associations in GLM and MLM was set as 1×10^{-5} and 1×10^{-4} , respectively. Considering that the FER in maize is controlled by many minor QTL with small effects, using a less stringent threshold of 1×10^{-4} or 1.0×10^{-5} in this study is reasonable. The candidate genes associated with FER resistance were identified according to the B73 reference genome information available in the MaizeGDB database (<https://www.maizegdb.org>). In GLM, the genes containing the SNPs significantly associated with FER resistance were considered as the putative candidate genes. In MLM, the genes containing the SNP associations or adjacent to the SNP associations in half of the linkage disequilibrium decay distance of the 874 inbred lines were considered as the putative candidate genes. The general feature format gene annotation file and functional annotation file of B73_RefGen_v2 were downloaded from the MaizeGDB database (<https://www.maizegdb.org>). The annotation analysis of the candidate genes was performed against the annotated B73 reference genome with the customized Perl scripts.

2.6. GP analysis

The GP analysis was conducted using the BGLR library [33] in the R program [32]. A total of 300,000 Markov chain Monte Carlo samples were collected, with 250,000 discarded as burn-in. Thinning was done by keeping one of every 10 samples. GP analyses were performed on the populations of CMLs, DTMA AM panel, SYN_DH, and all the 874 inbred lines. Within each population, prediction accuracies for FER performance were estimated with the genome-wide markers within each environment and in CombinedENV. Across all the populations, GP analysis was only performed in CombinedENV. The genome-wide marker datasets filtered with MAF greater than 0.05 and missing data rate less than 20% within each population, were used to perform GP analyses. A five-fold cross-validation scheme with 100 replications was used to generate the training and validation sets and assess the prediction accuracy. The average value of the correlations between the phenotypic and the genomic estimated breeding values was defined as the prediction accuracy.

In parallel, GP analyses with significantly associated SNPs were also performed to simulate MAS, where the significantly associated SNPs detected from the GLM method of the GWAS analysis in each population, defined as the self-significant SNPs (Self_sigSNP), were selected to perform GP on the same population. Moreover, all the unique significantly associated SNPs detected in the GLM method in all the GWAS analyses across all the populations, defined as the all-significant SNPs (All_sigSNP), were also selected to perform GS prediction on each population. The prediction accuracies were estimated

from the five-fold cross-validation scheme with 100 replications.

The transferability of the GP models across populations was tested, when training and validation sets were independent, and one or two populations were used as the training set to predict the other population as the validation set. The prediction accuracy was estimated as the correlation between the phenotypic and the genomic estimated breeding values of the validation set.

3. Results

3.1. Phenotypic variation and correlations among environments

Broad phenotypic variation of disease severity was observed within each population (Table 1, Fig. S1). In CombinedENV analyses, Fv% ranged from 3.82% to 90.84% with an overall mean of 31.76% in the CMLs population, from 15.95% to 96.61% with an overall mean of 45.40% in the DTMA AM panel, from 1.02% to 96.98% with an overall mean of 28.17% in the SYN_DH population, and from 9.26% to 68.94% with an overall mean of 24.66% for the parental lines of the SYN_DH population. The overall mean of Fv% varied in different populations: the DTMA AM panel had a higher value than

other populations, while the overall means of Fv% in the populations of SYN_DH and their parental lines were lower than those in the CMLs population and DTMA AM panel.

Within all the populations, the component of genetic variance (σ_g^2) was significant at P-value of 0.01 in all the single-environment ANOVA analyses. In the CombinedENV ANOVA analyses, both genetic variance (σ_g^2) and genotype \times environment variance (σ_{ge}^2) components were significant at P-value of 0.01 within all the populations, indicating that FER resistance is controlled by genetic factors, but it is also greatly affected by genotype \times environment interactions (Table 1). The estimated heritability of FER resistance in the single-environment analyses ranged from 0.72 to 0.86 in the CMLs population, from 0.68 to 0.91 in the DTMA AM panel, from 0.84 to 0.89 in the SYN_DH population, and from 0.69 to 0.89 for the parental lines of the SYN_DH population. The heritability of FER estimated in CombinedENV analyses in the populations of CMLs, DTMA AM panel, SYN_DH, and the parental lines of the SYN_DH were 0.87, 0.64, 0.87, and 0.76, respectively (Table 1). The estimated heritabilities of FER resistance were relatively high in both the single-environment and CombinedENV analyses, indicating that the phenotypic data was reliable for further genetic analyses.

In the single-environment analyses, the Pearson's correlation coefficients among environments were moderate to high

Table 1 – Information of FER response in the populations of CMLs, DTMA AM panel, SYN_DH, and SYN_DH parents in each environment and the combined-environment analyses, including mean value (Mean), minimum value (Min), maximum value (Max), standard deviation, estimates of genotypic variance (σ_g^2), genotype-by-environment variance (σ_{ge}^2), error variance (σ_e^2), and heritability (H^2).

Population	Environment	Mean (%)	Min (%)	Max (%)	SD	σ_g^2	σ_{ge}^2	H^2
CMLs	AF_18A	39.15	2.31	100	0.21	0.030**		0.75
	AF_18B	42.51	4.15	97.94	0.23	0.047**		0.85
	AF_19A	26.38	4.40	98.55	0.18	0.027**		0.86
	TL_18A	31.53	0	100	0.22	0.031**		0.71
	TL_18B	23.11	0	85.83	0.16	0.018**		0.77
	TL_19A	27.35	3.53	99.19	0.18	0.024**		0.72
	CombinedENV	31.76	3.82	90.84	0.13	0.020**	0.010**	0.86
DTMA AM panel	AF_08B	59.89	14.67	100	0.20	0.033**		0.83
	AF_10B	52.77	8.33	100	0.22	0.034**		0.73
	AF_11B	35.35	1.09	100	0.27	0.065**		0.91
	TL_14A	32.95	0	100	0.21	0.029**		0.68
	CombinedENV	45.4	15.95	96.61	0.16	0.015**	0.024**	0.64
SYN_DH	AF_15B	30.24	0.00	100	0.23	0.044**		0.85
	AF_16B	29.13	0.47	100	0.23	0.044**		0.88
	TL_15B	30.89	0	100	0.24	0.052**		0.88
	TL_16A	22.77	0	100	0.21	0.036**		0.84
	CombinedENV	28.17	1.02	96.98	0.19	0.033**	0.013**	0.87
SYN_DH parents	AF_17A	21.86	1.67	78.82	0.16	0.023**		0.86
	AF_19A	23.51	5.00	88.23	0.19	0.031**		0.87
	AF_19B	15.65	5.00	92.17	0.15	0.019**		0.89
	TL_19A	37.89	8.11	75.12	0.18	0.024**		0.69
	CombinedENV	24.66	9.26	68.94	0.13	0.013**	0.011**	0.76

CML, CIMMYT maize inbred line; DTMA AM: Drought Tolerant Maize for Africa (DTMA) association mapping (AM) panel; SYN_DH, double haploid (DH) lines derived from synthetic population (SYN).

Environment, Location_Season, for examples, AF_18A means the environment at Agua Fria (AF) experimental station during the winter season in 2017–2018, the winter season is defined as Cycle “A”; and TL_18B means the environment at Tlaltizapan (TL) experimental station during the summer season in 2018, the summer season is defined as Cycle “B”.

** Significant at P = 0.01. SD, standard deviation.

in the populations of CMLs ($r = 0.42\text{--}0.59$) and SYN_DH ($r = 0.55\text{--}0.74$), and low to moderate in the DTMA AM panel ($r = 0.15\text{--}0.49$). Within all the populations, much stronger Pearson's correlation coefficients were observed between the single-environment and CombinedENV, which ranged from 0.73 to 0.81 in the CMLs population, from 0.56 to 0.76 in the DTMA AM panel, and from 0.81 to 0.87 in the SYN_DH population (Fig. S1). Therefore, subsequent GWAS analyses were only performed using phenotypic data from CombinedENV analyses.

