Published June 28, 2019

Genome-Wide Analysis and Prediction of Resistance to Goss's Wilt in Maize

Julian S. Cooper, Brian R. Rice, Esperanza M. Shenstone, Alexander E. Lipka, and Tiffany M. Jamann*

Dep. of Crop Sciences, Univ. of Illinois at Urbana–Champaign, Urbana, IL 61801.

ABSTRACT Goss's bacterial wilt and leaf blight is one of the most important foliar diseases of maize (*Zea mays* L.). To date, neither large-effect resistance genes, nor practical chemical controls exist to manage the disease. Thus, the importance of discovering durable host resistance necessitates additional genetic mapping for this disease. Unfortunately, because of the biology of the pathogen and the highly significant genotype-by-environment interaction effect observed with Goss's wilt, consistent phenotyping across multiple years poses a hurdle for genetic studies and conventional breeding methods. The objective of this study was to perform a genome-wide association study (GWAS) to identify regions of the genome associated with Goss's wilt resistance as well as to use genomic prediction models to evaluate the utility of genomic selection (GS) in predicting Goss's wilt phenotypes in a panel of diverse maize lines. Using genome-wide association mapping, we were unable to identify any variants significantly associated with Goss's wilt. However, using genomic prediction we were able to train a model with an accuracy of 0.69. Taken together, this suggests that resistance to Goss's wilt is highly polygenic. In addition, when evaluating the accuracy of our prediction model under reduced marker density, it was shown that only 10,000 single nucleotide polymorphisms (SNPs), or ~20% of our total marker set, was necessary to achieve prediction accuracies similar to the full marker set. This is the first report of genomic prediction for a bacterial disease of maize, and these results highlight the potential of GS for disease resistance in maize.

Abbreviations: AUDPC, area-under-disease-progress curve; BLUP, best linear unbiased predictor; *Cmn, Clavibacter michiganensis* subsp. *nebraskensis*; FDR, false discovery rate; FHB, Fusarium head blight resistance; GBS, genotype-by-sequencing; GEBV, genomic estimated breeding value; GS, genomic selection; GWAS, genome-wide association study; MAS, marker-assisted selection; PEBV, phenotypically estimated breeding value; QTL, quantitative trait loci; RR-BLUP, ridge regression best linear unbiased prediction; SNP, single nucleotide polymorphism.

CORE IDEAS

- Goss's wilt is a complex, polygenic trait with no resistance genes or large-effect QTL.
- Genomic prediction accuracy of 0.69 achieved for Goss's wilt in panel of diverse inbred lines.
- Association mapping using a diverse panel of maize lines revealed no significant SNPs.

Since its discovery in 1969, no source of complete genetic control has been identified for Goss's bacterial wilt and leaf blight. The disease, caused by the gram-positive bacterium *Clavibacter michiganensis* subsp. *nebraskensis* (*Cmn*), produces foliar blight lesions and vascular wilt symptoms in susceptible maize varieties (Schuster, 1975). The bacterium overwinters in infected crop residue on the soil surface, where it relies on splashing rain for dissemination (Schuster, 1975). The pathogen is nonmotile (Vidaver and Mandel, 1974) and most likely to infect through wounded or damaged tissue; however, disease development has been observed in unwounded plants under high humidity conditions (Mallowa et al., 2016).

Goss's wilt is endemic from Louisiana, USA, to Alberta, Canada (Howard et al., 2015; Jackson et al., 2007; Singh et al., 2015). Yield losses >40% have been observed in susceptible maize hybrids as a result of Goss's wilt (Carson and Wicks, 1991). An estimated 12.7 billion kg of maize were lost to Goss's wilt in the United

Citation: Cooper, J.S., B.R. Rice, E.M. Shenstone, A.E. Lipka, and T.M. Jamann. 2019. Genome-wide analysis and prediction of resistance to Goss's wilt in maize. Plant Genome 12:180045. doi:10.3835/plantgenome2018.06.0045

Received 10 May 2018. Accepted 21 Feb. 2019.
*Corresponding author (tjamann@illinois.edu).

States and Canada between 2012 and 2015, making it the fourth most severe disease during this period behind northern corn leaf blight, gray leaf spot, and Fusarium stalk rot (Mueller et al., 2016). Cultural, chemical, and genetic control strategies have all been employed to curb Goss's wilt occurrence; however, no chemical control methods have been shown to decrease Goss's wilt occurrence at a statistically significant level (Mehl et al., 2015).

As with many plant pathogens, genetic resistance remains one of the best control strategies. Linkage mapping, which uses populations derived from controlled crosses to identify quantitative trait loci (QTL) associated with a trait of interest, has been used to identify several regions of the maize genome associated with Goss's wilt resistance. Singh et al. (2016) conducted joint linkage mapping for Goss's wilt resistance in three recombinant inbred line populations (B73×Oh43, B73×HP301, and B73×P39). Using these populations, 19 putative QTL on all chromosomes excluding eight were identified; however, the effect size of each QTL was small and none contributed >6% of the total observed phenotypic variance (Singh et al., 2016). An evaluation of the intermated B73 × Mo17 population identified seven putative QTL on chromosomes 1, 2, 5, 7, 8, and 10, both confirming Singh et al.'s findings and presenting the first report of a Goss's wilt QTL on chromosome 8 (Cooper et al., 2018). The QTL on chromosome 1 overlaps with the known multiple disease resistance locus *qMDR1.06*, which has been associated with reduced severity of Stewart's wilt, northern leaf blight, southern leaf blight, and a number of other maize diseases (Chung et al., 2010; Jamann et al., 2014; Wisser et al., 2006).

