



# Exploiting genomic tools for genetic dissection and improving the resistance to Fusarium stalk rot in tropical maize

Junqiao Song<sup>1,2,3</sup> · Yubo Liu<sup>2,4</sup> · Rui Guo<sup>2,5</sup> · Angela Pacheco<sup>2</sup> · Carlos Muñoz-Zavala<sup>2</sup> · Wei Song<sup>2,5</sup> · Hui Wang<sup>2,4</sup> · Shiliang Cao<sup>2,6</sup> · Guanghui Hu<sup>2,6</sup> · Hongjian Zheng<sup>4</sup> · Thanda Dhliwayo<sup>2</sup> · Felix San Vicente<sup>2</sup> · Boddupalli M. Prasanna<sup>7</sup> · Chunping Wang<sup>1</sup> · Xuecai Zhang<sup>2,8,9</sup>

Received: 18 December 2023 / Accepted: 7 March 2024 / Published online: 22 April 2024  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

## Abstract

**Key message** A stable genomic region conferring FSR resistance at ~250 Mb on chromosome 1 was identified by GWAS. Genomic prediction has the potential to improve FSR resistance.

**Abstract** Fusarium stalk rot (FSR) is a global destructive disease in maize; the efficiency of phenotypic selection for improving FSR resistance was low. Novel genomic tools of genome-wide association study (GWAS) and genomic prediction (GP) provide an opportunity for genetic dissection and improving FSR resistance. In this study, GWAS and GP analyses were performed on 562 tropical maize inbred lines consisting of two populations. In total, 15 SNPs significantly associated with FSR resistance were identified across two populations and the combined POP consisting of all 562 inbred lines, with the  $P$ -values ranging from  $1.99 \times 10^{-7}$  to  $8.27 \times 10^{-13}$ , and the phenotypic variance explained (PVE) values ranging from 0.94 to 8.30%. The genetic effects of the 15 favorable alleles ranged from  $-4.29$  to  $-14.21\%$  of the FSR severity. One stable genomic region at ~250 Mb on chromosome 1 was detected across all populations, and the PVE values of the SNPs detected in this region ranged from 2.16 to 5.18%. Prediction accuracies of FSR severity estimated with the genome-wide SNPs were moderate and ranged from 0.29 to 0.51. By incorporating genotype-by-environment interaction, prediction accuracies were improved between 0.36 and 0.55 in different breeding scenarios. Considering both the genome coverage and the threshold of the  $P$ -value of SNPs to select a subset of molecular markers further improved the prediction accuracies. These findings extend the knowledge of exploiting genomic tools for genetic dissection and improving FSR resistance in tropical maize.

## Introduction

The Food and Agriculture Organization (FAO) estimates that the world's population will surpass 9 billion by 2050 (Nations and United Nations. 2019). Fulfilling the food and feed demand, the average genetic gain per year of maize should be accelerated to more than 2%, which is a big challenge under the effect of climate change (Prasanna et al. 2021). With the change in precipitation, temperature, and humidity, crop diseases become the key factor that affects genetic gain. Stalk rot and ear rot are the two crucial major maize diseases having the highest impact on climate change (Prasanna et al. 2021).

Stalk rot, one complex fungal disease, can be caused by *Fusarium verticillioides* (*F.v.*), *F. graminearum* (*Gibberella*), *Colletotrichum graminicola* (*Anthracnose*), and *Pythium aphanidermatum*, as well as some bacterial species of *Erwinia* with similar symptoms caused by *Fusarium spp.* (Chambers 1987). Fusarium stalk rot (FSR), caused by *F.v.*, is one of the most disastrous diseases worldwide, especially in the tropical and subtropical zones (Savary et al. 2019; Chivasa et al. 2021). In South and Central America, the incidence of FSR is usually above 50% (Christensen et al. 2014). Generally, the incidence of FSR ranges from 30 to 70%, and it could surge to 90% in some specific years in India, China, and the Philippines (Duan et al. 2019). FSR can cause 38–100% yield loss of maize; furthermore, it can produce low molecular-weight secondary metabolites known as mycotoxins in the grain and the plant, bringing fatal harm to humans and other animals (Maize AICRP 2014; Subedi et al. 2016; Mueller et al. 2022). As one of the most aggressive pathogens, *F.v.* can infect any part of maize from the

---

Communicated by Mingliang Xu.

---

Junqiao Song and Yubo Liu contribute equally to this work and share the first authorship.

---

Extended author information available on the last page of the article

beginning to the end of the cropping season, and it keeps alive on the residue of maize or the rotation crops during winter (Munkvold 2003; White 1999). Currently, the efficiency of controlling FSR disease in the field was relatively low, because the effective fungicides to control FSR disease are rare (Zhu et al. 2021; Holland et al. 2020). Alternatively, development and deployment of maize varieties with genetic resistance to FSR are the most cost-effective and environment-friendly approach.

Novel genetic tools, such as GWAS (Genome-wide association study), provide the opportunity to dissect the genetic architecture of FSR resistance for the development of maize varieties with genetic resistance to FSR more rapid and effective. Various genetic studies had reported that FSR resistance is a complex quantitative inherited trait, hundreds of QTL (quantitative trait loci) and genomic regions associated with resistance to stalk rot caused by different pathogens have been reported, including *qRfg1* (Yang et al. 2010; Wang et al. 2017), *qRfg2* (Zhang et al. 2012; Ye et al. 2018), *qRfg3* (Ma et al. 2017) and *Rgsr8.1* (Chen et al. 2017) for *Gibberella* stalk rot, *Rpi1* (Yang et al. 2005), *RpiQI319-1*, *RpiQI319-2* (Song et al. 2015), *RpiX178-1* and *RpiX178-2* for *Pythium* stalk rot (Duan et al. 2019), and *Rcg1* (Jung et al. 1994) for *Anthraco*se stalk rot. The causal genes of *qRfg1* and *qRfg2* have been cloned, and functional markers were developed for the implementation of marker-assisted selection for improving stalk rot resistance.

A key genomic region on chromosome 6 at 168 Mb conferring FSR resistance, with the PVE values ranging from 6.16 to 8.38%, was identified by a GWAS analysis recently, which was further validated by linkage mapping in two  $F_{2:3}$  populations (Rashid et al. 2022). The candidate gene conferring FSR resistance in this crucial region is annotated as a nucleic acid binding protein, playing an integral part in gene silencing pathways, and responding to diverse abiotic stress tolerances in maize (Zhai et al. 2019; Qian et al. 2011). However, no more GWAS research was reported to dissect the genetic architecture of maize FSR resistance; more GWAS researches are required for a comprehensive understanding of the genetic architecture of FSR resistance in different genetics and breeding populations.

Genomic prediction (GP), another novel genomic tool, provides opportunities for improving breeding efficiency and accelerating the development of maize varieties with genetic resistance to FSR. GP was also known as genomic selection (GS), which offers an attractive alternative to conventional breeding or marker-assisted selection (Meuwissen et al. 2001). In GS, the effects of all the molecular markers across the entire genome were estimated to predict the genomic estimated breeding value (GEBV) of candidates to be selected (Vivek et al. 2017; Jannink et al. 2010; Meuwissen et al. 2001). Prediction accuracy was always used to evaluate the effectiveness of GS for improving the target

trait; it was reported that key factors affecting prediction accuracy include trait heritability, prediction model, marker density, genotype-by-environment interaction ( $G \times E$ ), the relationship between the training and testing population, etc. (Edriss et al. 2017; Crossa et al. 2017; Guo et al. 2020; Mageto et al. 2020). However, the potential of exploiting GP for improving FSR resistance has not been reported; the potential of GS for improving FSR resistance needs to be further investigated by estimating the prediction accuracy of FSR resistance in different breeding scenarios and assessing the effects of different factors on the estimation of the prediction accuracy.

In this study, GWAS and GP analyses were performed on 562 tropical and subtropical maize inbred lines, where all of the maize inbred lines were screened in four environments under artificial inoculation to evaluate their response to FSR resistance and genotyped with genotyping-by-sequencing. The main objectives of the present study are to: (1) dissect the genetic architecture of FSR resistance, identify the significantly associated single-nucleotide polymorphisms (SNPs) and stable genomic regions conferring FSR resistance, and estimate the genetic effects of the favorable alleles and haplotypes in improving FSR resistance; (2) evaluate the prediction accuracy of FSR resistance in different breeding scenarios to explore the potential of GP for improving FSR resistance by performing the predictions with various cross-validation schemes and different prediction models, within and across populations; and (3) investigate the effects of key factors on estimation of prediction accuracy of FSR resistance, including to incorporate the  $G \times E$  into prediction, utilize the combined phenotypic datasets from the same year or same location for GP analysis, and select a subset of molecular markers by considering both the genome coverage and the threshold of the  $P$ -value of SNPs for prediction.

## Materials and methods

### Plant materials

In the present study, 562 tropical and subtropical maize inbred lines from two populations were used to conduct GWAS to dissect the genetic architecture of FSR resistance and estimate the prediction accuracy of FSR resistance under different scenarios. The first population, designated as the CIMMYT maize lines (CML) panel, consists of 280 tropical and subtropical maize inbred lines developed by the International Maize and Wheat Improvement Center (CIMMYT). From 1984 to 2023, CIMMYT developed and released a total of 647 CMLs, which represent a significant portion of the genetic diversity of CIMMYT maize germplasm (Wu et al. 2016). In the present study, the lowland

tropical, mid-altitude/subtropical lines between CML300 and CML603 were selected for screening their response to FSR resistance.

The second population, designated as the Drought Tolerant Maize for Africa (DTMA) panel, consisted of 282 tropical and subtropical inbred lines developed by CIMMYT. These lines originated from different breeding programs of CIMMYT, consisting of different kinds of lines with tolerance or resistance to an array of abiotic and biotic stresses (Yuan et al. 2019).

## Experimental design

The CML and DTMA populations were screened for FSR resistance at two CIMMYT experimental stations, Agua Fria (AF), in the state of Puebla in Mexico (97°38' W, 20°28'N, 110 m above sea level); and Tlaltizapan (TL), in the state of Morelos, Mexico (99°07'W, 18°4'N; 940 m above sea level). The CML population was planted in AF and TL in the summer season of 2018 and 2019. The DTMA population was planted in AF and TL in the summer season of 2014 and 2019. A randomized complete block design was used for all experiments with three replications per location and a single-row plot per replication. Each plot was 2.5 m long with 11 plants. The distance between rows was 0.80 m, and the distance between plants in a plot was 0.25 m.

The environment was defined as a combination of year and location. Therefore, each population would have phenotypic data from four environments and 12 data points. For example, the CML population was screened in four environments, designated as 2018AF, 2018TL, 2019AF, and 2019TL, respectively.

## Artificial inoculation and evaluation

All the 562 inbred lines were artificially inoculated with the pathogen of *Fusarium verticillioides* (*F.v*), which was the main pathogen causing stalk rot in maize in Mexico (Prasanna et al. 2021). It was cultured in fresh potato dextrose agar plates in which sterile toothpicks were inserted. The culture was incubated at 25 °C for 2 weeks, and the infected toothpicks were used for inoculation (Lal and Singh 1984). Fourteen days after flowering, all plants in each plot were inoculated by inserting infected toothpicks into a drilled hole on the first stem segment (approximately 0.1 m above the soil surface). A recent study reported that the toothpick inoculation method is effective with similar performance as other widely used inoculation methods, such as soil inoculation, drilling inoculation, and needle injection (Asiedu et al. 2024).

At the harvest, the plants were cutoff at the height of the cob approx. 0.50 to 1.00 m high above the ground, and the stalks were split longitudinally through the points of inoculation. Disease severity was estimated by the formula below:

$$\text{FSR severity (\%)} = \frac{\text{visible lesion area/}}{\text{whole longitudinal cut area}} \times 100\%$$

The FSR severity ranges from 0 to 100%. The FSR severity close to 0% (no visible disease symptoms or lesions identifiable on the stalk) means that the line has the highest level of resistance to FSR, i.e., the lowest FSR severity, whereas the FSR severity close to 100% means that the line has the lowest level or no resistance to FSR, i.e., the highest FSR severity.

## Phenotypic data analysis

For the CML and DTMA population, the best linear unbiased estimate (BLUE) values and broad sense heritability ( $H^2$ ) of FSR severity were analyzed within the single environment analysis and the combined analysis across environments (CombinedENV) by META-R software (<https://hdl.handle.net/10883/20997>) (Alvarado et al. 2020) using the mixed linear model. The mixed linear model applied in META-R was implemented with the 'lme4' (Bates et al. 2015) R-package using the function of 'lmer'. Meanwhile, the estimation of variance components in the mixed linear model was used the function of 'reml'. The formula was as follows:

$$\text{Individual environment : } y_{ijk} = \mu + g_i + e_j + r_k e_j + \varepsilon_{ijk}$$

$$\text{Combined environments : } y_{ijk} = \mu + g_i + e_j + g e_{ij} + r_k e_j + \varepsilon_{ijk}$$

where  $y_{ijk}$  is the FSR severity,  $\mu$  is the overall mean,  $g_i$ ,  $e_j$ , and  $g e_{ij}$  are the effects of the  $i$ th genotype,  $j$ th environment, and  $i$ th genotype by  $j$ th environment interaction, respectively.  $r_k e_j$  is the effect of the  $k$ th replication within the  $j$ th environment.  $\varepsilon_{ijk}$  is the residual effect of the  $i$ th genotype,  $j$ th environment, and  $k$ th replication. Genotype is treated as the fixed effect, whereas all other effects are declared as random effects. Moreover, there is no  $g e_{ij}$  (interaction between genotype and environment) in the single environment analysis.