Based on phenotypic data, a few maize inbred lines were identified as sources of resistance to FER in each of the populations (Table S1). These sources of FER resistance were CML287, CML300, CML360, CML362, CML424, CML452, CML479, CML481, CML482, CML575, and CML577 among the CMLs, which had Fv% consistently lower than 10% in the CombinedENV analysis. In the DTMA AM panel, the sources of FER resistance identified were DTMA55, DTMA56, DTMA176, DTMA177, DTMA185, DTMA191, DTMA195, DTMA196, DTMA215, DTMA252, which had Fv% consistently lower than 20% in the CombinedENV analysis. In the SYN_DH population, the FER resistant line identified were RCGS60, RCGS178, RCGS206, RCGSY20, RCGSY76, RCGSY99, and RCGSY154, as these lines had Fv% consistently lower than 5% in the CombinedENV analysis.

3.2. Population structure analysis

The population structure within each population was illustrated based on the first two principal components in Fig. 1, a heat map and a dendrogram of the kinship matrix in Fig. S2. The results showed little population structure and low levels of genetic relatedness among the inbred lines in the CMLs population and DTMA AM panel. Much population structure and higher levels of genetic relatedness were observed among the inbred lines in the SYN_DH population. A higher level of genetic relatedness was observed between the CMLs population and the DTMA AM panel, and a lower level of genetic relatedness was observed between the SYN_DH population and the other two populations.

The average linkage disequilibrium decay distance over all ten chromosomes in the populations of CMLs, DTMA AM panel, and all the 874 inbred lines with $r^2 = 0.1$ were 5.22, 5.66, and 7.31 kb, respectively. In the SYN_DH population, the average linkage disequilibrium decay distance over all ten chromosomes with $r^2 = 0.1$ was greater than 50 kb.

3.3. GWAS analysis for FER resistance using the GLM method

The results of GWAS for FER resistance using the GLM method and CombinedENV phenotypic data are shown in Fig. 2 and Table S2. In total, 41, 25, 26, and 3034 SNPs, significantly associated with FER resistance at the P -value threshold of 1×10^{-5} , were identified in the CMLs population, DTMA AM panel, SYN_DH population, and in all the 874 inbred lines, respectively. These SNPs significantly associated with FER resistance were distributed on all the ten maize chromosomes. The average MAF across all the SNP associations was 0.22, 0.17, 0.27, and 0.24 in the CMLs population, DTMA AM panel, SYN_DH

population, and in all the 874 inbred lines, respectively (Table S2).

In the CMLs population, the greatest number of marker-trait associations, i.e. 10 SNPs, were detected on chromosome 5. The average PVE (phenotypic variation explained) value of these 41 associations was 10.55% with a range from 9.46% to 13.23%, and 23 of them had a PVE value greater than 10%. The association of SNP S7_145770513 had the lowest P -value of 2.21×10^{-7} , and the greatest PVE value of 13.23%. In the DTMA population, the greatest number of associations, i.e. 12 SNPs, were detected on chromosome 5. The average PVE value of the 25 associations was 7.83% with a range from 6.73% to 11.74%, and only two of them had a PVE value greater than 10%. The association of SNP S3_56477054 had the lowest P -value of 6.80×10^{-9} , and the greatest PVE value of 11.74%. The association of SNP S1_299329113 had the second-lowest P -value of 5×10^{-8} , and the second greatest PVE value of 10.31%. In the SYN_DH population, the greatest number of associations, i.e. 10 SNPs, were detected on chromosome 3. The average PVE value of the 26 associations was 6.77% with a range from 6.14% to 8.39%. The association of SNP S3_202836307 had the lowest P -value of 2.78×10^{-7} , and the greatest PVE value of 8.39%. In the GWAS analysis of all the 874 inbred lines, the number of associations per chromosome ranged from 175 on chromosome 10 to 559 on chromosome 1, with an average value of 303 associations per chromosome, and 604 of the 3034 association had P -values lower than 1×10^{-7} . The average PVE value of the 3034 associations was 3% with a range from 2.33% to 6.92%, and 49 associations had a PVE value greater than 5%. The association of SNP S3_141271697 had the lowest P -value of 4.99×10^{-14} , and the greatest PVE value of 6.92% (Table S2).

The number of overlapping SNP associations between different populations is shown in Fig. S3. No overlapping SNP association was observed among the CMLs population, DTMA AM panel, and SYN_DH population. The number of overlapping SNP associations with all the 874 inbred lines was 11, 1, and 7 in the populations of CMLs, DTMA AM panel, and SYN_DH, respectively.

Compared with the GWAS analyses within each population, GWAS analysis with all the 874 inbred lines detected much more marker-trait associations, with lower P -values and smaller PVE values. This result indicates that bigger population size is required to conduct GWAS to dissect the polygenic trait controlled by many genes with minor allele effects, such as FER resistance in maize. However, the quantile–quantile plots from all the GWAS analyses using the GLM method implied that the false-positive rates have to be further reduced through controlling both the population structure and family relatedness in the GWAS statistical model (Fig. 2).

3.4. GWAS analysis for FER resistance using the MLM method

The results of GWAS for FER resistance using the MLM method and CombinedENV phenotypic data are shown in Fig. 3, Tables 2 and 3. In total, 12, 5, 22, and 19 SNPs, significantly associated with FER resistance at the P -value threshold of 1×10^{-4} , were identified in the CMLs population, DTMA AM panel, SYN_DH population, and in all the 874 inbred lines,

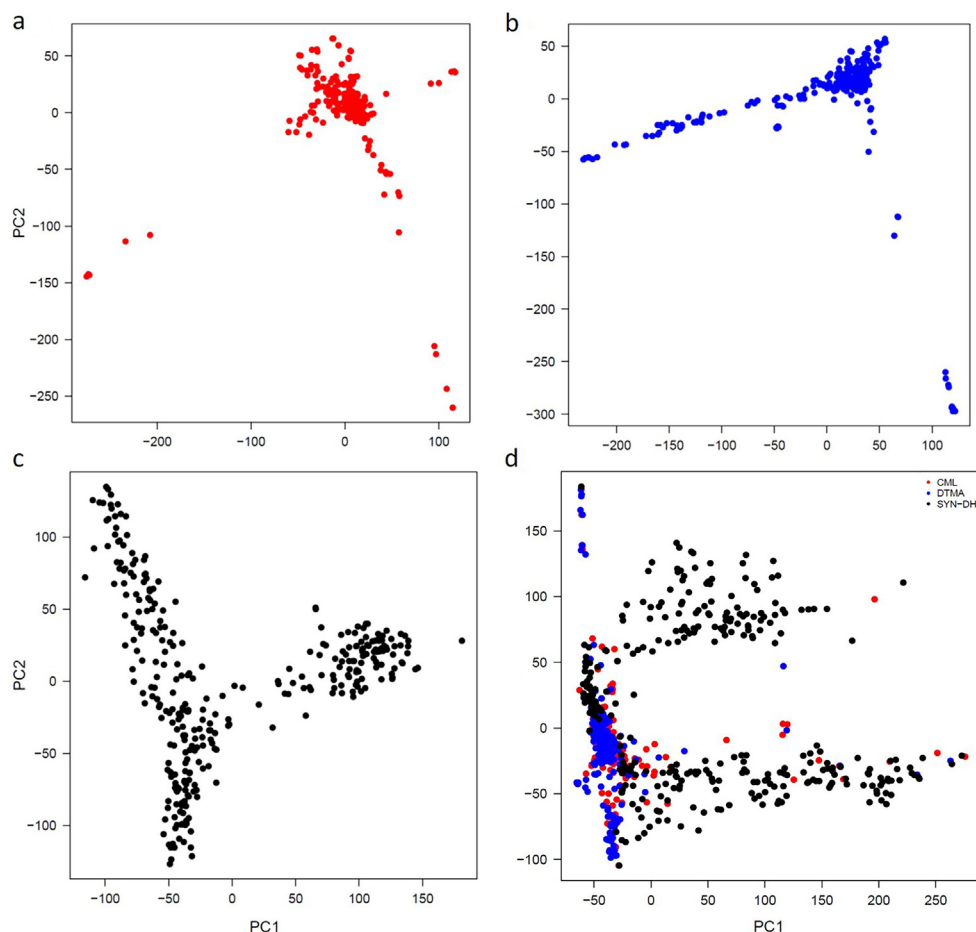


Fig. 1 – Genetic relatedness among the inbred lines visualized using a principal component (PC) analysis in the populations of (a) CMLs, (b) DTMA AM panel, (c) SYN_DH, and (d) in all the 874 inbred lines. The horizontal and vertical axes are the first and second principal components, respectively.

respectively. The average MAF across all the association SNPs was 0.18, 0.15, 0.26, and 0.28 in the CMLs population, DTMA AM panel, SYN_DH population, and in all the 874 inbred lines, respectively (Tables 2, 3).