The environment with an estimated heritability below 0.40 was excluded from the CominbedENV analysis. The  $H^2$  of FSR severity in individual environment analysis and CombinedENV analysis were calculated as:

$$\text{Individual environment analysis : } H^2 = \frac{V_g}{V_g + \frac{V_e}{k}}$$

$$\text{CombinedENV analysis : } H^2 = \frac{V_g}{V_g + \frac{V_{ge}}{j} + \frac{V_e}{jk}}$$

where  $V_g$  is genetic variance,  $V_{ge}$  is the variance of interaction between genotype and environment,  $V_e$  is error variance,

$j$  is the number of environments, and  $k$  is the number of replications within each environment.

In addition, description statistics of phenotypic data analysis was carried out in IBM SPSS Statistics, version 22.0 (IBM Corp. 2022). The distributions of FSR severity in the individual environment and CombinedENV analysis were plotted in R (R Core Team 2020) using the ‘*ggplot2*’ package (Wickham 2016). The Pearson correlations of FSR severity among each single and combined environment in the populations of CML and DTMA were calculated using the BLUE values and visualized in R using the package ‘*ggcorrplot*’ (Wickham et al. 2016). Moreover, the top 10 lines with the lowest FSR severity and the bottom 10 lines with the highest FSR severity were identified within each population.

### Genotyping, GBS, and SNP calling

Total genomic DNA was extracted from bulked young leaves for all lines using a CTAB procedure (Doyle and Doyle 1987). Genotyping was performed at Cornell University Biotechnology Resource Center (Ithaca, NY). Genomic DNA was digested with the restriction enzyme of *ApeKI*. Genotyping-by-sequencing (GBS) libraries were constructed in the 96-plex and sequenced on Illumina HiSeq2000 (Elshire et al. 2011). SNP calling was performed using the TASSEL GBS Pipeline, where the GBS Version 2.7 TOPM (tags on physical map) file downloaded from Panzea ([www.panzea.org](http://www.panzea.org)) was used to anchor reads to the maize reference genome of B73 RefGen\_v2 (Glaubitz et al. 2014; Wang et al. 2020). For each inbred line, 955,690 SNPs were called, 955,120 SNPs of them were evenly distributed on the ten maize chromosomes, while the other 570 SNPs were without position information.

### GWAS analysis

Before GWAS analysis, quality control of the genotypic data is an important step to ensure the accuracy of the later analysis. The combined population consisting of all the 562 inbred lines was abbreviated as CominbedPOP. The raw GBS datasets were filtered with a minor allele frequency (MAF) above 0.05, missing data rate below 30%, and heterozygosity rate below 5% in TASSEL V5.0 in the populations of CML, DTMA, and CominbedPOP, respectively. Then, the imputation was performed with the default parameters in TASSEL 5.0 (Bradbury et al. 2007) using the LD KNNi method (Money et al. 2015). The imputed GBS datasets and the BLUE values of FSR severity were used to conduct GWAS analyses in all three populations mentioned above.

Bayesian information and linkage disequilibrium iteratively nested keyway (BLINK) model (Huang et al 2019)

was chosen to detect the associations between the SNPs and FSR severity in the GWAS analysis, because this model effectively reduces the false positives. In addition to the capability to incorporate principal components (PCs) and kinship (K) as covariates to reduce false positives, BLINK iteratively incorporated associated markers as covariates to eliminate their unclear connection to the individuals. Moreover, the SNPs sampled in the BLINK model were selected according to linkage disequilibrium, optimized for Bayesian information content (BIC), and re-examined across multiple iterative to reduce false positives. The BLINK conducted two fixed effect models and one filtering process (Huang et al. 2019).

The GWAS analysis using the BLINK model was run by GAPIT version 3 (Wang et al. 2021) in R. Meanwhile, PC analysis and K matrix calculation were also conducted using default parameters. The default threshold of  $P$ -value in BLINK for selecting SNPs significantly associated with FSR resistance was defined as the Bonferroni multiple test threshold of  $0.05/n$ , and  $n$  is the number of SNPs. The results of PC analysis were visualized in R using the first two principal components by the package ‘*ggplot2*’ (Wickham et al. 2016). The Manhattan and quantile–quantile (QQ) plots of GWAS results were drawn by the ‘*complot*’ package (Van den Ende et al. 2019) in R.

### Genetic effect of the favorable allele

For each SNP significantly associated with FSR resistance detected by GWAS, the allele with a lower average FSR severity was assigned the favorable allele, whereas another allele with a higher average FSR severity was assigned the unfavorable allele. The formula for calculating the effect of each favorable allele was shown below:

Effect of each favorable allele = Average FSR severity of the lines carrying the favorable allele – Average FSR severity of the lines carrying the unfavorable allele.

### Candidate gene analysis

The average linkage disequilibrium (LD) decay for each chromosome was measured in TASSEL V5.0, using sliding window analysis with a window size of 50 SNPs. Squared Pearson correlation coefficient ( $r^2$ ) between vectors of SNPs was used to assess the level of LD decay on each chromosome, and the average LD decay distance across ten chromosomes at  $r^2 = 0.1$  was used to measure the LD decay distance in the populations of CML, DTMA, and CominbedPOP (Yan et al. 2009). The LD decay results were plotted against physical distance (kb) in R by the package ‘*ggplot2*’ (Wickham et al. 2016).

Considering LD decay distance, the interval of the physical position of SNP  $\pm$  LD decay distance was defined as a

genomic region. The overlapped or partially overlapped genomic regions were joined together as one region. Putative genes located in all the genomic regions were considered as candidate genes conferring FSR resistance. Annotation of candidate genes was performed on NCBI (<https://www.ncbi.nlm.nih.gov>) and MaizeGDB (<https://www.maizegdb.org>).

### Haplotype analysis within the genomic regions conferring FSR resistance

Haplotype blocks in genomic regions associated with FSR resistance were built using LDBlockShow software (Dong et al. 2021) based on standardized disequilibrium coefficients ( $D'$ ) (Flint-Garcia et al. 2003), and the significantly associated SNPs selected at the threshold of  $P$ -value at  $10^{-3}$  within each genomic region were used to build the haplotype blocks.

### Genomic prediction analysis

A five-fold cross-validation (CV) scheme with 20 replications was used to generate the training and validation sets randomly and assess the prediction accuracy. The average value of Pearson correlations between the true breeding values and the genomic estimated breeding values in the testing population was defined as the prediction accuracy (Liu et al. 2021). GP analysis was conducted using whole genome-wide SNPs and the BLUE values of FSR severity from single environment analysis in the populations of CML, DTMA, and CombinedPOP, respectively. The GP analysis was conducted using the BGLR library (Pérez et al. 2014) in the R program, where deviance information criterion (DIC) value was calculated for each model at the same time. The lower DIC value means that the model was more precise (Tomohiro 2011).

Two CV schemes were applied. The first scheme of CV1 was used to mimic one breeding scenario that predicts the newly developed lines, and it means that these lines have not been observed in any environment. The second scheme of CV2 was used to mimic sparse testing, in which some lines were observed in some environments but absent in others (Mageto et al. 2020).

To compare the prediction accuracy between phenotypic selection and GS and to assess the effects of incorporating genotype-by-environment interactions in improving prediction accuracy, three prediction models were applied. The first prediction model of M1 is a phenotypic prediction model, where the effects of the environment and lines were employed for prediction. The second prediction model of M2 is a general GP model where the effects of molecular markers were employed for prediction. The third prediction model of M3 is an extension of M2 incorporating  $G \times E$

into prediction. More details of these three models were described in Method S1.

To evaluate the effects of year and location on estimation of prediction accuracy, the phenotypic data of FSR severity were analyzed within the CML population and the DTMA population by combining the data from the same location (CombinedAF and CombineTL) or the same year (Combined2018 and Cominbed2019 in the CML population, Combined2014 and Cominbed2019 in the DTMA population). Within each population, the prediction accuracy was estimated using the BLUE values of FSR severity from the same location or the same year.

To investigate the GP accuracy estimated with the significantly associated SNPs conferring FSR resistance, different numbers of SNPs detected by GWAS at different thresholds of the  $P$ -value of  $10^{-3}$ ,  $10^{-4}$ , and  $10^{-5}$  were selected for conducting GP analyses with M3 in CV2, only the unique SNPs across all the GWAS analyses were selected for GP analyses.

GP accuracy was also estimated between the CML population and the DTMA population, by training one population to predict the other as a testing population, where both the genome-wide SNPs and the significant SNPs conferring FSR resistance detected by GWAS at a  $P$ -value threshold of  $10^{-3}$  were used for GP analyses with all three prediction models and CV2 scheme.

## Results

### Phenotypic variation of FSR severity and correlation analysis

The FSR severity had broader variations and higher average values in the CML population than those in the DTMA population across the individual environment and CombinedENV analyses, except for in 2019TL (Table 1, Fig. 1a and c). In the CombinedENV, the FSR severity in the CML population ranged from 29.17 to 92.50%, with an overall mean of 56.24%. The FSR severity in the DTMA population ranged from 17.41 to 79.86%, with an overall mean of 46.70%. The phenotypic differences between these two populations indicated their differences in genetic variations responding to FSR resistance, and in the disease, pressure occurred in different years.

The estimated heritabilities of FSR severity were medium to high in both populations, ranging from 0.67 to 0.85 in the CML population, and from 0.53 to 0.79 in the DTMA population by excluding the lowest heritability of 0.38 observed in the environment of 2014AF. In the CombinedENV analysis, the heritability of FSR severity in the populations of CML and DTMA was 0.77 and 0.55, respectively (Table 1).

The Pearson correlation coefficients of FSR severity among all the individual and combined environments were

**Table 1** Summary information of Fusarium stalk rot (FSR) severity in the populations of CIMMYT Maize Line (CML) and Drought Tolerant Maize for Africa (DTMA) in each individual environment and the combined environment analyses

Population	Environment <sup>a</sup>	Minimum (%)	Maximum (%)	Mean (%)	Standard error	Skewness	Kurtosis	V <sub>G</sub>	V <sub>G×E</sub>	V <sub>e</sub>	H <sup>2</sup>
CML	2018AF	36.67	100.00	66.24	0.013	-1.22	1.83	0.015***		0.013	0.77
	2018TL	40.00	93.33	63.31	0.012	-1.66	2.64	0.007***		0.011	0.67
	2019AF	16.67	96.67	50.34	0.008	0.29	-0.62	0.025***		0.013	0.85
	2019TL	10.00	86.67	45.11	0.007	0.20	-0.69	0.021***		0.019	0.76
DTMA	CombinedENV	29.17	92.50	56.24	0.006	0.43	0.12	0.010***	0.007***	0.014	0.77
	2014AF	18.67	88.00	48.49	0.003	0.29	-0.44	0.007***		0.034	0.38
	2014TL	11.11	91.11	43.78	0.004	0.60	0.23	0.011***		0.029	0.53
	2019AF	13.33	90.00	46.32	0.008	0.18	-0.37	0.019***		0.015	0.79
	2019TL	0.00	80.00	49.64	0.006	-0.10	-0.45	0.013***		0.017	0.70
	CombinedENV	17.41	79.86	46.70	0.003	0.17	0.09	0.006***	0.008***	0.020	0.55

<sup>a</sup>The individual environment was defined as a combination of year and location. The CIMMYT experimental stations of Agua Fria and Tlaltizapan were abbreviated as AF and TL, respectively. The combined analysis across all environments was abbreviated as CombinedENV

positive and moderate to high (Fig. 1b and d). The Pearson correlation coefficients between the CombinedENV and the individual environments were higher than those between the individual environments in both populations. In the CML population, the correlation coefficients between the individual environments ranged from 0.26 to 0.56, and the correlation coefficients between the CombinedENV and the individual environments ranged from 0.56 to 0.82. In the DTMA population, the correlation coefficients between the individual environments ranged from 0.15 to 0.44, and the correlation coefficients between the CombinedENV and the individual environments ranged from 0.64 to 0.77.