In the CMLs population, the SNP associations were distributed on chromosomes 2, 3, 4, 6, 7, 8, and 9. The number of SNP associations per chromosome ranged from 1 to 3. The greatest number of associations were detected on chromosomes 6 and 9. The average PVE value of the 12 associations was 7.01% with a range from 6.34% to 7.71%. The association of SNP S8_16910712 had the lowest P -value of 1.39×10^{-5} , and the greatest PVE value of 7.71%. In the DTMA AM panel, the SNP associations were distributed on chromosomes 1, 3, 5, 7, and 9, with only one SNP association per chromosome. The average PVE value of the five associations was 5.01% with a range from 4.52% to 5.96%. The association of SNP S3_56477054 had the lowest P -value of 6.73×10^{-6} , and the greatest PVE value of 5.96%. In the SYN_DH population, the SNP associations were distributed on chromosomes 1, 2, 3, 6, 7, 8, and 9. The number of SNP associations per chromosome ranged from 1 to 6, and the greatest number of associations were detected on chromosome 3. The average PVE value of the 22 associations was 4.88% with a range from 4.47% to 5.91%. The association of SNP S1_260142965 had

the lowest P -value of 7.80×10^{-6} , and the greatest PVE value of 5.91% (Table 2).

The GWAS analysis in all the 874 inbred lines using the MLM method detected 19 SNP associations, which were distributed on all the maize chromosomes, except on chromosome 7. The number of SNP associations per chromosome ranged from 1 to 5, and the greatest number of associations were detected on chromosomes 1 and 9. The average PVE value of the 19 associations was 1.60% with a range from 1.39% to 2.04%. The association of SNP S9_115952576 had the lowest P -value of 2.33×10^{-6} , and the greatest PVE value of 2.04% (Table 3).

The number of overlapping SNP associations between different populations is shown in Fig. S3, only two overlapping SNP associations, i.e., S3_142736417 and S6_159221992, were observed between SYN_DH population and in all the 874 inbred lines. The presence of a few overlapping SNPs between different populations implied that FER resistance in maize is indeed a complex trait, and it is highly affected by the genetic background of the populations used for analysis.

The quantile–quantile plots from all the GWAS analyses using the MLM method implied that the population structure and family relatedness were well controlled, and the false-positive rates were further reduced (Fig. 3). In the same popu-

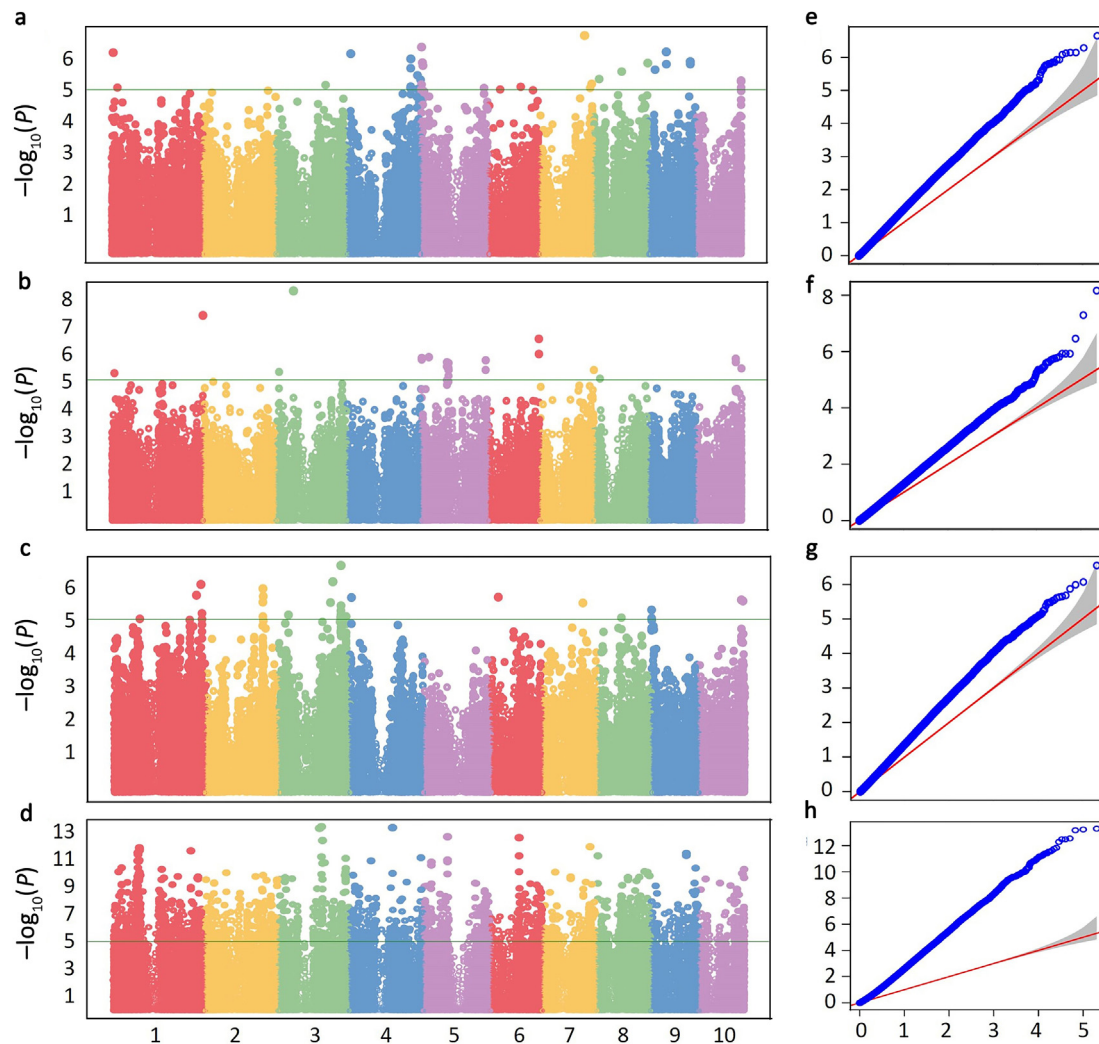


Fig. 2 – GWAS results for FER resistance using the GLM method and the Manhattan plots of the populations of (a) CML, (b) DTMA AM panel, (c) SYN_DH, and (d) all the 874 inbred lines. The horizontal axis shows the physical positions on the 10 maize chromosomes, the vertical axis shows the value of $-\log_{10}(P)$. The horizontal line indicates the genome-wide significance with a moderate stringent threshold of $-\log_{10}(1 \times 10^{-5})$. The quantile–quantile (Q-Q) plots of the populations of (e) CMLs, (f) DTMA AM panel, (g) SYN_DH, and (h) all the 874 inbred lines. The x-axis represents the expected $-\log_{10}(P)$, and the y-axis represents the observed $-\log_{10}(P)$.

lation, GWAS analysis detected fewer SNP associations and smaller PVE values using the MLM method than using the GLM method, which confirmed that FER resistance in maize was controlled by many genes with minor effects. Similar to the GLM method, the MLM method with all the 874 inbred lines detected more associations with lower PVE values than within the CMLs population and DTMA AM panel.

3.5. Putative candidate genes associated with FER resistance in maize detected by GWAS

According to the genome and annotation information of the B73 reference, GWAS analysis using the GLM method identified 1309 putative candidate genes that contained the SNPs significantly associated with FER resistance: 19 genes from the CMLs population, 11 genes from the DTMA AM panel, 13 genes from the SYN_DH population, and 1266 genes from

the population of all the 874 inbred lines. No overlapping candidate genes were observed among the CMLs population, DTMA AM panel, and SYN_DH population. The number of overlapping candidate genes with the population of all the 874 inbred lines were 7, 1, and 4 in the CMLs population, DTMA AM panel, and SYN_DH population, respectively (Fig. S3).