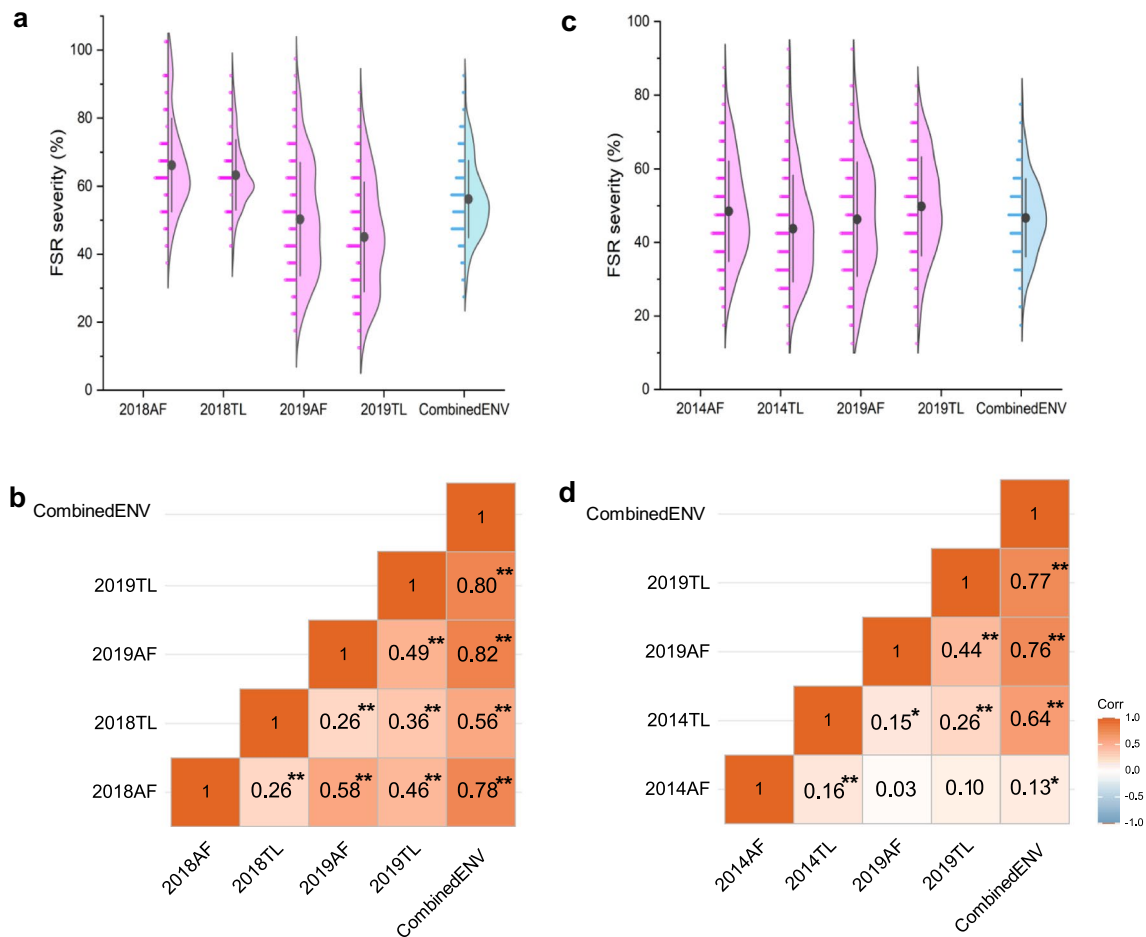
Within each population, the top ten resistant lines with the lowest FSR severity values and the bottom ten susceptible lines with the highest FSR severity values were identified based on the BLUE values in CombinedENV analysis (Supplementary Table 1). The FSR severity values of the top ten resistant lines ranged from 29.17% to 37.50% in the CML population (CML552, CML596, CML581, CML601, CML389, CML478, CML582, CML600, CML307, and CML401), and from 17.41% to 28.15% in DTMA population (DTMA175, DTMA143, DTMA180, DTMA155, DTMA187, DTMA192, DTMA146, DTMA261, DTMA60, and DTMA145). The bottom ten susceptible lines with the highest FSR severity values ranged from 78.33% to 92.50% in the CML population (CML511, CML360, CML590, CML334, CML467, CML584, CML585, CML591, CML329, and CML362), and from 67.78% to 79.86% in DTMA population (DTMA107, DTMA241, DTMA166, DTMA65, DTMA231, DTMA98, DTMA13, DTMA120, DTMA126, and DTMA29).

**Population structure analysis and LD decay distance**

After QC, 215,914, 209,111, and 221,190 SNPs were selected to perform further genetic analysis in the population of CML, DTMA, and CominbedPOP, respectively. The high-quality SNPs were distributed evenly on ten chromosomes in all three populations. The average MAF was 0.22, 0.19, and 0.20 in the population of CML, DTMA, and CominbedPOP, respectively. The average missing rate was 9%, 3%, and 6% in the population of CML, DTMA, and CominbedPOP, respectively (Fig. S1).

The result of population structure in all three populations was illustrated by the PCA plot, where the first two principal components of PC1 and PC2 together explained a total of 5.6%, 8.1%, and 4.7% of the phenotype variation in the populations of CML, DTMA, and CombinedPOP, respectively (Fig. 2a–c). All three populations have been divided into two clusters of tropical lines and subtropical lines based on their pedigree information.

The average LD decay distance at  $r^2 = 0.10$  across the ten chromosomes was 3.60 kb, 3.47 kb, and 2.83 kb in the



**Fig. 1** **a** The violin plot for the distribution of Fusarium stalk rot (FSR) severities in the CML population. **b** The phenotypic correlations of FSR severity among different environments in the CML population. **c** The violin plot for the distribution of FSR severities in the DTMA population. **d** The phenotypic correlations of FSR severity among different environments in the DTMA population. The individ-

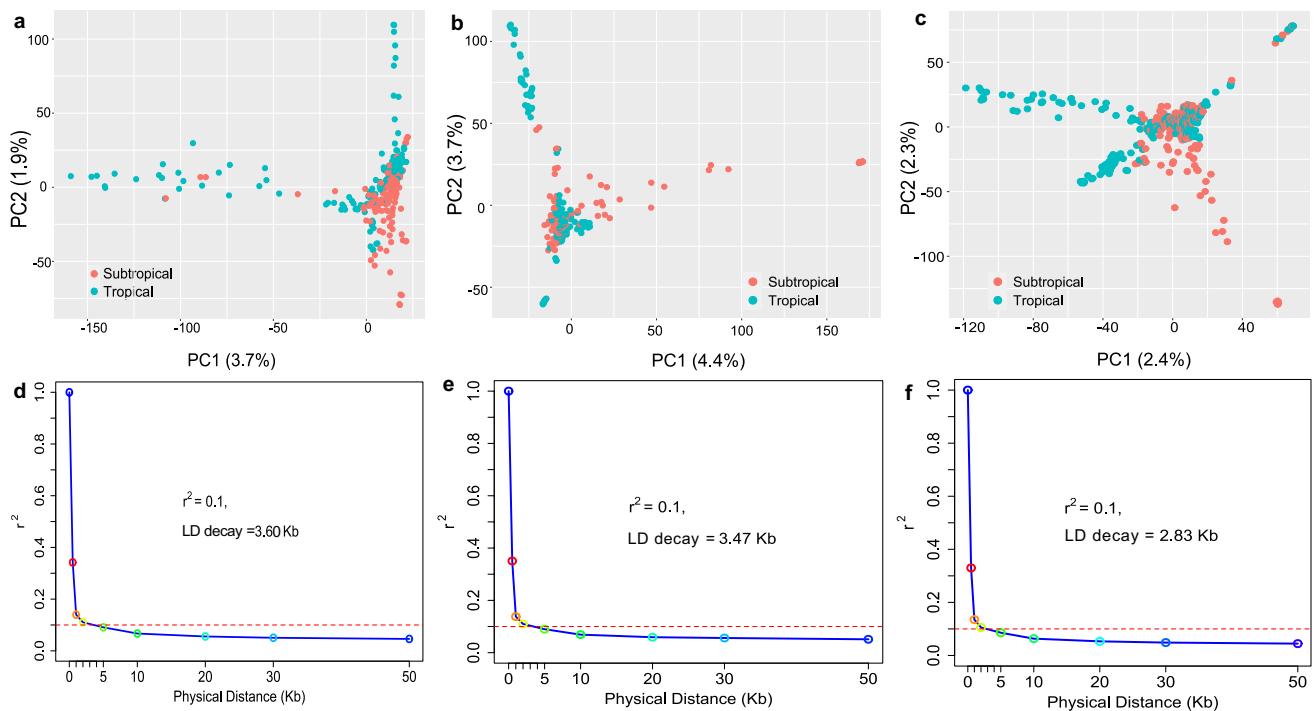
ual environment was defined as a combination of year and location. Agua Fria and Tlaltizapan were abbreviated as AF and TL, respectively. The combined analysis across all environments was abbreviated as CombinedENV. Correlation coefficients with \*\* represent extremely significant correlations, and correlation coefficients with \* represent significant correlations

populations of CML, DTMA, and CombinedPOP, respectively (Fig. 2d–f).

### Significantly associated SNPs and genomic regions conferring FSR resistance detected by GWAS and annotation of candidate genes

In total, 15 SNPs significantly associated with FSR resistance were detected in GWAS analyses across all three populations at the  $P$ -value threshold of  $0.05/n$  ( $n$  is the number of genome-wide SNPs used for GWAS), i.e.,  $2.3 \times 10^{-7}$ ,  $2.4 \times 10^{-7}$ , and  $2.3 \times 10^{-7}$  in the populations of CML, DTMA, and CombinedPOP, respectively (Table 2, Fig. 3a, c and e). The QQ plots from the three GWAS analyses indicated that the population structure was well controlled, and the BLINK model applied in the present study is powerful to identify reliable SNPs conferring FSR

resistance (Fig. 3b, d and f). These 15 SNPs significantly associated with FSR resistance were distributed on all ten chromosomes, only except for on chromosomes 9 and 10. Five of them were detected in the CML population, which were located on chromosomes 1, 2, 3, and 5. Seven of them were detected in the DTMA population, which were distributed on chromosomes 1, 3, 4, 6, 7, and 8. Three of them were detected in the CombinedPOP, which were concentrated on chromosomes 1 and 4. The  $P$ -values of the 15 significantly associated SNPs ranged from  $1.99 \times 10^{-7}$  to  $8.27 \times 10^{-13}$ , whose phenotypic variance explained (PVE) values ranged from 0.94 to 8.30%, with an average PVE value of 3.63% (Table 2). The significantly associated SNP of S6\_112215613 detected in the DTMA population had the lowest  $P$ -value of  $8.27 \times 10^{-13}$ , with a PVE value of 2.09% and a MAF of 0.32. Among these 15 SNPs, four of them had PVE values greater than 5%. The significantly



**Fig. 2** **a, b, c** Principal component (PC) analysis plots in the populations of CML, DTMA, and the CombinedPOP consisting of all 562 inbred lines. The tropical lines and subtropical lines in PCA plots

were colored in different colors. **d, e, f** The linkage disequilibrium decay plots in the populations of CML, DTMA, and the CombinedPOP

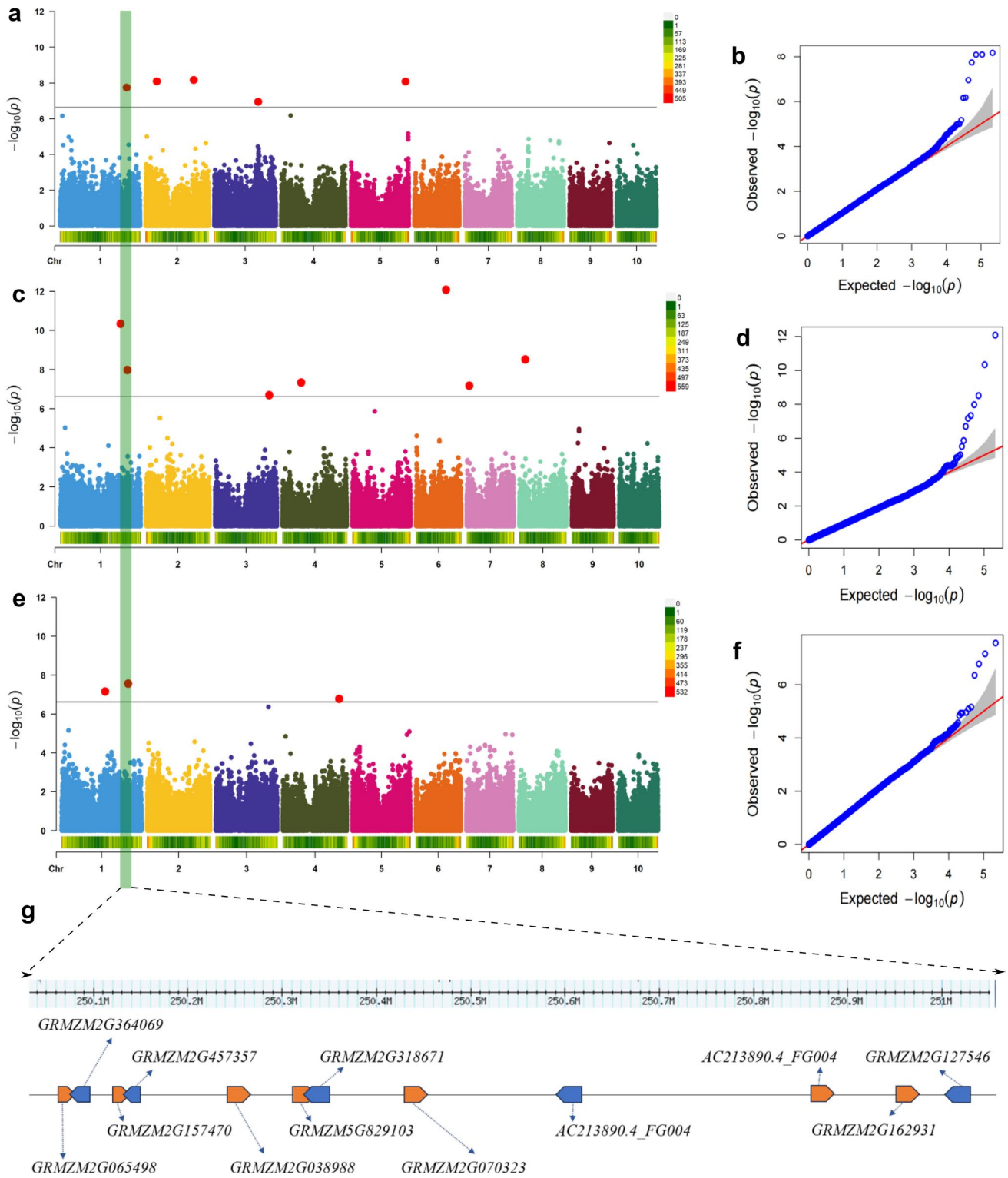
**Table 2** Information of significantly associated SNPs conferring FSR resistance detected by GWAS using BLINK model in populations of CML, DTMA, and CombinedPOP