In total, GWAS analyses using the MLM method identified 39 putative candidate genes that either contained or were adjacent to the SNPs significantly associated with FER resistance in half of the linkage disequilibrium decay distance of 7.31 kb: 8 genes from the CMLs population, 4 genes from the DTMA AM panel, 14 genes from the SYN_DH population, and 13 genes from the population of all the 874 inbred lines. One candidate gene of GRMZM2G146020, transcription factor PosF21, shared between the SYN_DH population and the population of all the 874 inbred lines (Fig. S3). Several putative

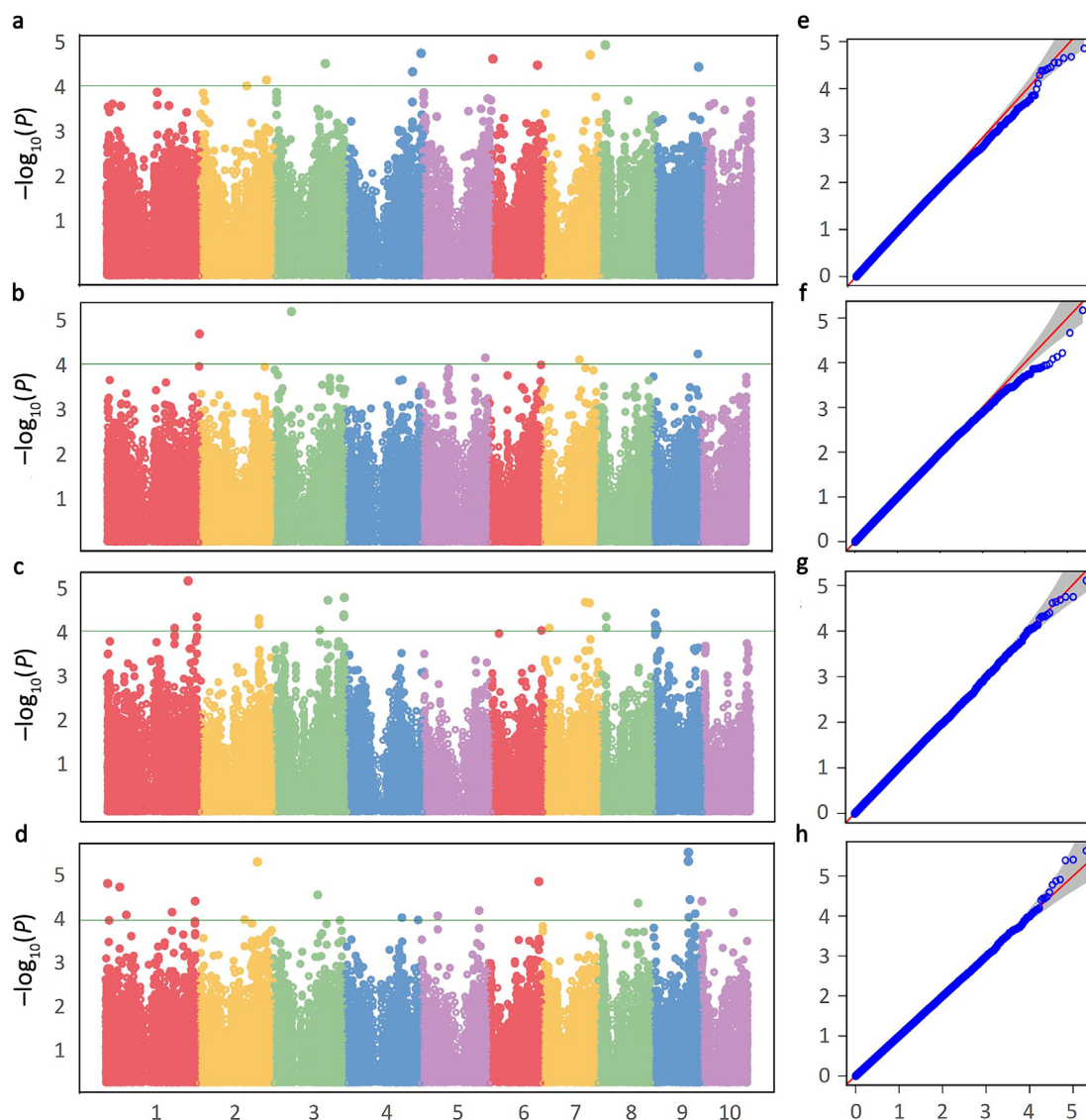


Fig. 3 – GWAS results for FER resistance using the MLM method and the Manhattan plots of the populations of (a) CMLs, (b) DTMA AM panel, (c) SYN_DH, and (d) all the 874 inbred lines. The horizontal axis shows the physical positions on the 10 maize chromosomes. The vertical axis shows the value of $-\log_{10}(P)$. The horizontal line indicates the genome-wide significance with a moderate stringent threshold of $-\log_{10}(1 \times 10^{-4})$. The quantile–quantile (Q-Q) plots of the populations of (e) CMLs, (f) DTMA AM panel, (g) SYN_DH, and (h) all the 874 inbred lines. The x-axis represents the expected $-\log_{10}(P)$, and the y-axis represents the observed $-\log_{10}(P)$.

candidate genes associated with FER resistance were of interest and include *GRMZM5G879570* in bin 7.03, *GRMZM2G025997* in bin 8.02, *GRMZM2G061314* in bin 9.07, and *GRMZM2G127416* in bin 10.03. *GRMZM5G879570* in bin 7.03 was detected from the SYN_DH population, which encodes a member of the RLCK VII-4 subfamily of receptor-like cytoplasmic kinases that have been shown to phosphorylate MAPKKK5 Ser-599 and MEKK1 Ser-603, both play important roles in PRR-mediated resistance to bacterial and fungal pathogens. *GRMZM2G025997* in bin 8.02 is a putative ring zinc finger domain superfamily protein, which was adjacent to the SNP association of S8_16910712 detected from the CMLs population, this SNP association had the lowest P-value and the

greatest PVE value in the GWAS analysis with MLM method. *GRMZM2G061314* in bin 9.07 detected from the DTMA AM panel encodes a receptor-like protein, which consists of a leucine-rich receptor-like repeat, a transmembrane region, and a short cytoplasmic region, with no kinase domain. This gene plays a possible role in fungal defense. *GRMZM2G127416* in bin 10.03 belongs to secondary cell wall glycosyltransferase family 47, which may be involved in cell and plant tissue development. Other putative candidate genes had gene function or expressing patterns showing a possible role in stress response, such as binding proteins including *GRMZM5G877773*, *GRMZM2G134976*, *GRMZM2G181390*, *GRMZM2G017400*, and *GRMZM2G154029* involved in posttranslational regulation;

Table 2 – Significantly associated SNPs and the annotation of the candidate genes revealed by the GWAS using the MLM method in the populations of CMLs, DTMA AM panel, and SYN_DH.