SNP	Population	Bin	Allele	<i>P</i> -value	PVE (%)	MAF	No	Genomic region
<b>S1_250295975</b>	CML	1.09	T/C	1.81E-08	2.16	0.22	1	Chr1:250,089,724–251044933
<b>S1_251041463</b>	DTMA	1.09	T/G	1.04E-08	5.18	0.10		
<b>S1_250092554</b>	CombinedPOP	1.08	C/T	2.73E-08	3.14	0.24		
S2_41485521	CML	2.04	G/A	8.07E-09	8.30	0.11	2	Chr2:41,481,921–41,489,121
S2_180835270	CML	2.06	C/T	6.76E-09	2.88	0.39	3	Chr2:180,831,670–180838870
S3_165448326	CML	3.05	C/T	1.11E-07	4.27	0.29	4	Chr3:165,444,726–165,451,926
S5_204626484	CML	5.06	T/C	8.16E-09	5.56	0.09	5	Chr5:204,608,587–204630084
S1_225626297	DTMA	1.07	G/A	4.55E-11	3.17	0.30	6	Chr1:225,622,827–225,629,767
S3_202767355	DTMA	3.07	T/C	1.99E-07	4.07	0.06	7	Chr3:202,763,885–202770825
S4_69998604	DTMA	4.05	G/C	4.59E-08	0.94	0.47	8	Chr4:69,995,134–70,002,074
S6_112215613	DTMA	6.04	T/C	8.27E-13	2.09	0.32	9	Chr6:112,212,143–112,219,117
S7_9545142	DTMA	7.01	C/G	6.67E-08	1.76	0.36	10	Chr7:9,541,672–9,548,612
S8_21865355	DTMA	8.03	C/A	3.04E-09	1.91	0.27	11	Chr8:21,861,885–21,868,825
S1_163787906	CombinedPOP	1.05	A/G	6.95E-08	2.07	0.39	12	Chr1:163,785,076–163790736
<b>S4_211481644</b>	CombinedPOP	4.09	G/A	1.66E-07	6.98	0.05	13	Chr4:211,478,814–211,484,474

associated SNP of S2\_41485521 detected in the CML population had the largest PVE value of 8.30%, with a *P*-value of  $1.99 \times 10^{-7}$  and MAF of 0.11. The significantly associated SNP of S4\_211481644 detected in CombinedPOP had the second largest PVE value of 6.98%, with a *P*-value of

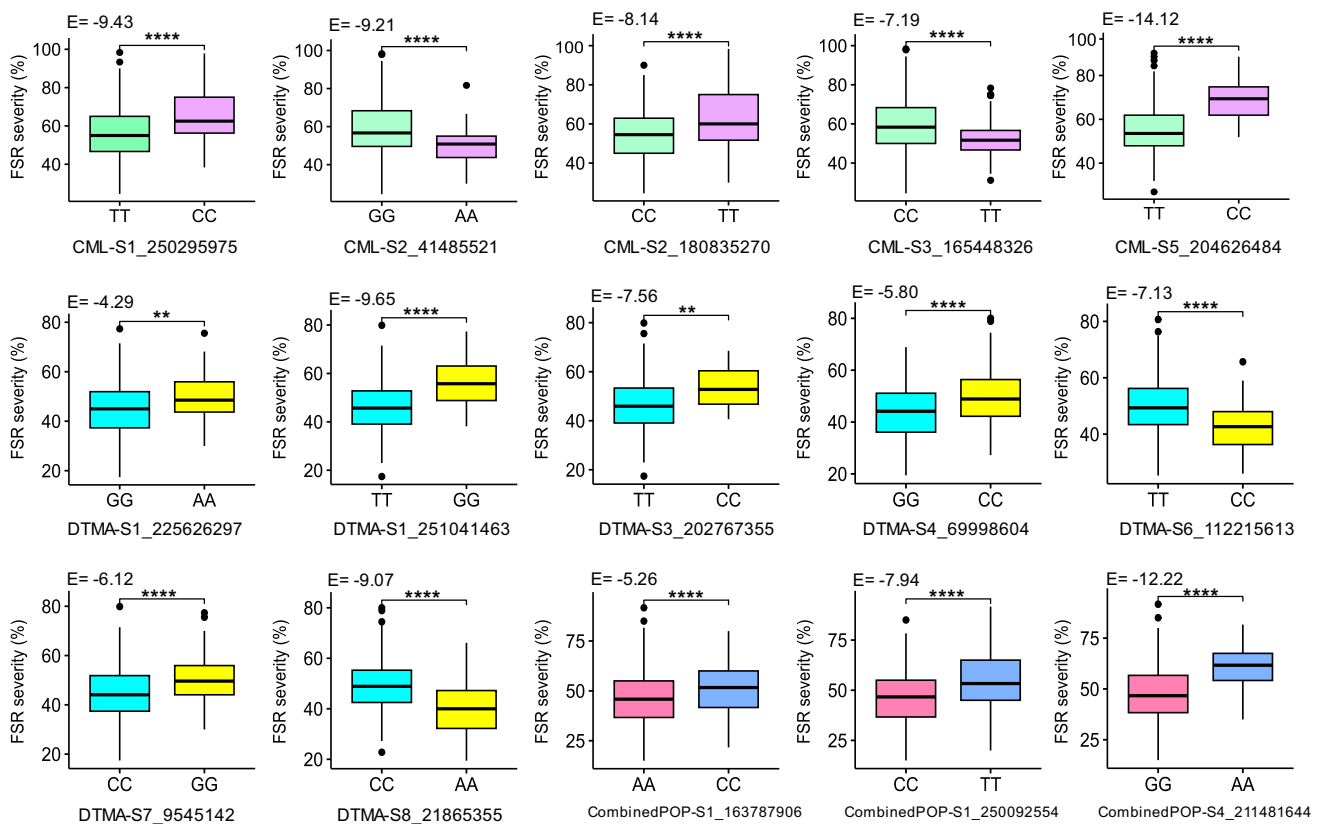
$1.66 \times 10^{-7}$  and a MAF of 0.05. These results showed that FSR resistance in tropical maize is controlled by multiple QTL (quantitative trait loci) with small to medium effects, and it is highly influenced by the genetic background of the populations studied.





**Fig. 3** **a-f** Manhattan and Quantile–Quantile (QQ) plots of the GWAS result using the BLINK model in populations of CML (**a, b**), DTMA (**c, d**), and CombinedPOP (**e, f**). The green bar showed the stable

genomic region conferring FSR resistance at 250Mb on chromosome 1 across all three populations. **g** The distribution of annotated candidate genes based on their physical positions in this region



**Fig. 4** The distribution, effect, and comparison analysis of FSR severity between the materials carrying favorable and unfavorable alleles for five significantly associated SNPs in the CML population (green

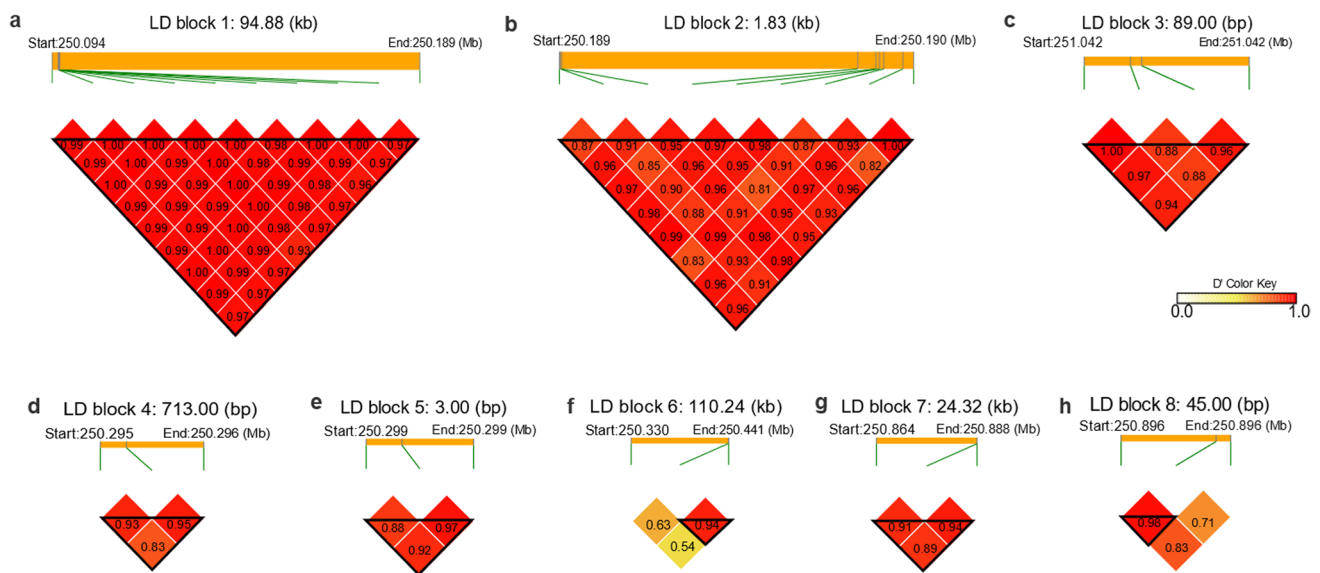
and pink), seven SNPs in the DTMA population (indigo and yellow), and three SNPs in the CominbedPOP (rose madder and blue)

The genetic effects of the 15 SNPs significantly associated with FSR resistance had significant or extremely significant differences between the favorable and unfavorable alleles in all the populations (Fig. 4). In the CML population, the genetic effect differences between the favorable and unfavorable alleles for each SNP ranged from  $-7.19$  to  $-14.12\%$ , with an average genetic effect of  $-9.62\%$ , while those values in the DTMA population ranged from  $-4.29$  to  $-9.65\%$ , with an average genetic effect of  $-7.09\%$ . The genetic effect differences between the favorable and unfavorable alleles for each SNP in the CombinedPOP ranged from  $-5.26$  to  $-12.22\%$ , with an average genetic effect of  $-8.47\%$ . For the 15 SNPs significantly associated with FSR resistance, the frequencies of the favorable alleles of 11 SNPs were greater than 0.50, and the frequencies of the favorable alleles of the rest of 4 SNPs, i.e., S2\_41485521 and S3\_165448326 in CML population, and S6\_112215613 and S8\_21865355 in the DTMA population, were smaller than 0.50.

Based on the information of LD decay distance within each population and the physical position of the significantly associated SNPs detected, 13 key genomic regions conferring FSR resistance were identified (Table 2), including five genomic regions in CML located in bins 1.09, 2.04, 2.06,

3.05, and 5.06, seven genomic regions in DTMA located in bins 1.07, 1.09, 3.07, 4.05, 6.04, 7.01, and 8.03, and three genomic regions in CombinedPOP located in bins 1.04, 1.08, and 4.09. In total, nine haplotype blocks were built for all 13 key genomic regions, only except for the key genomic region at  $\sim 250$  Mb on chromosome 1 detected across all three populations (Fig. S2).

In these genomic regions, 26 putative candidate genes conferring FSR resistance were identified (Supplementary Table 2). *GRMZM2G414537*, *GRMZM2G059106*, and *GRMZM2G042027* were associated with transmembrane transport, which may be involved in the invasion and anti-invasion process between pathogen and plant (Sailer et al 2018). Furthermore, *AC213890.4\_FG004*, *GRMZM2G070323*, and *GRMZM2G002555* have the function of protein phosphorylation, known as a switch or coordinator, which can help crops to make responses to some specific stresses by coordinating the expression of some functional genes (Yao and Xu 2017).