SNP Name	Bin	Allele	P-value	PVE (%)	MAF	Candidate gene	Annotation
CMLs							
S2_211168475	2.08	G/C	7.683E-05	6.34	0.12	GRMZM2G358381	Pentatricopeptide repeat-containing protein mitochondrial
S3_161730661	3.05	C/G	3.465E-05	6.97	0.12	GRMZM2G103812	Selenium-binding protein 3
S4_207486641	4.09	A/G	5.118E-05	6.66	0.13		
S4_234653907	4.09	G/C	2.088E-05	7.38	0.47	GRMZM2G073059	Cyclase
S6_146700213	6.05	G/A	3.713E-05	6.92	0.11		
S6_4350527	6.01	A/G	2.763E-05	7.15	0.19		
S6_4350528	6.01	T/C	2.763E-05	7.15	0.19		
S7_145770513	7.03	C/T	2.248E-05	7.32	0.10	GRMZM2G485559	Blue copper protein
						GRMZM2G006117	Transmembrane 9 superfamily member 9
						GRMZM2G006416	Probable protein phosphatase 2C 21
S8_16910712	8.02	G/A	1.39E-05	7.71	0.22	GRMZM2G025997	Putative ring zinc finger domain superfamily protein
S9_138615621	9.06	T/G	4.093E-05	6.84	0.18	GRMZM2G095727	Two-component response regulator-like APRR3
						GRMZM2G095868	Ypt/Rab-GAP domain of gyp1p superfamily protein
S9_138615643	9.06	A/C	4.093E-05	6.84	0.18	GRMZM2G095727	Two-component response regulator-like APRR3
						GRMZM2G095868	Ypt/Rab-GAP domain of gyp1p superfamily protein
S9_138615701	9.06	C/T	3.991E-05	6.86	0.19	GRMZM2G095727	Two-component response regulator-like APRR3
						GRMZM2G095868	Ypt/Rab-GAP domain of gyp1p superfamily protein
DTMA AM panel							
S1_299329113	1.12	G/A	2.126E-05	5.29	0.20		
S3_56477054	3.04	T/A	6.726E-06	5.96	0.06	GRMZM2G409309	
S5_206046656	5.07	T/G	7.381E-05	4.58	0.18	GRMZM5G877773	NAD(P)-binding Rossmann-fold superfamily protein
S7_121545012	7.02	T/A	8.185E-05	4.52	0.11		
S9_149540390	9.07	G/C	5.998E-05	4.70	0.21	GRMZM2G061280	Ras-related protein RABA4a
						GRMZM2G061314	Receptor Like Proteins consists of a leucine-rich receptor-like repeat, a transmembrane region, and a short cytoplasmic region, with no kinase domain
SYN_DH							
S1_217252370	1.07	T/C	8.65E-05	4.53	0.37		
S1_260142965	1.09	T/C	7.798E-06	5.91	0.10	GRMZM2G134976	Myosin-binding protein 7
S1_288742469	1.11	T/C	4.83E-05	4.86	0.48	GRMZM2G163912	Protein WVD2-like 3
S1_288742499	1.11	C/T	8.49E-05	4.54	0.48	GRMZM2G163912	Protein WVD2-like 3
S2_185868849	2.06	G/C	5.36E-05	4.80	0.27		
S2_185869223	2.06	G/C	7.24E-05	4.63	0.34		
S3_142736417	3.05	T/C	9.49E-05	4.47	0.20	GRMZM2G146020	Transcription factor PosF21
S3_168922004	3.06	G/T	2.052E-05	5.35	0.23	GRMZM2G035092	MADS-box transcription factor family protein
						GRMZM5G807064	BTB/POZ and TAZ domain-containing protein 3

SNP Name	Bin	Allele	P-value	PVE (%)	MAF	Candidate gene	Annotation
S3_219293102	3.09	C/T	4.327E-05	4.92	0.09	GRMZM2G047813	Histone H2A
S3_219398619	3.09	G/A	4.74E-05	4.87	0.11	GRMZM2G061890	RNA polymerase III subunit RPC82 family protein
						GRMZM2G061723	IQ domain-containing protein IQM5
S3_220645282	3.09	T/C	1.78E-05	5.43	0.44	GRMZM2G130356	Alpha/beta-Hydrolases superfamily protein
S3_220645287	3.09	A/G	1.78E-05	5.43	0.44	GRMZM2G130356	Alpha/beta-Hydrolases superfamily protein
S6_159221992	6.06	G/A	9.613E-05	4.47	0.34		
S7_130321001	7.03	G/A	2.29E-05	5.29	0.09	GRMZM2G181390	Movement protein binding protein 2C
S7_143397458	7.03	C/G	2.39E-05	5.26	0.06	GRMZM5G879570	Probable receptor-like protein kinase
S7_15155655	7.02	C/G	8.67E-05	4.52	0.12		
S8_20516145	8.02	A/G	4.78E-05	4.86	0.32	GRMZM2G049695	Putative MYB DNA-binding domain superfamily protein
S8_20516249	8.02	T/A	8.41E-05	4.54	0.36	GRMZM2G049695	Putative MYB DNA-binding domain superfamily protein
S9_1590568	9.00	T/C	7.81E-05	4.58	0.19	GRMZM2G000587	Probable galacturonosyltransferase-like 9
S9_1590571	9.00	G/C	7.35E-05	4.62	0.19	GRMZM2G000587	Probable galacturonosyltransferase-like 9
S9_1590608	9.00	C/T	3.97E-05	4.97	0.17	GRMZM2G000587	Probable galacturonosyltransferase-like 9
S9_7585423	9.01	A/G	9.44E-05	4.48	0.32	GRMZM2G120302	Aap1; acylaminoacyl-peptidase1: endospERM cDNA 5C09A06, similar to acylaminoacyl-peptidase

SNP name, chromosome position, for example, S2_211168475 means the SNP is on chromosome 2, the physical position is 211168475 bp.

PVE, phenotypic variation explained; MAF, minor allele frequency.

transcription factors including GRMZM2G146020 and GRMZM2G035092 involved with signal recognition.

3.6. Genomic prediction accuracies estimated from the five-fold cross-validation schemes

The prediction accuracies of FER resistance estimated from the five-fold cross-validation schemes are shown in Fig. 4. In each population studied in the present study, the prediction accuracy estimated with the genome-wide markers in the CombinedENV is moderate, which is higher than those estimated within the individual environment. In the CombinedEnv, the prediction accuracies estimated with genome-wide markers in the CMLs population, DTMA AM panel, SYN_DH population, and all the 874 inbred lines were 0.46, 0.53, 0.32, and 0.57, respectively. In the single-environment analyses, the prediction accuracies estimated with the genome-wide markers ranged from 0.25 to 0.42 in the CMLs population within the six individual environments, from 0.28 to 0.42 in the DTMA AM panel within the four individual environments, and from 0.17 to 0.30 in the SYN_DH population within the four individual environments.

3.7. Genomic prediction accuracies estimated with the SNPs significantly associated with FER resistance

The prediction accuracies of FER resistance estimated with the significantly associated SNPs in all the populations were greater than 0.50 in the CombinedEnv analyses, which were relatively higher than those estimated with the genome-wide markers within each population (Fig. 4). When the significantly associated SNPs detected in each population were used for predicting the same population, the prediction accuracy of FER resistance in the CMLs population, DTMA AM panel, SYN_DH population, and all the 874 inbred lines were 0.74, 0.62, 0.63, and 0.65, respectively. In total, 3108 unique significantly associated SNPs were detected in the GLM model in all the GWAS analyses across all the populations. The prediction accuracy of FER resistance estimated with these 3108 significantly associated SNPs in the CMLs population, DTMA AM panel, SYN_DH population, and all the 874 inbred lines were 0.66, 0.66, 0.55, and 0.67, respectively. Within the same population, similar prediction accuracies were observed in the two different types of significantly associated SNP datasets.

3.8. Genomic prediction accuracies estimation when the training and validation sets were independent

The prediction accuracies estimated in the validation sets are shown in Table 4, when the training and validation sets were independent, and when one or two of the populations were used as the training set to predict the other population as the validation set. The prediction accuracies between the CMLs population and the DTMA AM panel were moderate, which were 0.45 using the CMLs population as the training set to predict the DTMA AM panel as the validation set, and vice versa. The prediction accuracies were low and ranged from 0.04 to 0.21 between the SYN_DH population and the other two populations. The prediction accuracy in the SYN_DH population was 0.19 and 0.04, when the SYN_DH

Table 3 – Significantly associated SNPs and the annotation of the candidate genes revealed by the GWAS using the MLM method in the population comprising all the 874 inbred lines.