**Fig. 5** Eight haplotype blocks carrying favorable genetic effects conferring FSR resistance located in the stable genomic region at ~250 Mb on chromosome 1 detected in all three populations

### A stable genomic region associated with FSR resistance at ~250 Mb on chromosome 1

Across all three populations, a stable genomic region conferring FSR resistance was consistently detected at ~250 Mb on chromosome 1, in an interval of 0.95 Mb from 250,089,724 bp to 251,044,933 bp (Table 2) based on the reference genome of B73 RefGen\_v2. In addition, a significantly associated SNP of S1\_253271793 close to this genomic region was also identified in the GWAS analysis in the DTMA population using the phenotypic data from the environment 2019AF (data not given), confirming the importance of this genomic region. The  $P$ -values of the three significantly associated SNPs detected in this genomic region ranged from  $1.04 \times 10^{-8}$  to  $2.73 \times 10^{-8}$ , with PVE values ranging from 2.16 to 5.18%, and the differences of the genetic effects between the favorable and the unfavorable alleles ranging from  $-7.94$  to  $-9.65\%$  (Table 2 and Fig. 4).

In total, 21, 16, and 20 haplotype blocks were identified in the stable genomic region conferring FSR resistance at ~250 Mb on chromosome 1 in the populations of CML, DTMA, and CombinedPOP, respectively. Among them, 28 unique haplotype blocks were repeatedly detected in at least two populations, including nine unique haplotype blocks repeatedly detected across all the three populations. Moreover, eight of these nine unique haplotypes carried favorable genetic effects conferring FSR resistance. The information of physical position and the  $D'$  of each pair of SNPs within each of eight haplotype blocks is shown in Fig. 5. The eight haplotype blocks were distributed evenly in this stable genomic region, with the length ranging from 3 bp

to 110.24 kb, and the number of SNPs ranging from 2 to 10. The average  $D'$  value of each pair of SNPs across each haplotype block was higher than 0.90, indicating that all the SNPs within each haplotype block were all strongly linked (Fig. 5).

In total, 12 putative candidate genes in this genomic region were identified, the distribution of these candidate genes based on their physical positions in this genomic region is shown in Fig. 3g, and their annotated functions associated with the responses to biotic or abiotic stresses in crops are listed in Supplementary Table 2. Among them, *GRMZM2G457357*, *GRMZM2G364069*, and *GRMZM5G829103* were associated with zinc ion binding and zinc finger proteins. Zinc finger proteins maintain a finger-like spatial configuration by binding to zinc ions through amino acids in the peptide chain, which are widely distributed on the plasma membrane under abiotic stress, and may have functions as sensors or abscisic acid (ABA) receptors in abiotic stress signaling (Han et al 2020). *GRMZM2G457357* and *GRMZM2G027991* belonged to the ubiquitin–proteasome system (UPS), which is a rapid regulatory mechanism for selective protein degradation and plays crucial roles in an integral part of plant adaptation to stresses, such as drought, salinity, cold, nutrient deprivation, and pathogens (Xu et al. 2019). *GRMZM2G070323* and *AC213890.4\_FG004* have phosphoprotein phosphatase activity as mentioned above. Additionally, the rest of the putative candidate genes were directly associated with the intracellular signal transduction or response to abiotic stresses. Results of the annotated functions of these candidate genes revealed the

importance of this genomic region at ~250 Mb on chromosome 1 involving in positive responses to biotic and abiotic stresses, in particular in improving FSR resistance.

### Prediction accuracies of FSR severity estimated from different breeding scenarios

Prediction accuracies of FSR severity estimated from different breeding scenarios, including two CV schemes and three prediction models, are shown in Fig. 6a–c. Results showed that prediction accuracies of FSR severity estimated from the CV2 were higher than those values estimated from the CV1 across all the populations and prediction models, indicating that the predictions could benefit from previous records of lines whose FSR severity values have already been observed in other environments. Across all the CV schemes and populations, GP model of M2 obtained higher accuracies of FSR severity than those values obtained from the phenotypic prediction model of M1, and the GP model of M3 incorporating  $G \times E$  obtained the highest accuracies of FSR severity.

In the CV1 of mimicking the breeding scenario that predicts the newly developed lines never been tested in any environment, the accuracies of FSR severity estimated with the phenotypic prediction model of M1 were close to zero across all the three populations. The accuracies of FSR severity estimated with the GP model of M2 were 0.36, 0.29, and 0.34 in the populations of CML, DTMA, and CombinedPOP, respectively. The accuracies of FSR severity estimated with the GP model of M3 incorporating  $G \times E$  were 0.40, 0.36, and 0.36 in the populations of CML, DTMA, and CombinedPOP, respectively.

In the CV2 to mimic breeding scenario of sparse testing, in which some newly developed lines were observed in some environments but absent in others, the accuracies of FSR severity estimated with the phenotypic prediction model of M1 were 0.48, 0.29, and 0.48 in the populations of CML, DTMA, and CombinedPOP, respectively. The accuracies of FSR severity estimated with the GP model of M2 were 0.51, 0.34, and 0.51 in the populations of CML, DTMA, and CombinedPOP, respectively. Among all the three prediction models, the highest accuracies of FSR severity were observed in the GP model of M3 incorporating  $G \times E$ , which were 0.55, 0.42, and 0.53 in the populations of CML, DTMA, and CombinedPOP, respectively.

### Prediction accuracies of FSR severity estimated with the phenotypic data from four individual environments and the combined phenotypic data from the same location or same year

Prediction accuracies of FSR severity estimated with the BLUE values from the combined analysis across the same

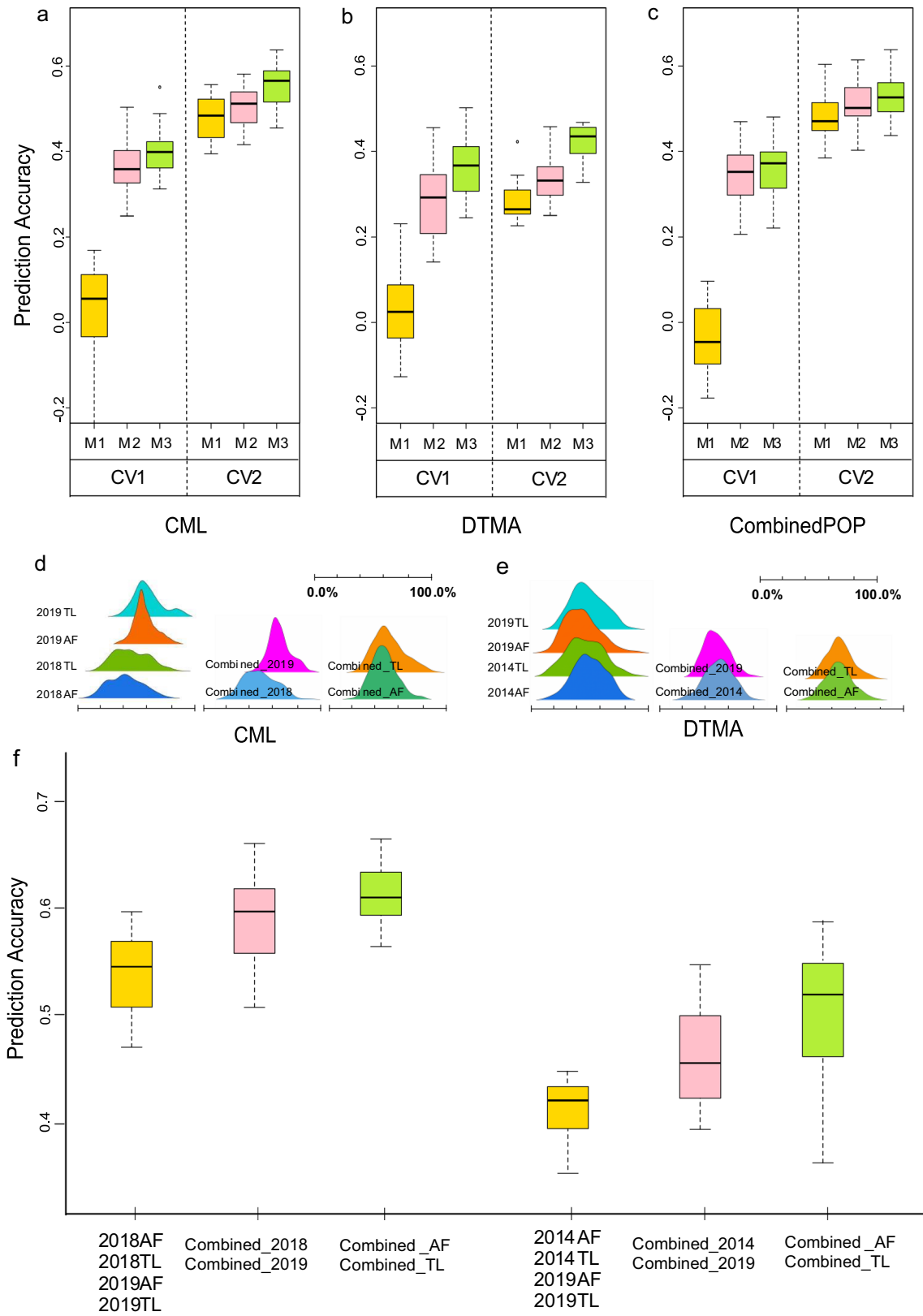
location or same year were higher than those values estimated with the BLUE values from the combined analysis across four individual environments. In the CML population, the prediction accuracy of FSR severity in M3 and CV2 estimated with the BLUE values from the combined analysis across four individual environments was 0.55, which was improved to 0.60 using the BLUE values from the combined analysis across years, i.e., Cominbed\_2018 and Cominbed\_2019, and to 0.62 using the BLUE values from the combined analysis across locations, i.e., Cominbed\_AF and Cominbed\_TL, (Fig. 6d and f). Similar trend was also observed in the DTMA population (Fig. 6e, f).

### Prediction accuracies of FSR severity estimated with different numbers of significant SNPs detected by GWAS

Prediction accuracies of FSR severity estimated with the whole genome-wide SNPs, 2105, 197, and 26 significantly associated SNPs selected at the thresholds of  $P$ -values at  $10^{-3}$ ,  $10^{-4}$ , and  $10^{-5}$  are shown in Fig. 7a–c. Prediction accuracies of FSR severity estimated with 2105 significantly associated SNPs selected at the thresholds of  $P$ -values at  $10^{-3}$  were the highest across all the three populations, and the prediction accuracies estimated with the whole genome-wide SNPs were higher than those values estimated with 197 and 26 significantly associated SNPs selected at the thresholds of  $P$ -values at  $10^{-4}$  and  $10^{-5}$  across all the three populations. Prediction accuracies of FSR severity in CV2 and M3 estimated with 2105 significantly associated SNPs selected at the threshold of  $P$ -value at  $10^{-3}$  were 0.69, 0.60, and 0.69 in populations of CML, DTMA, and CombinedPOP, respectively.

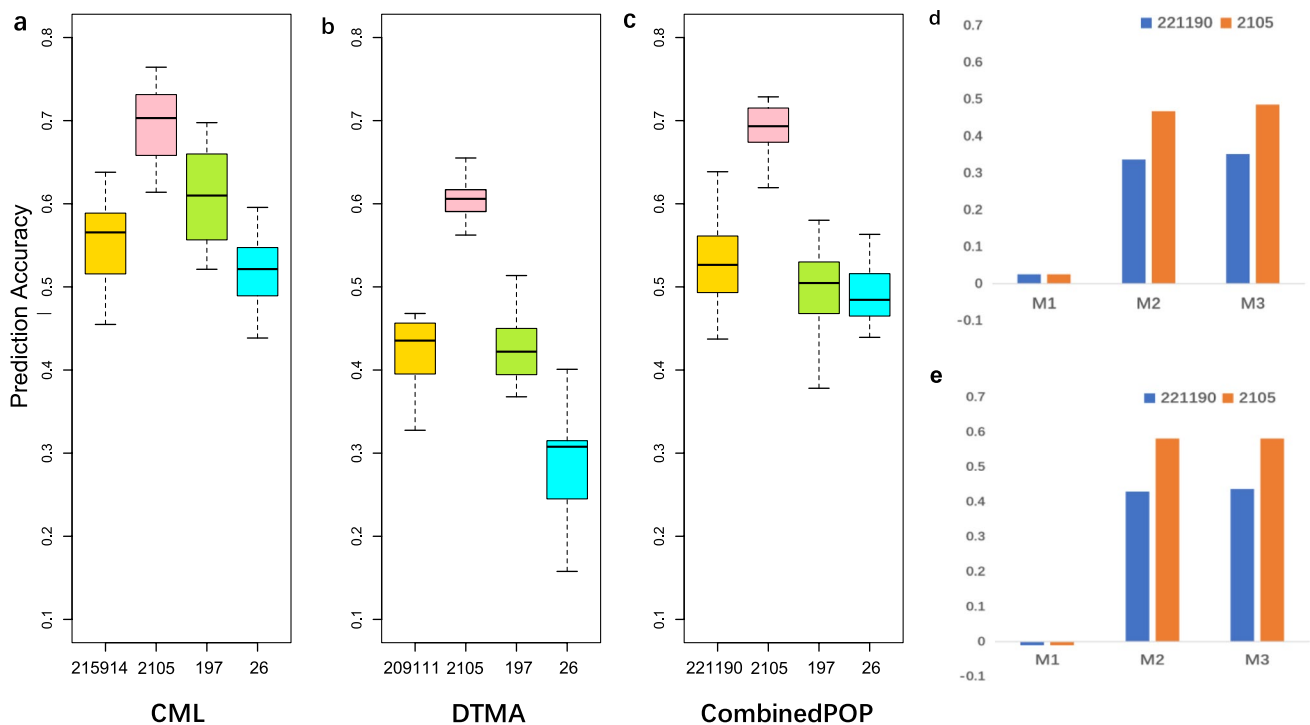
### Prediction accuracies of FSR severity estimated across different populations