SNP Name	Bin	Allele	P-value	PVE (%)	MAF	Candidate gene	Annotation
S1_215350780	1.07	G/C	6.80258E-05	1.45	0.21		
S1_289064255	1.11	A/G	3.62493E-05	1.56	0.16	GRMZM2G072156	Cactin
S1_47993337	1.03	G/A	1.65098E-05	1.69	0.37	GRMZM2G017536 GRMZM2G017400	DUF4378 domain protein Calmodulin-binding protein
S1_69095566	1.04	G/C	7.90317E-05	1.42	0.08		
S1_9355044	1.05	A/G	1.32781E-05	1.73	0.19	GRMZM5G860938	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
S2_187053202	2.06	G/A	4.01302E-06	1.94	0.28	GRMZM2G066516	Ternary complex factor MIP1-like protein
S3_142736417	3.05	T/C	2.5078E-05	1.62	0.26	GRMZM2G146020	Transcription factor PosF21
S4_180206370	4.08	G/A	9.2945E-05	1.39	0.20		
S5_185646038	5.05	C/G	6.26077E-05	1.46	0.30	GRMZM2G028830	RmlC-like cupins superfamily protein
S5_53522776	5.03	T/A	8.17344E-05	1.41	0.49		
S6_159221992	6.06	G/A	1.20711E-05	1.75	0.46		
S8_130513588	8.05	G/A	3.96428E-05	1.54	0.26	GRMZM2G155889	DUF724 domain-containing protein 2
S9_115952569	9.04	C/A	3.89388E-06	1.95	0.30		
S9_115952576	9.04	T/G	2.32544E-06	2.04	0.31		
S9_117051071	9.04	C/T	9.0096E-05	1.40	0.49		
S9_120747394	9.04	C/T	3.29508E-05	1.57	0.43	GRMZM2G154029	Putative HLH DNA-binding domain superfamily protein isoform 1%3B; Putative HLH DNA-binding domain superfamily protein isoform 2
S9_137847927	9.06	A/C	7.40253E-05	1.43	0.15	GRMZM2G128971	Fatty acid desaturase 7
S10_103085578	10.04	C/T	6.93807E-05	1.44	0.11		
S10_2656869	10.03	G/A	3.607E-05	1.56	0.22	GRMZM2G127416 GRMZM2G127393 GRMZM2G427014	Secondary cell wall glycosyltransferase family; Trigger factor-like protein TIG Chloroplastic 40S ribosomal protein S16

SNP name, chromosome_position, for example, S1_215350780 means the SNP is on chromosome 1, the physical position is 215350780 bp. PVE, phenotypic variation explained; MAF, minor allele frequency.

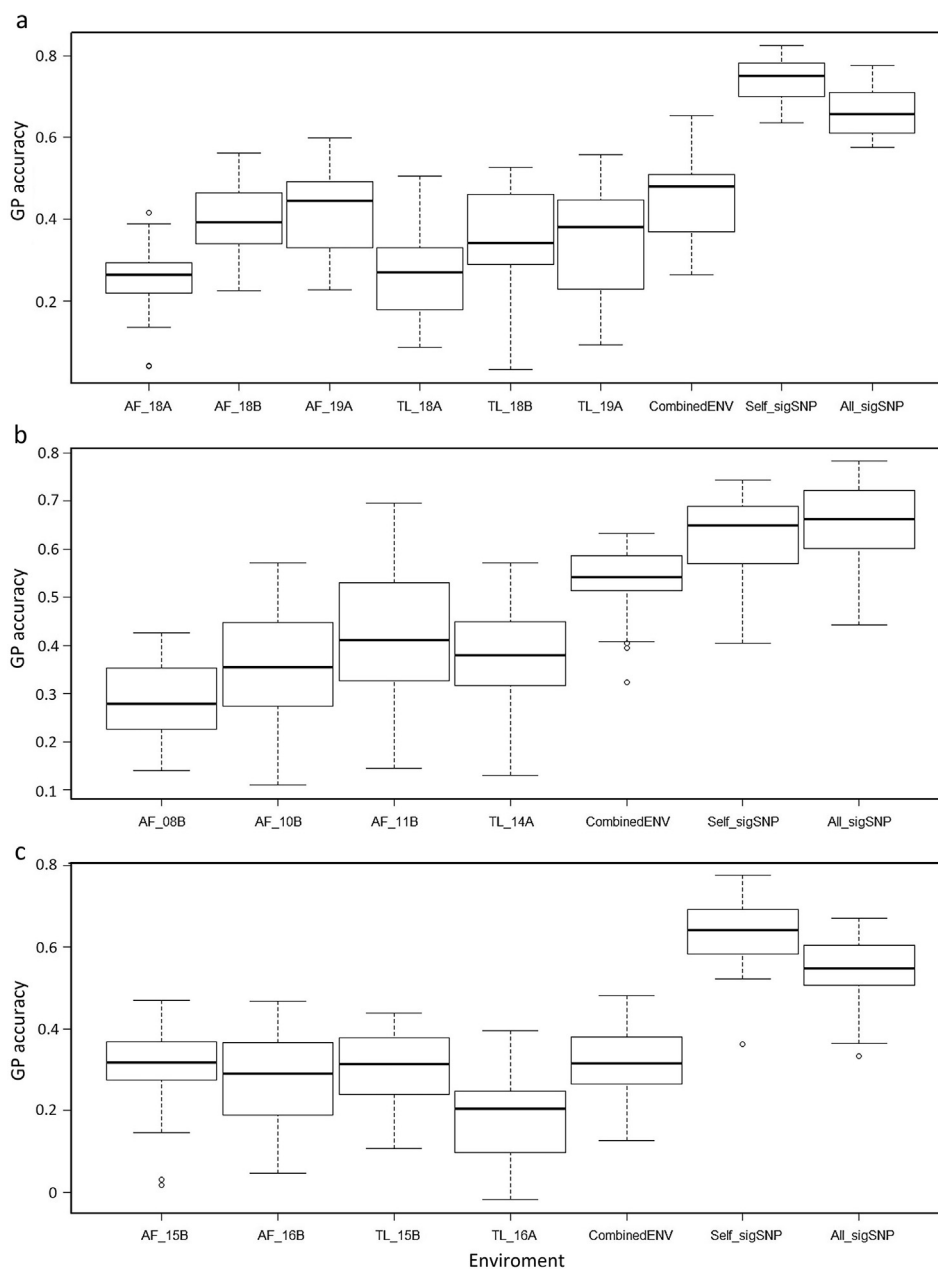


Fig. 4 – Genomic prediction (GP) accuracies for FER resistance estimated from the five-fold cross-validation schemes in the populations of (a) CMLs, (b) DTMA AM panel, and (c) SYN_DH with the genome-wide markers in each environment and the combined-environment (CombinedENV) analyses, and with the significantly associated SNPs detected in each population (Self_sigSNP) and across all the populations (All_sigSNP) in the combined-environment analyses.

population was predicted by the CMLs population and the DTMA AM panel, respectively. The prediction accuracy was 0.21 in the CMLs population and 0.10 in the DTMA AM panel, when the SYN_DH population was used as the training set to do prediction. The prediction accuracies in the CMLs population, DTMA AM panel, and SYN_DH population was 0.52, 0.45, and 0.17, when the other two populations were used as the training sets. Higher prediction accuracies were observed between the CMLs population and the DTMA AM panel, and lower prediction accuracies were observed between the SYN_DH population and the other two populations. This

result is in line with the population structure and genetic relatedness analyses.

4. Discussion

Previous studies showed that most stable sources of resistance to FER are mainly from tropical maize germplasm [6,7]. In the present study, a total of 874 tropical and subtropical CIMMYT maize inbred lines were phenotyped in multiple-location trials for screening their FER responses. Broad phenotypic variation of disease severity was observed, and sev-

Table 4 – The genomic prediction accuracies estimated in the validation sets, when the training and validation sets were independent, and one or two of the populations were used as the training set to predict the other population as the validation set.

Training set	Validation set	Prediction accuracy
CMLs	DTMA AM panel	0.45
CMLs	SYN_DH	0.19
DTMA AM panel	CMLs	0.45
DTMA AM panel	SYN_DH	0.04
SYN_DH	CMLs	0.21
SYN_DH	DTMA AM panel	0.10
CMLs & DTMA AM panel	SYN_DH	0.17
CMLs & SYN_DH	DTMA AM panel	0.45
DTMA AM panel & SYN_DH	CMLs	0.52

eral donor lines showing stable FER resistance were identified. In all the multiple-location trials, the heritabilities of FER resistance estimated from the replicated trials are high in both the single-environment and CombinedENV analyses, revealing that phenotypic selection is effective for improving FER resistance, and the phenotypic data was reliable for further genetic analyses to identify favorable alleles associated with improved FER resistance for implementing MAS. However, it is still difficult to effectively incorporate the resistance alleles from the tropical and subtropical donor lines to the elite breeding lines through phenotypic selection, because the heritability of individual plants against FER is relatively low, and resistance to FER in maize is highly influenced by large genotype-by-environment interactions [1]. The results of this study confirmed that most of the Pearson's correlation coefficients among environments were low to moderate in all the populations, and the component of genotype-by-environment variance was significant at P -value of 0.01 in all the ANOVA analyses, which indicated that multiple location trials are required to improve FER resistance through the phenotypic selection to eliminate the effect of genotype-by-environment interaction. However, improving FER resistance through phenotypic selection could be expensive and time-consuming.