Prediction accuracies of FSR severity estimated across different populations, i.e., training the CML population to predict the DTMA population as the testing population, or vice versa, are shown in Fig. 7d–e. Moderate prediction accuracies of FSR severity were observed across population in the GP model of M2 and the GP model of M3 incorporating  $G \times E$ , when the whole genome-wide SNPs or 2105 significantly associated SNPs selected at the threshold of  $P$ -value at  $10^{-3}$  were used for prediction. By training the CML population to predict the DTMA population as testing population, the prediction accuracies of FSR severity across populations in CV2 and M3 estimated with the whole genome-wide SNPs and the 2105 significantly associated SNPs selected at the threshold of  $P$ -value at  $10^{-3}$  were 0.35 and 0.49, respectively. By training the DTMA population to predict the CML population as testing population, the



**Fig. 6 a-c** Prediction accuracies of FSR severity estimated using three prediction models of M1, M2, and M3, in two cross-validation (CV) schemes in the populations of CML, DTMA, and CombinedPOP. **d-e** The distribution of FSR severity in the environments used in the CML

and DTMA populations when conducted GP. **f** The prediction accuracy of FSR severity estimated in CML (left side) and DTMA (right side) populations using CV2 and M3



**Fig. 7** a-c Prediction accuracies of FSR severity estimated with the different number of significantly associated SNPs selected from GWAS results at  $P$ -value thresholds of  $10^{-3}$  (2105 SNPs),  $10^{-4}$  (197 SNPs), and  $10^{-5}$  (26 SNPs) in populations of CML, DTMA, and CombinedPOP using CV2 and M3. d-e Prediction accuracies of FSR

severity estimated with the whole genome-wide 221,190 SNPs, and 2105 SNPs detected by GWAS at the  $P$ -value threshold of  $1 \times 10^{-3}$  in M1, M2, and M3 by training the CML population to predict the DTMA population or by training the DTMA population to predict the CML population

prediction accuracies of FSR severity across populations in CV2 and M3 estimated with the whole genome-wide SNPs and the 2105 significantly associated SNPs selected at the threshold of  $P$ -value at  $10^{-3}$  were 0.44 and 0.59, respectively.

## Discussion

Similar to the observations were reported in other maize stalk rots caused by different pathogens (Pè et al. 1993; Duan et al. 2019; Mu et al. 2019; Jung et al. 1994), this study also showed that the FSR resistance in maize is a complex quantitative trait with medium to high heritabilities and significantly affected by  $G \times E$ , it is controlled by multiple loci with minor effects. In total, 15 SNPs significantly associated with FSR resistance were identified across all three populations, with the  $P$ -values ranging from  $1.99 \times 10^{-7}$  to  $8.27 \times 10^{-13}$ , and PVE values ranging from 0.94 to 8.30%. These findings extend the knowledge of understanding the genetic architecture of FSR resistance in tropical maize.

In a previous study of Rashid et al. (2022), 342 tropical/sub-tropical maize inbred lines in the CAAM panel were also used to conduct GWAS analysis for genetic mapping the FSR resistance. Although this previous and the

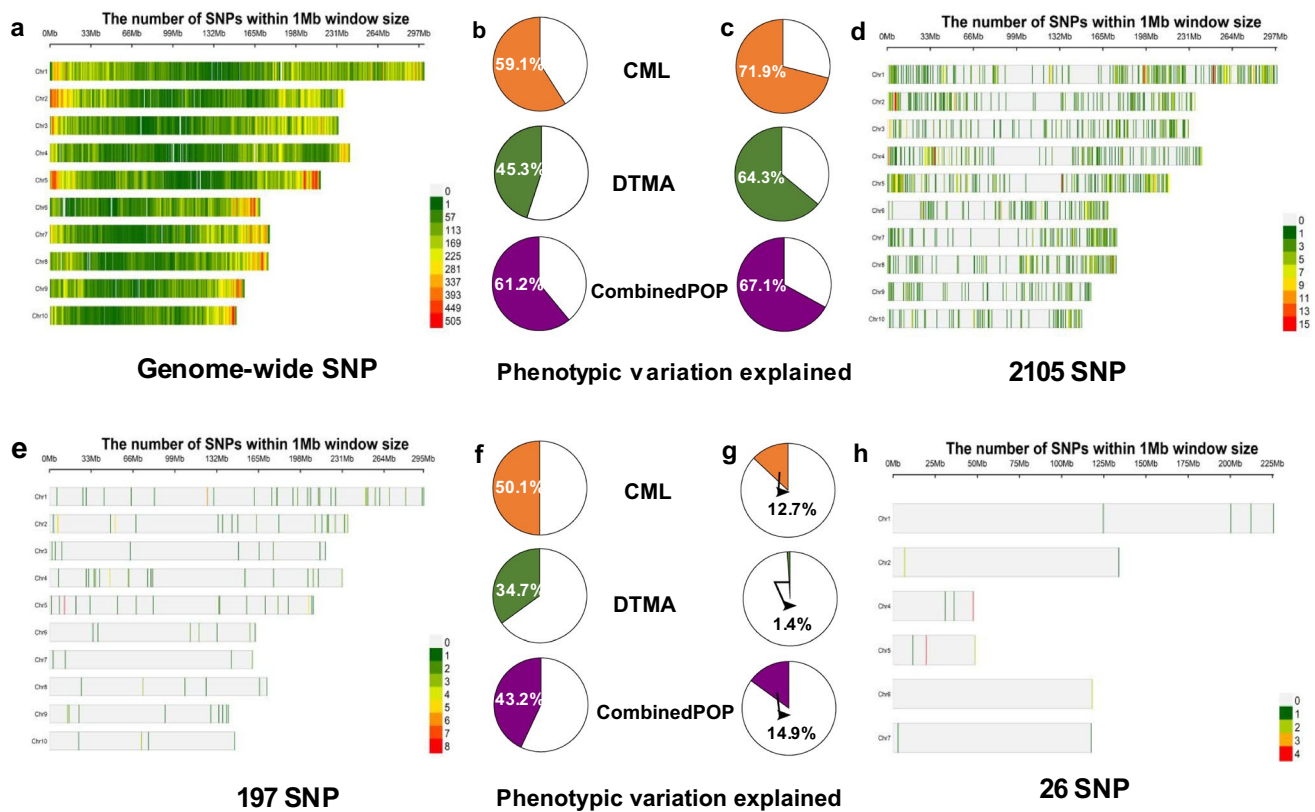
present studies both focused on the genetic dissection of the resistance to FSR in tropical maize, the genetic mapping populations, FSR screening environments, and validation approaches were different between these two studies, resulting in the detection of different loci significantly associated with the FSR resistance. In the previous study of Rashid et al. (2022), the maize inbred lines in the CAAM panel for GWAS analysis are adapted to Asian tropical ecologies with predominantly yellow kernel color. In the present study, the CML and DTMA populations represented a broader genetic diversity of tropical maize germplasm developed by CIMMYT with mostly white kernel color. The CAAM panel was screened for FSR resistance in two locations in India in the previous study, whereas the CML and DTMA populations used in the present study were screened for FSR resistance in two locations in Mexico. Moreover, the validation approaches were different in these two studies, the peak at 168 Mb on chromosome 6 detected by GWAS was validated in further QTL mapping analyses in the previous study of Rashid et al. (2022), and the peak at  $\sim 250$  Mb on chromosome 1 detected by the present study was validated across two independent GWAS populations. Although there are several major differences between these two studies, the importance of the crucial genomic region conferring FSR

resistance at 168 Mb on chromosome 6 identified by Rashid et al. (2022) was also partially validated in the present study (Supplementary Table 3), six SNPs highly linked with FSR resistance in the genomic region between 162 and 168 Mb on chromosome 6 were detected by the mixed linear model in GWAS analysis using the populations from the present study (Supplementary Table 3). The  $P$ -value of these six SNPs ranged from  $2.88 \times 10^{-5}$  to  $7.07 \times 10^{-4}$ , and the PVE values ranged from 2.60 to 6.13%, indicating the consistency of the results across different studies.

In addition, a stable genomic region conferring FSR resistance at ~250 Mb on chromosome 1 was identified by the present study, providing additional valuable information on crucial genomic region conferring FSR resistance to further investigate the possibility of developing trait markers for deployment in breeding programs. Several QTL associated with stalk rot resistance caused by different pathogens have been reported on chromosome 1 as well. In our previous study, a genomic region in bin 1.06 on chromosome 1 conferring FSR resistance was identified in different GWAS panels developed by CIMMYT (Song et al. 2024). In addition, QTL in bin 1.03 and 1.09 associated with *Pythium* stalk rot resistance were reported previously by Song et al. in 2015 and Duan et al. in 2019, respectively. In this crucial genomic region conferring stalk resistance, several putative candidate genes highly associated with responses to abiotic and biotic stresses were also reported.

One of the novel genomic tools of GP, also known as GS, had been reported as an effective approach for crop improvement, despite the prediction accuracies were highly affected by the trait heritability, prediction model, marker density, genotype-by-environment interaction, the relationship between the training and testing population, etc. (Sitonik et al. 2019; Nyaga et al. 2019; Zhang et al. 2017). Extensive research has been conducted on evaluating the potential of the utilization of GP for improving breeding efficiency for various traits, including the major maize diseases (Yu et al. 2022; Guo et al. 2020; Beyene et al. 2015; Oakey et al. 2016). In the previous studies, prediction accuracy for resistance to northern corn leaf blight reached to 0.70, for resistance to maize lethal necrosis ranged from 0.46 to 0.86, and for resistance to *Fusarium* ear rot ranged from 0.46 to 0.67 (Technow et al. 2013; Sitonik et al. 2019; Kuki et al. 2020; Liu et al. 2021; Holland et al. 2020). In the present study, GP accuracies of FSR severity estimated with the whole genome-wide SNPs were moderate and ranged from 0.29 to 0.51 in different breeding scenarios across two CV schemes and three populations. Moreover, moderate prediction accuracies of FSR severity estimated with the whole genome-wide SNPs were also observed across populations. These results show the high potential of GP for improving the FSR resistance.

For further improving the prediction accuracy of FSR resistance, the effects of the key factors on estimations of prediction accuracy were investigated in the present study, focusing on how to sample the most informative molecular markers, and incorporate the effects of environments and  $G \times E$  into prediction models. Generally, GP using the whole genome-wide SNPs was expected to achieve the highest prediction accuracy (Massman 2013; Lorenz et al. 2011; Budhlakoti et al. 2020). However, the low-density markers were also cost-effective for the implementation of GS (Zhang et al. 2015; Abed et al. 2018; Werner et al. 2018). In the present study, the prediction accuracies of FSR severity estimated with the 2105 significantly associated SNPs selected at the threshold of  $P$ -value at  $10^{-3}$  were higher than those prediction accuracies estimated with the whole genome-wide SNPs across all the three populations, whereas the prediction accuracies of FSR severity estimated with the 197 and 26 significantly associated SNPs selected at the thresholds of  $P$ -values at  $10^{-4}$  and  $10^{-5}$  were lower than those prediction accuracies estimated with the whole genome-wide SNPs across all populations. These results indicated that considering both the genome coverage and the threshold of the  $P$ -value of SNPs to select a subset of molecular markers could improve the GP accuracy. Similar trends were also observed in the total PVE values estimated with the different SNP datasets (Fig. 8). Across all populations, the total PVE values estimated the 2105 significantly associated SNPs selected at the threshold of  $P$ -value at  $10^{-3}$  were higher than those estimated with the whole genome-wide SNPs, whereas the total PVE values estimated with the 197 and 26 significantly associated SNPs selected at the thresholds of  $P$ -values at  $10^{-4}$  and  $10^{-5}$  were lower than those total PVE values estimated with the whole genome-wide SNPs. These results indicated that the subset of SNPs representing more genotypic information conferring the FSR resistance could achieve higher prediction accuracy. Furthermore, analysis of the variance components estimated with different SNP datasets also validated that the prediction accuracies of FSR severity estimated with the 2105 significantly associated SNPs selected at the threshold of  $P$ -value at  $10^{-3}$  were highest across all populations, due to the minimum value of residual variance and the lowest deviance information criterion (DIC) value representing more effective and precise prediction (Supplementary Table 4). The previous studies showed that GP model incorporating  $G \times E$  achieved higher prediction accuracy (Burgueño et al. 2012; Guo et al. 2013; Jarquín et al. 2017; Zhang et al. 2015; Monteverde et al. 2018). In the present study, the analysis of variance components of the random effects in each prediction model and the percentage of the total variance explained by each random effect were performed (Supplementary Tables 5 and 6). In Supplementary Table 5, GP model of M3 incorporating  $G \times E$  had the lowest percentage of the total variance



**Fig. 8** a, d, e, h The distribution on each chromosome of the whole genome-wide SNPs (a), 2105 SNPs (d), 197 SNPs (e), 26 SNPs (h) selected from GWAS results at  $P$ -value thresholds of  $10^{-3}$ ,  $10^{-4}$ , and  $10^{-5}$ . b, c, f, g The total phenotypic variation explained (PVE) was

estimated in the populations of CML, DTMA, and CombinedPOP using the whole genome-wide SNPs (b), 2105 SNPs (c), 197 SNPs (f), and 26 SNPs (g), respectively

explained by the residual effect, the highest percentage of the total variance explained together by the environment and the genotyped-by-environment interaction, and the lowest DIC values among all the three models in both populations. These results support the conclusion that GP model of M3 incorporating  $G \times E$  could improve the prediction accuracy. Similar results are also observed in Supplementary Table 6, the GP model employing the phenotypic data combined the same year or same location had the lower percentage of the total variance explained by the residual effect, the higher percentage of the total variance explained together by the environment and the  $G \times E$ , compared with the GP model employing the phenotypic data from individual environment. These results enhance the understanding of exploiting GP for improving FSR resistance.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00122-024-04597-x>.