GWAS is a powerful strategy for genetic dissection of complex traits in plants. In this study, GWAS was conducted with both GLM and MLM methods to identify the specific genomic regions and SNPs conferring FER resistance in tropical maize germplasm. The results showed that resistance to FER in maize is controlled by several QTL with small effects, and is highly influenced by the genetic background of the populations studied. The GLM method identified 3034 SNPs significantly associated with FER resistance in all the 874 inbred lines at the P -value threshold of 1×10^{-5} , and 604 of the 3034 association had P -values lower than 1×10^{-7} . The average PVE value of the 3034 associations was 3%, and only 49 associations had a PVE value greater than 5%. The MLM method identified 19 SNPs significantly associated with FER resistance in all the 874 inbred lines at the P -value threshold of 1×10^{-4} , the average PVE value of the 19 associations was 1.60% with a range from 1.39% to 2.04%. Both GLM and MLM methods revealed a few overlapping SNP associations among the populations studied. No overlapping SNP association was observed between the CMLs population and the DTMA AM panel, even through the

genetic relatedness between these two populations was not far, it implied that the statistical power of GWAS should be increased furtherly to map the stable minor QTL with small effects across populations. These results are consistent with the observations of previous studies [34], which implied that MAS for improving FER resistance may not be very effective.

Several GWAS analyses have been conducted in maize to detect genomic regions conferring FER resistance [13,14,34], and a few stable genomic regions were identified across these studies, although these genomic regions have minor effects. Previous and present studies indicate that bins 3.04/05, 7.02/04, 9.06/07, and 10.03/04 were enriched in SNPs significantly associated with FER resistance. In the GWAS analyses with the MLM method, the SNPs in bin 3.04/3.05 significantly associated with FER resistance were revealed across all the four populations used in the present study. In the DTMA AM panel, the association of SNP S3_56477054 at bin 3.04 had the lowest P -value and the greatest PVE value. In the SYN_DH population, the greatest number of SNP associations were detected on chromosome 3, including S3_142736417 in bin 3.05. In all the 874 inbred lines, the SNP association of S3_142736417 was detected as well, which was one of the two overlapping SNP associations observed between the SYN-DH population and in all the 874 inbred lines. These results show that the genomic region in bin 3.04/05 is stable, which is consistent with the observations of the previous studies [7,13,14]. The genomic region in bin 7.02/04 conferring the FER resistance was reported previously [12,13], the results of this study confirm the previous observations. The genomic region associated with FER resistance in the CMLs population, DTMA AM panel, and SYN_DH population was in bin 7.03, 7.02, and 7.02/03, respectively. The genomic region at bin 9.06/07 associated with FER resistance was previously reported [14,34]. The genomic region in bin 9.06/07 associated with FER resistance was observed in this study in all the four populations, except for the SYN_DH population. The genomic region in bin 10.03/04 associated with FER resistance was detected in the SYN_DH population and all the 874 inbred lines, which was also reported in previous studies [13,14,34]. Besides, the genomic region in bin 9.00/01 detected in the SYN_DH population and the genomic regions in bin 9.04 detected in all the 874 inbred lines are new and stable. Thus, these results indicate that it could be effective to improve the FER resistance in maize by selecting multiple stable genomic regions simultaneously.

Previously published studies and the present research show that the detected genomic regions have minor genetic effects and low frequencies of minor alleles, and the overlapping of genomic regions across different studies is rare. These results indicated that the statistical power of genetic mapping should be increased, and the genetic mapping resolution should be improved in further GWAS analyses. In this study, GWAS analyses using both GLM and MLM methods detected more associations at lower *P*-values in all the 874 inbred lines, although the PVE values of these associations were smaller, indicating that the statistical power of genetic mapping can be dramatically increased by enlarging the population size. Besides increasing the statistical power of genetic mapping, the false-positive rate in the GWAS also should be controlled. The quantile–quantile plots implied that the GWAS using the GLM method did not adequately control the false-positive rates, and a high number of false-positive associations were detected. Therefore, the genomic regions and candidate genes containing the SNP associations detected from the GLM method were not further explored, which were mainly used for further GP analyses. In contrast, the quantile–quantile plots implied that the GWAS analyses using the MLM method did not well control the false-negative rates, SNPs with lower significance may not have been detected, but this should not affect the identification of SNPs and genomic regions significantly associated with FER resistance.

Multiparent Advanced Generation Inter-crossed (MAGIC) population and Complete-diallel design plus Unbalanced Breeding-like Inter-Cross (CUBIC) population have proved to be a valuable method to improve the statistical power of genetic mapping. This is mostly due to the high phenotypic diversity, enrichment of favorable alleles with low frequency, clear population structure pattern, and rapid linkage disequilibrium decay [12,35]. The SYN_DH population used in this study is a multi-parental synthetic population, which is similar to the MAGIC and CUBIC populations. The analyses of GWAS with MLM method detected more significantly associated SNPs in the SYN_DH population than in all the 874 inbred lines, showing that the statistical mapping power was improved in the SYN_DH population. Besides, several SNP associations detected in the SYN_DH population have relatively low MAF, i.e. S1_260142965, S3_219293102, S7_130321001, and S7_143397458 having MAF less than 0.10, implying that SYN_DH population has the advantage to detect rare alleles associated with the target trait. Moreover, the significantly associated SNPs detected in the SYN_DH populations have no overlapping with those detected in the CMLs population and DTMA AM panel, indicating the presence of unique phenotypic and genotypic diversity of the SYN_DH population, and the low level of genetic relatedness presented between the SYN_DH population and the other two populations.

GS is effective for the improvement of complex traits in maize [16,17]. In each population studied in the present study, the prediction accuracies of FER resistance estimated with the genome-wide markers in the CombinedENV analyses were moderate, and the prediction accuracies estimated in the CombinedENV analyses were higher than those estimated within the individual environment. These results show that GS is a promising tool for improving FER resistance in maize.

However, the training set for predicting FER performance needs to be phenotyped in multi-location trials to eliminate the effect of large genotype-by-environment interaction to achieve good prediction accuracy. The prediction accuracies of FER resistance estimated with the significantly associated SNPs were greater than those estimated with the genome-wide markers across all the studied populations, suggesting that GWAS information helped in improving the prediction accuracy. When the training and validation sets were independent, higher prediction accuracies were observed between the CMLs population and the DTMA AM panel, and lower prediction accuracies were observed between the SYN_DH population and the other two populations, indicating that the closely related training and validation sets produced higher prediction accuracy. However, increasing the population size of the training set by adding distantly related materials does not help to improve prediction accuracy. When the SYN_DH population was incorporated into the training set, the prediction accuracies observed between the CMLs population and the DTMA AM panel did not change or slightly increased. The training set comprising of both populations produced similar prediction accuracy, as either the CMLs population or the DTMA panel was used as the training set to predict the SYN_DH population.

Declaration of competing interest

Authors declare that there are no conflicts of interest.

Acknowledgements

The authors gratefully acknowledge the financial support from the MasAgro project funded by Mexico's Secretary of Agriculture and Rural Development (SADER), the Genomic Open-source Breeding Informatics Initiative (GOBII) (grant number OPP1093167) supported by the Bill & Melinda Gates Foundation, and the CGIAR Research Program (CRP) on maize (MAIZE). MAIZE receives W1&W2 support from the Governments of Australia, Belgium, Canada, China, France, India, Japan, the Republic of Korea, Mexico, Netherlands, New Zealand, Norway, Sweden, Switzerland, the United Kingdom, USA, and the World Bank. The authors also thank the National Natural Science Foundation of China (grant number 31801442), the CIMMYT–China Specialty Maize Research Center Project funded by the Shanghai Municipal Finance Bureau, and the China Scholarship Council.

Appendix A. Supplementary data

Supplementary data this article can be found online at <https://doi.org/10.1016/j.cj.2020.08.008>.

REFERENCES

- [1] Gaikpa DS, Miedaner T. Genomics-assisted breeding for ear rot resistances and reduced mycotoxin contamination in maize: methods, advances and prospects. *Theor. Appl. Genet.* 2019;132:2721–39.

- [2] Logrieco A, Mulè G, Moretti A, Bottalico A. Toxigenic *Fusarium* species and mycotoxins associated with maize ear rot in Europe. *Eur. J. Plant Pathol.* 2002;108:597–609.
- [3] Missmer SA, Suarez L, Felkner M, Wang E, Merrill AH, Rothman KJ, Hendricks KA. Exposure to fumonisins and the occurrence of neutral tube defects along the Texas-Mexico border. *Environ. Health Perspect.* 2006;114:237–41.
- [4] Munkvold GP. Epidemiology of *Fusarium* diseases and their mycotoxins in maize ears. *Eur. J. Plant. Pathol.* 2003;109:705–13.
- [5] Mesterházy Á, Lemmens M, Reid LM. Breeding for resistance to ear rots caused by *Fusarium* spp. in maize—a review. *Plant Breed.* 2012;131:1–19.
- [6] Zila CT, Ogut F, Romay MC, Gardner CA, Buckler ES, Holland JB. Genome-wide association study of *Fusarium* ear rot disease in the U.S.A. maize inbred line collection. *BMC Plant Biol.* 2014;14:372.
- [7] Chen J, Shrestha R, Ding J, Zheng H, Mu C, Wu J, Mahuku G. Genome-wide association study and QTL mapping reveal genomic loci associated with *Fusarium* ear rot resistance in tropical maize germplasm. *G3-Genes Genomes Genet.* 2016;6:3803–15.
- [8] Chen J, Ding J, Li H, Li Z, Sun X, Li J, Wang R, Dai X, Dong H, Song W, Chen W, Xia Z, Wu J. Detection and verification of quantitative trait loci for resistance to *Fusarium* ear rot in maize. *Mol. Breeding* 2012;30:1649–56.
- [9] Ding J-Q, Wang X-M, Chander S, Yan J-B, Li J-S. QTL mapping of resistance to *Fusarium* ear rot using a RIL population in maize. *Mol. Breeding* 2008;22:395–403.
- [10] Maschietto V, Colombi C, Pirona R, Pea G, Strozzi F, Marocco A, Rossini L, Lanubile A. QTL mapping and candidate genes for resistance to *Fusarium* ear rot and fumonisin contamination in maize. *BMC Plant Biol.* 2017;17:20.
- [11] Yan J, Warburton M, Crouch J. Association mapping for enhancing maize (*Zea mays* L.) genetic improvement. *Crop Sci.* 2011;51:433–49.
- [12] Butrón A, Santiago R, Cao A, Samayoa LF, Malvar RA. QTLs for resistance to *Fusarium* ear rot in a multiparent advanced generation intercross (MAGIC) maize population. *Plant Dis.* 2019;103:897–904.
- [13] Coan MMD, Senhorinho HJC, Pinto RJB, Scapim CA, Tessmann DJ, Williams WP, Warburton ML. Genome-wide association study of resistance to ear rot by *Fusarium verticillioides* in a tropical field maize and popcorn core collection. *Crop Sci.* 2018;58:564–78.
- [14] Yao L, Li Y, Ma C, Tong L, Du F, Xu M. Combined genome-wide association study and transcriptome analysis reveal candidate genes for resistance to *Fusarium* ear rot in maize. *J. Integr. Plant Biol.* 2020;62:1535–51.
- [15] Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 2001;157:1819–29.
- [16] Crossa J, Pérez P, Hickey J, Burgueño J, Ornella L, Cerón-Rojas J, Zhang X, Dreisigacker S, Babu R, Li Y, Bonnett D, Mathews K. Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity* 2014;112:48–60.
- [17] Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de los Campos G, Burgueño J, González-Camacho JM, Pérez-Elizalde S, Beyene Y, Dreisigacker S, Singh R, Zhang X, Gowda M, Roorkiwal M, Rutkoski J, Varshney RK. Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 2017;22:961–75.
- [18] Gowda M, Das B, Makumbi D, Babu R, Semagn K, Mahuku G, Olsen MS, Bright JM, Beyene Y, Prasanna BM. Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. *Theor. Appl. Genet.* 2015;128:1957–68.
- [19] Sitonik C, Suresh LM, Beyene Y, Olsen MS, Makumbi D, Oliver K, Das B, Bright JM, Mugo S, Crossa J, Tarekegne A, Prasanna BM, Gowda M. Genetic architecture of maize chlorotic mottle virus and maize lethal necrosis through GWAS, linkage analysis and genomic prediction in tropical maize germplasm. *Theor. Appl. Genet.* 2019;132:2381–99.
- [20] Cao S, Loladze A, Yuan Y, Wu Y, Zhang A, Chen J, Huestis G, Cao J, Chaikam V, Olsen M, Prasanna BM, San Vicente F, Zhang X. Genome-wide analysis of tar spot complex resistance in maize using genotyping-by-sequencing SNPs and whole-genome prediction. *Plant Genome* 2017;10. <https://doi.org/10.3835/plantgenome2016.10.0099>.
- [21] Han S, Miedaner T, Utz HF, Schipprack W, Schrag TA, Melchinger AE. Genomic prediction and GWAS of Gibberella ear rot resistance traits in dent and flint lines of a public maize breeding program. *Euphytica* 2018;214:1–20.
- [22] Bian Y, Holland JB. Enhancing genomic prediction with genome-wide association studies in multiparental maize populations. *Heredity* 2017;118:585–93.
- [23] Liu X, Wang H, Wang H, Guo Z, Xu X, Liu J, Wang S, Li W-X, Zou C, Prasanna BM, Olsen MS, Huang C, Xu Y. Factors affecting genomic selection revealed by empirical evidence in maize. *Crop J.* 2018;6:341–52.
- [24] Wang N, Liu B, Liang X, Zhou Y, Song J, Yang J, Yong H, Weng J, Zhang D, Li M, Nair S, Vicente FS, Hao Z, Zhang X, Li X. Genome-wide association study and genomic prediction analyses of drought stress tolerance in China in a collection of off-PVP maize inbred lines. *Mol. Breed.* 2019;39:113.
- [25] Wu Y, San Vicente F, Huang K, Dhliwayo T, Costich DE, Semagn K, Sudha N, Olsen M, Prasanna BM, Zhang X, Babu R. Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. *Theor. Appl. Genet.* 2016;129:753–65.
- [26] Yuan Y, Cairns JE, Babu R, Gowda M, Makumbi D, Magorokosho C, Zhang A, Liu Y, Wang N, Hao ZF, Vicente FS, Olsen MS, Prasanna BM, Lu Y, Zhang X. Genome-wide association mapping and genomic prediction analyses reveal the genetic architecture of grain yield and flowering time under drought and heat stress conditions in maize. *Front. Plant Sci.* 2019;9:1919.
- [27] Drepper WJ. Comparison of methods for inoculation of ears and stalks of maize with *Fusarium moniliforme*. *Plant Dis.* 1990;74:952–6.
- [28] Grau CR. Resistance of soybean cultivars to *Sclerotinia sclerotiorum*. *Plant Dis.* 1982;66:506–8.
- [29] Alvarado G, Rodríguez FM, Pacheco A, Burgueño J, Crossa J, Vargas M, Pérez-Rodríguez P, Lopez-Cruz MA. META-R: A software to analyze data from multi-environment plant breeding trials. *Crop J.* 2020;8:745–56.
- [30] Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 2011;6. e19379
- [31] Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z. GAPIT: genome association and prediction integrated tool. *Bioinformatics* 2012;28:2397–9.
- [32] R Core Team, R: A language and environment for statistical computing, <https://www.R-project.org/>, 2019 (Accessed April 1, 2019).
- [33] Pérez P, de los Campos G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 2014;198:483–95.
- [34] Guo Z, Zou C, Liu X, Wang S, Li W-X, Jeffers D, Fan X, Xu M, Xu Y. Complex genetic system involved in *Fusarium* ear rot

resistance in maize as revealed by GWAS, bulked sample analysis, and genomic prediction. *Plant Dis.* 2020. <https://doi.org/10.1094/pdis-07-19-1552-re>.

- [35] Liu HJ, Wang X, Xiao Y, Luo J, Qiao F, Yang W, Zhang R, Meng Y, Sun J, Yan S, Peng Y, Niu L, Jian L, Song W, Yan J, Li C, Zhao Y,

Liu Y, Warburton ML, Zhao J, Yan J. CUBIC: an atlas of genetic architecture promises directed maize improvement. *Genome Biol.* 2020;21:20.