**Acknowledgements** The authors gratefully acknowledge the financial support received from the Accelerating Genetic Gains for Maize and Wheat Improvement (AGG) project [INV-003439] funded by Bill & Melinda Gates Foundation (B&MGF), U.S. Agency for International

Development (USAID) and Foundation for Agricultural Research (FFAR), the AGG supplemental project (Amend. No. 9 MTO 069033) funded by USAID, One CGIAR Initiative on Accelerated Breeding, and the CGIAR Research Program on Maize (MAIZE). The authors also gratefully acknowledge the financial support received from Shennong Laboratory (Award Number SN01-2022-01). Junqiao Song was financially supported by the China Scholarship Council.

**Author contribution statement** XZ and CW initiated and designed the overall study. FSV, TD, CMZ, and XZ performed and coordinated the field experiments and phenotypic data collection. XZ, HZ, WS, and BMP contributed to genotypic data generation. JS, YL, AP, RG, HW, SC, and GH conducted the data analysis. JS, YL, CW, and XZ interpreted the results and wrote the manuscript. All authors contributed to manuscript editing.

**Funding** The research was funded by the Accelerating Genetic Gains for Maize and Wheat Improvement (AGG) project [INV-003439], AGG supplemental project (Amend. No. 9 MTO 069033), One CGIAR Initiative on Accelerated Breeding, Project of Shennong Laboratory (SN01-2022-01) and CGIAR Research Program (CRP) on Maize.

**Data availability** The datasets generated and analyzed during the current study are available online at CIMMYT Research Data & Software Repository at: <https://hdl.handle.net/11529/10549066>.



## Declarations

**Conflict of interest** The authors declare no competing interests.

## References

- Abed A, Pérez-Rodríguez P, Crossa J, Belzile F (2018) When less can be better: how can we make genomic selection more cost-effective and accurate in barley? *Theor Appl Genet* 131:1873–1890. <https://doi.org/10.1007/S00122-018-3120-8>
- Alvarado GR, Francisco P, Angela B, Juan C, Jose V, Mateo PR, Paulino L (2020) META-R: A software to analyze data from multi-environment plant breeding trials. *CROP J*. <https://doi.org/10.1016/j.cj.2020.03.010>
- Asiedu DD, Akohoue F, Frank S, Koch S, Lieberherr B, Oyiga B, Kessel B, Presterl T, Miedaner T (2024) Comparison of four inoculation methods and three *Fusarium* species for phenotyping stalk rot resistance among 22 maize hybrids (*Zea mays*). *Plant Pathol*. <https://doi.org/10.1111/ppa.13874>
- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *J Statist Softw* 67(1):1–48
- Beyene Y, Semagn K, Mugo S, Tarekegne A, Babu R, Meisel B, Sehabiague P, Makumbi D, Magorokosho C, Oikeh S, Gakunga J, Vargas M, Olsen M, Prasanna BM, Banziger M, Crossa J (2015) Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. *Crop Sci* 55(1):154–163. <https://doi.org/10.2135/cropsci2014.07.0460>
- Bradbury P, Zhang Z, Kroon D, Casstevens T, Ramdoss Y, Buckler E (2007) TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>
- Budhlakoti N, Rai A, Mishra DC, Jaggi S, Kumar M, Rao AR (2020) Comparative study of different non-parametric genomic selection methods under diverse genetic architecture. *Int J Pharma Bio Sci* 80:395–401
- Burgueño J, de los Campos G, Weigel K, Crossa J, (2012) Genomic prediction of breeding values when modeling genotype × Environment interaction using pedigree and dense molecular markers. *Crop Sci* 52:707–719. <https://doi.org/10.2135/CROPSCI2011.06.0299>
- Chambers KR (1987) Stalk rot of maize: host-pathogen interaction. *J Phytopathol* 118:103–108
- Chen Q, Song J, Du W, Xu L, Jiang Y, Zhang J, Xiang X, Yu G (2017) Identification, Mapping, and Molecular Marker Development for *Rgsr8.1*: A New Quantitative Trait Locus Conferring Resistance to *Gibberella* Stalk Rot in Maize (*Zea mays* L.). *Front Plant Sci* 8: 1355. <https://doi.org/10.3389/fpls.2017.01355>
- Chivasa W, Mutanga O, Burgueño J (2021) UAV-based high-throughput phenotyping to increase prediction and selection accuracy in maize varieties under artificial MSV inoculation. *Comput Electron Agric* 184:106128. <https://doi.org/10.1016/j.compag.2021.106128>
- Christensen S, Nemchenko A, Park Y, Borrego E, Huang P, Schmelz E, Kunze S, Feussner I, Yalpani N, Meeley R, Kolomiets M (2014) The novel monocot-specific 9-lipoxygenase ZmLOX12 is required to mount an effective jasmonate-mediated defense against *Fusarium verticillioides* in maize. *Mol Plant-Microbe Interact* 27(11):1263–1276. <https://doi.org/10.1094/mpmi-06-13-0184-r>
- Crossa J, Perez-Rodriguez P, Cuevas J, Montesinos-Lopez O, Jarquin D, de Los CG et al (2017) Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci* 11:961–975. <https://doi.org/10.1016/j.tplants.2017.08.011>
- Doyle J, Doyle J (1987) A rapid procedure for DNA purification from small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–15
- Dong SS, He WM, Ji JJ, Zhang C, Guo Y, Yang TL (2021) LDBlock-Show: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Brief Bioinform*. <https://doi.org/10.1093/bib/bbaa227>
- Duan C, Song F, Sun S, Guo C, Zhu Z, Wang X (2019) Characterization and molecular mapping of two novel genes resistant to phythium stalk rot in maize. *Phytopathology* 109(5):804–809. <https://doi.org/10.1094/phyto-09-18-0329-r>
- Edriss V, Gao Y, Zhang X, Jumbo MB, Makumbi D, Olsen MS et al (2017) Genomic prediction in a large African maize population. *Crop Sci* 57:2361–2371. <https://doi.org/10.2135/cropsci2016.08.0715>
- Elshire R, Glaubitz J, Sun Q, Poland J, Kawamoto K, Buckler E et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high-diversity species. *PLoS ONE* 6:e19379. <https://doi.org/10.1371/journal>
- Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol* 54:357–374. <https://doi.org/10.1146/annurev.arplant.54.031902.134907>
- Glaubitz J, Casstevens T, Lu F, Harriman J, Elshire R, Sun Q et al (2014) TASSEL-GBS: a high-capacity genotyping by sequencing analysis pipeline. *PLoS ONE* 9:e90346. <https://doi.org/10.1371/journal.pone.0090346>
- Guo R, Dhlwayo T, Mageto E, Palacios-Rojas N, Lee M, Yu D et al (2020) Genomic prediction of kernel zinc concentration in multiple maize populations using genotyping-by-sequencing and repeat amplification sequencing markers. *Front Plant Sci* 11:534. <https://doi.org/10.3389/fpls.2020.00534>
- Guo Z, Tucker D, Wang D et al (2013) Accuracy of across-environment genome-wide prediction in maize nested association mapping population. *G3 (Bethesda)*, 3(2): 263–272. <https://doi.org/10.1534/g3.112.005066>
- Han G, Lu C, Guo J, Qiao Z, Sui N, Qiu N et al (2020) C2H2 zinc finger proteins: master regulators of abiotic stress responses in plants. *Front Plant Sci* 11:115. <https://doi.org/10.3389/fpls.2020.00115>
- Holland J, Marino T, Manching H, Wisser R (2020) Genomic prediction for resistance to *Fusarium* ear rot and fumonisin contamination in maize. *Crop Sci* 60(4):1863–1875. <https://doi.org/10.1002/csc2.20163>
- Huang M, Liu X, Zhou Y, Summers R, Zhang Z (2019) BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience* 8(2):154. <https://doi.org/10.1093/gigascience/giy154>
- IBM (2022) IBM SPSS statistics algorithms. IBM Corporation. [https://www.ibm.com/docs/en/SSLVMB\\_29.0.0/pdf/IBM\\_SPSS\\_Statistics\\_Algorithms.pdf](https://www.ibm.com/docs/en/SSLVMB_29.0.0/pdf/IBM_SPSS_Statistics_Algorithms.pdf). Accessed 19 January 2023
- Jannink J, Lorenz A, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomics* 9:166–177. <https://doi.org/10.1093/bfpg/elq001>
- Jarquín D, Lemes da Silva C, Gaynor R, Poland J, Fritz A et al (2017) Increasing genomic-enabled prediction accuracy by modeling genotype · environment interactions in Kansas wheat. *Plant Genome* 10:2. <https://doi.org/10.3835/plantgenome2016.12.0130>
- Jung M, Weldekidan T, Schaff D, Paterson A, Tingey S, Hawk J (1994) Generation-means analysis and quantitative trait locus mapping of anthracnose stalk rot genes in maize. *Theor Appl Genet* 89(4):413–418. <https://doi.org/10.1007/bf00225375>
- Kuki M, Pinto R, Bertagna F, Tessmann D, Amaral A, Scapim C, Holland J (2020) Association mapping and genomic prediction for ear rot disease caused by *Fusarium verticillioides* in a tropical maize germplasm. *Crop Sci*. <https://doi.org/10.1002/csc2.20272>
- Lal S, Singh IS (1984) Breeding for resistance to downy mildews and stalk rots in maize. *Theor Appl Genet* 69:111–119
- Liu Y, Hu G, Zhang A, Loladze A, Hu X, Wang H, Qu J, Zhang X et al (2021) Genome-wide association study and genomic prediction

- of Fusarium ear rot resistance in tropical maize germplasm. *The Crop Journal* 9(2):325–341. <https://doi.org/10.1016/j.cj.2020.08.008>
- Lorenz A, Chao S, Asoro F, Heffner E, Hayashi T, Iwata H et al (2011) Genomic selection in plant breeding. *Adv Agron* 110:77–123. <https://doi.org/10.1016/B978-0-12-385531-2.00002-5>
- Ma C, Ma X, Yao L, Liu Y, Du F, Yang X, Xu M (2017) *qRfg3*, a novel quantitative resistance locus against *Gibberella* stalk rot in maize. *Theor Appl Genet* 130(8):1723–1734. <https://doi.org/10.1007/s00122-017-2921-5>
- Mageto E, Crossa J, Pérez-Rodríguez P, Dhliwayo T, Palacios-Rojas N, Lee M, Guo R, San Vicente F, Zhang X, Hindu V (2020) Genomic Prediction with Genotype by Environment Interaction Analysis for Kernel Zinc Concentration in Tropical Maize Germplasm. *G3 (Bethesda)*. 10(8):2629–2639. <https://doi.org/10.1534/g3.120.401172>
- Maize AICRP (2014) Annual report of maize pathology of all India coordinated research program on Maize. IIMR, New Delhi
- Massman J, Jung H-J, Bernardo R (2013) Genome-wide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. *Crop Sci*. <https://doi.org/10.2135/cropsci2012.02.0112>
- Meuwissen T, Hayes B, Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4):1819–1829. <https://doi.org/10.1093/genetics/157.4.1819>
- Money D, Gardner K, Migicovsky Z, Schwaninger H, Zhong G, Myles S (2015) LinkImpute: Fast and Accurate Genotype Imputation for Nonmodel Organisms. *G3 (Bethesda)*. 5(11): 2383–90. <https://doi.org/10.1534/g3.115.021667>
- Monteverde E, Rosas J, Blanco P, Pérez de Vida F, Bonnacarrère V et al (2018) Multi-environment models increase the prediction accuracy of complex traits in advanced breeding lines of rice. *Crop Sci* 58:1519–1530. <https://doi.org/10.2135/cropsci2017.09.0564>
- Mu C, Gao J, Zhou Z et al (2019) Genetic analysis of cob resistance to *F. verticillioides*: another step towards the protection of maize from ear rot. *Theor Appl Genet* 132:1049–1059. <https://doi.org/10.1007/s00122-018-3258-4>
- Mueller D, Wise K, Sisson A (2022) Corn disease loss estimates from the United States and Ontario Canada. *Plant Health Progress*. <https://doi.org/10.31274/cpn-20220328-0>
- Munkvold G (2003) Epidemiology of Fusarium diseases and their mycotoxins in maize ears. *Eur J Plant Pathol* 109:705–713. <https://doi.org/10.1023/A:1026078324268>
- Nyaga C, Gowda M, Beyene Y, Muriithi W, Makumbi D, Olsen MS, Suresh L, Bright J, Das B, Prasanna BM (2019) Genome-wide analyses and prediction of resistance to MLN in large tropical maize germplasm. *Genes (Basel)*. <https://doi.org/10.3390/genes11010016>
- Oakey H, Cullis B, Thompson R, Comadran J, Halpin C, Waugh R (2016) Genomic Selection in Multi-environment Crop Trials. *G3 (Bethesda)*. 6(5):1313–26. <https://doi.org/10.1534/g3.116.027524>
- Pè M, Gianfranceschi L, Taramino G, Tarchini R, Angelini P, Dani M et al (1993) Mapping quantitative trait loci (QTLs) for resistance to Gibberellae infection in maize. *Mol Gen Genet* 241:11–16. <https://doi.org/10.1007/BF00280195>
- Pérez P, de los Campos G. (2014) Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198:483–495. <https://doi.org/10.1534/genetics.114.164442>
- Prasanna B, Cairns J, Zaidi P et al (2021) Beat the stress: breeding for climate resilience in maize for the tropical rainfed environments. *Theor Appl Genet* 134:1729–1752. <https://doi.org/10.1007/s00122-021-03773-7>
- Qian W, He X, Chan E, Xu H, Zhang J (2011) Measuring the evolutionary rate of protein-protein interaction. *Proc Natl Acad Sci USA* 108:8725–8730
- R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rashid Z, Babu V, Sharma S, Singh P, Nair S (2022) Identification and validation of a key genomic region on chromosome 6 for resistance to Fusarium stalk rot in tropical maize. *Theor Appl Genet* 135(12):4549–4563. <https://doi.org/10.1007/s00122-022-04239-0>
- Sailer C, Babst-Kostecka A, Fischer M et al (2018) Transmembrane transport and stress response genes play an important role in adaptation of *Arabidopsis halleri* to metalliferous soils. *Sci Rep* 8:16085. <https://doi.org/10.1038/s41598-018-33938-2>
- Savary S, Willocquet L, Pethybridge S, Esker P, McRoberts N, Nelson A (2019) The global burden of pathogens and pests on major food crops. *Nat Ecol Evol* 3(3):430–439. <https://doi.org/10.1038/s41559-018-0793-y>
- Sitonik C, Suresh L, Beyene Y, Olsen M, Makumbi D, Oliver K, Das B, Bright J, Mugo S, Crossa J, Tarekegne A, Prasanna B, Gowda M (2019) 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* 132(8):2381–2399. <https://doi.org/10.1007/s00122-019-03360-x>
- Song F, Xiao M, Duan C, Li H, Zhu Z, Liu B, Sun S, Wu X, Wang X (2015) Two genes conferring resistance to Pythium stalk rot in maize inbred line Qi319. *Mol Gen Genomics* 290(4):1543–1549. <https://doi.org/10.1007/s00438-015-1019-5>
- Song JQ, Pacheco A, Alakonya A et al (2024) Genome-wide association mapping and genomic prediction of stalk rot in two mid-altitude tropical maize populations. *Crop J*. <https://doi.org/10.1016/j.cj.2024.02.004>
- Subedi S, Subedi H, Neupane S (2016) Status of maize stalk rot complex in western belts of Nepal and its integrated management. *J Maize Res Develop* 2(1):30–42. <https://doi.org/10.3126/jmrd.v2i1.16213>
- Technow F, Bürger A, Melchinger A (2013) Genomic prediction of northern corn leaf blight resistance in maize with combined or separated training sets for heterotic groups. *G3 (Bethesda)*. 3(2):197–203. <https://doi.org/10.1534/g3.112.004630>
- Tomohiro A (2011) Predictive Bayesian Model Selection. *Am J Math Manag Sci* 31(1–2):13–38. <https://doi.org/10.1080/01966324.2011.10737798>
- United Nations, Department of Economic and Social Affairs, Population Division (2019) World Population Prospects 2019: Highlights (ST/ESA/SER.A/423). <https://www.un.org/development/desa/en/>
- Van den Ende T, Abe Nijenhuis FA, van den Boorn HG, Ter Veer E, Hulshof MCM, Gisbertz SS, van Oijen MGH, van Laarhoven HWM (2019) COMplot, a graphical presentation of complication profiles and adverse effects for the curative treatment of gastric cancer: a systematic review and meta-analysis. *Front Oncol* 9:684. <https://doi.org/10.3389/fonc.2019.00684>
- Vivek B, Krishna G, Vengadassan V, Babu R, Zaidi P, Kha L, Mandal S, Grudloyma P, Takalkar S, Krothapalli K, Singh I, Ocampo E, Xingming F, Burgueño J, Azrai M, Singh R, Crossa J (2017) Use of genomic estimated breeding values results in rapid genetic gains for drought tolerance in maize. *Plant Genome*. <https://doi.org/10.3835/plantgenome2016.07.0070>
- Wang J, Zhang Z (2021) GAPIT version 3: Boosting power and accuracy for genomic association and prediction. *Genom Proteom Bioinf* 19(4):629–640. <https://doi.org/10.1016/j.gpb.2021.08.005>
- Wang C, Yang Q, Wang W, Li Y, Guo Y, Zhang D, Ma X, Song W, Zhao J, Xu M (2017) A transposon-directed epigenetic change in ZmCCT underlies quantitative resistance to *Gibberella* stalk rot in maize. *New Phytol*. 215(4):1503–1515. <https://doi.org/10.1111/nph.14688>
- Wang N, Yuan Y, Wang H, Yu D, Liu Y, Zhang A, Gowda M, Nair S, Hao Z, Lu Y, San Vicente F, Prasanna B, Li X, Zhang X (2020) Applications of genotyping-by-sequencing (GBS) in maize

- genetics and breeding. *Sci Rep* 10:16308. <https://doi.org/10.1038/s41598-020-73321-8>
- Werner C, Voss-Fels K, Miller C, Qian W, Hua W, Guan C et al (2018) Effective genomic selection in a narrow-gene pool crop with low-density markers: Asian rapeseed as an example. *Plant Genome* 11:170084. <https://doi.org/10.3835/plantgenome2017.09.0084>
- White D (1999) Compendium of corn diseases, 3rd edn. APS Press. <https://doi.org/10.1094/9780890544945>
- Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.
- Wu Y, San Vicente F, Huang K, Dhaliwayo T, Costich Denise E, Semagn K, Sudha N, Olsen M, Prasanna B, Zhang X, Babu R. (2016) Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. *Theor Appl Genet* 129:753–765. <https://doi.org/10.1007/s00122-016-2664-8>
- Xu F, Xue H (2019) The ubiquitin-proteasome system in plant responses to environments. *Plant Cell Environ* 42(10):2931–2944. <https://doi.org/10.1111/pce.13633>
- Yan J, Shah T, Warburton ML, Buckler E, McMullen M, Crouch J (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS ONE* 4(12):e8451. <https://doi.org/10.1371/journal.pone.0008451>
- Yang D, Jin D, Wang B, Zhang D, Nguyen H, Zhang C, Chen S (2005) Characterization and mapping of Rpi1, a gene that confers dominant resistance to stalk rot in maize. *Mol Genet Genomics* 274(3):229–234. <https://doi.org/10.1007/s00438-005-0016-5>
- Yang Q, Yin G, Guo Y, Zhang D, Chen S, Xu M (2010) A major QTL for resistance to *Gibberella* stalk rot in maize. *Theor Appl Genet* 121(4):673–687
- Yao Q, Xu D (2017) Bioinformatics analysis of protein phosphorylation in plant systems biology using P3DB. *Methods Mol Biol* 1558:127–138. [https://doi.org/10.1007/978-1-4939-6783-4\\_6](https://doi.org/10.1007/978-1-4939-6783-4_6)
- Ye J, Zhong T, Zhang D, Ma C, Wang L, Yao L, Zhang Q, Zhu M, Xu M (2018) The Auxin regulated protein *ZmAuxRPI* co-ordinates the balance between root growth and stalk rot disease resistance in maize. *Mol Plant* 12(3):360–373. <https://doi.org/10.1016/j.molp.2018.10.005>
- Yu G, Cu Y, Jia Y, Zhou K, Wan X, Wen Y, Xu Y, Yang K, Zhang X, Li P, Yang Z, Xu Y, Xu C (2022) Comparison of sequencing-based and array-based genotyping platforms for genomic prediction of maize hybrid performance. *CROP J*. <https://doi.org/10.1016/j.cj.2022.09.004>
- Yuan Y, Cairns J, Babu R, Gowda M, Makumbi D, Magorokosho C, Zhang A, Liu Y, Wang N, Hao Z, San Vicente F, Olsen M, Prasanna B, Lu Y, Zhang X (2019) 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*. <https://doi.org/10.3389/fpls.2018.01919>
- Zhai L, Teng F, Zheng K et al (2019) Expression analysis of Argonaute genes in maize (*Zea mays* L.) in response to abiotic stress. *Heredity* 156:27. <https://doi.org/10.1186/s41065-019-0102-z>
- Zhang D, Liu Y, Guo Y, Yang Q, Ye J, Chen S, Xu M (2012) Fine-mapping of *qRfg2*, a QTL for resistance to *Gibberella* stalk rot in maize. *Theor Appl Genet* 124:585–596
- Zhang X, Pérez-Rodríguez P, Semagn K, Beyene Y, Babu R et al (2015) Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. *Heredity* 114:291–299. <https://doi.org/10.1038/hdy.2014.99>
- Zhang A, Wang H, Beyene Y, Semagn K, Liu Y et al (2017) Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize populations. *Front Plant Sci* 8:1916. <https://doi.org/10.3389/fpls.2017.01916>
- Zhu M, Tong L, Xu M et al (2021) Genetic dissection of maize disease resistance and its applications in molecular breeding. *Mol Breeding* 41:32. <https://doi.org/10.1007/s11032-021-01219-y>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## Authors and Affiliations

Junqiao Song<sup>1,2,3</sup>  · Yubo Liu<sup>2,4</sup> · Rui Guo<sup>2,5</sup> · Angela Pacheco<sup>2</sup> · Carlos Muñoz-Zavala<sup>2</sup>  · Wei Song<sup>2,5</sup> · Hui Wang<sup>2,4</sup> · Shiliang Cao<sup>2,6</sup> · Guanghui Hu<sup>2,6</sup> · Hongjian Zheng<sup>4</sup> · Thanda Dhaliwayo<sup>2</sup>  · Felix San Vicente<sup>2</sup>  · Boddupalli M. Prasanna<sup>7</sup>  · Chunping Wang<sup>1</sup>  · Xuecai Zhang<sup>2,8,9</sup> 

✉ Chunping Wang  
chunpingw@haust.edu.cn

✉ Xuecai Zhang  
xc.zhang@cgiar.org

Junqiao Song  
songjunqiao2010@163.com

Yubo Liu  
liuyubo0218@163.com

Rui Guo  
guorui10484053@126.com

Angela Pacheco  
R.A.PACHECO@cgiar.org

Carlos Muñoz-Zavala  
C.Munoz@cgiar.org

Wei Song  
sw1717@126.com

Hui Wang  
wanghui19840109@163.com

Shiliang Cao  
caoshiliang2003@126.com

Guanghui Hu  
Gh\_Hu75@126.com

Hongjian Zheng  
hjzh6188@163.com

Thanda Dhliwayo  
D.Thanda@cgiar.org

Felix San Vicente  
f.sanvicente@cgiar.org

Boddupalli M. Prasanna  
b.m.prasanna@cgiar.org

- <sup>1</sup> Henan University of Science and Technology, Luoyang 471000, Henan, China
- <sup>2</sup> International Maize and Wheat Improvement Center (CIMMYT), 56237 Texcoco, Mexico
- <sup>3</sup> Anyang Academy of Agricultural Sciences, Anyang 455000, Henan, China
- <sup>4</sup> CIMMYT-China Specialty Maize Research Center, Crop Breeding and Cultivation Research Institute, Shanghai Academy of Agricultural Sciences, Shanghai 200063, China
- <sup>5</sup> Institute of Cereal and Oil Crops, Hebei Academy of Agricultural and Forestry Sciences, Shijiazhuang 050035, Hebei, China
- <sup>6</sup> Institute of Maize Research, Heilongjiang Academy of Agricultural Sciences, Harbin 150070, Heilongjiang, China
- <sup>7</sup> International Maize and Wheat Improvement Center (CIMMYT), Village Market, P. O. Box 1041, Nairobi 00621, Kenya
- <sup>8</sup> Institute of Crop Sciences, Chinese Academy of Agricultural Sciences (CAAS), CIMMYT-China Office, 12 Zhongguancun South Street, Beijing 100081, China
- <sup>9</sup> Nanfan Research Institute, CAAS, Sanya 572024, Hainan, China