Another method of genetic mapping is association mapping, which takes advantage of historical recombination to identify markers in linkage disequilibrium with a trait of interest (Myles et al., 2009). The diverse subpopulations of maize, including stiff stalk, nonstiff stalk, popcorn, sweet corn, tropical, and mixed varieties, offer high allelic diversity that can be unlocked by association mapping (Flint-Garcia et al., 2005). In addition, high heritability has been reported for Goss's wilt resistance (Schaefer and Bernardo, 2013; Singh et al., 2016). This makes Goss's wilt ideal for the multi-year, multi-environment trials required for association mapping. When a GWAS for Goss's wilt was conducted on a panel of historically important maize lines from Minnesota, nine significant associations were identified (Schaefer and Bernardo, 2013).

Ideally, favorable alleles of the QTL identified for Goss's wilt resistance in maize could be directly incorporated into maize breeding programs via marker-assisted selection (MAS). However, given the absence of simply inherited resistance genes for Goss's wilt and the lack of large-effect QTL, the use of MAS is not ideal for this disease. The genetic architecture for Goss's wilt appears to be polygenic in nature, consisting of multiple loci, each with a small effect (Cooper et al., 2018; Schaefer and Bernardo, 2013; Singh et al., 2016). To effectively account for these multiple small-effect loci in a breeding program for resistance to Goss's wilt, the efficacy of GS needs to be explored. Using genome-wide molecular markers to predict the phenotypic value of an individual based on all genotypic effects (Meuwissen et al., 2001), GS is based on the theory that quantitative traits are highly polygenic and thus their variation is best captured by the modeling of all genome-wide markers. If the genetic variance can be explained by the available marker data, then it becomes possible to quantify the additive contribution of numerous, small effect loci to the phenotypic variation (Goddard and Hayes, 2007). If a sufficiently large correlation between the actual trait values and the genomic estimated breeding values (GEBVs) are observed across all prediction subsets, then GS can be used to immediately and significantly increase selection gains per unit time and expedite the breeding cycle (Heffner et al., 2010; Wong and Bernardo, 2008).

Genomic selection has been used with varying success for a number of plant diseases, most notably for modeling resistance to wheat stem rust (Poland and Rutkoski, 2016; Rutkoski et al., 2011). It was estimated that a prediction accuracy of 0.56 to 0.62 could reduce breeding cycles in wheat (*Triticum aestivum* L.) by up to twofold (Rutkoski et al., 2011, 2012). Similar studies using GS for Fusarium head blight resistance (FHB) of wheat and barley (*Hordeum vulgare* L.) found prediction accuracies between 0.41 and 0.68 using *k*-fold cross-validation (Lorenz et al., 2012; Rutkoski et al., 2012). Lorenz et al. (2012) found genomic predictions comparable to observed phenotypic means and estimated that the cost of phenotyping for FHB was four times the cost of genotyping.

In maize, GS has been evaluated for improvement of resistance to several major pathogens including Gibberella ear rot, northern corn leaf blight, southern corn leaf blight, and gray leaf spot. Riedelsheimer et al. (2013) used GS to predict disease severity of Gibberella ear rot and deoxynivalenol concentration in five double-haploid families. Validation was high within full-sib families (0.65–0.70) but was lower for both severity (0.25–0.60) and deoxynivalenol concentration (0.05–0.70) when comparing across families (Riedelsheimer et al., 2013). A study by Technow et al. (2013) evaluated the use of GS to predict northern corn leaf blight severity. Two distinct maize inbred heterotic groups were examined—dent and flint—from the University of Hohenheim breeding program. Validation within each group was high (0.64–0.71) but when attempting to predict GEBVs across heterotic groups, this accuracy dropped sharply (0.11–0.29) (Technow et al., 2013). In an evaluation of the nested association mapping populations, southern corn leaf blight displayed a prediction accuracy of 0.50 to 0.52 (Bian and Holland, 2017). However, in that same population, the prediction accuracy of gray leaf spot was only 0.22 to 0.25 (Bian and Holland, 2017). This study shows that GS may not be equally effective for all diseases.

Goss's wilt is a strong candidate for GS for a number of reasons. The genetic architecture of Goss's wilt resistance is polygenic and no large-effect loci have been identified (Cooper et al., 2018; Schaefer and Bernardo, 2013; Singh et al., 2016). Additionally, the inoculation of Goss's wilt represents a significant challenge in obtaining accurate and

consistent phenotypic data. Phenotyping for Goss's wilt is labor intensive. The nonmotile nature of the pathogen necessitates individual wounding and application of inoculum to each plant, which also increases the error observed within the experiment. In areas where Goss's wilt has not yet spread, an accurate genomic prediction model would allow screening of future commercial lines without the risk of introducing Goss's wilt to the local community.

The combined use of association mapping and genomic prediction may offer the best strategy for identifying loci associated with Goss's wilt resistance and testing whether populations can be improved using GS. Statistically significant QTL identified using association mapping would offer insight into the genetic mechanisms governing resistance, while GS would increase the efficiency of breeding for resistance in the absence of large-effect QTL. The objectives of this study were to (i) perform a GWAS on the Goodman maize diversity panel to identify putative Goss's wilt quantitative trait nucleotides and (ii) test the accuracy of GS for population improvement to Goss's wilt.

MATERIALS AND METHODS

Field Design

The Goodman maize diversity panel (Flint-Garcia et al., 2005) was grown at the Crop Science Research and Education Center in Urbana, IL, in single-row plots during the 2016 and 2017 summer field seasons. Each plot was 3.2 meters long, with 0.76-meter alleys and a row spacing of 0.762 m. Plots were machine planted at a density of 20 kernels per row, and standard agronomic practices for central Illinois were followed. Initial seed for the diversity panel was obtained from the Germplasm Resources Information Network. In 2016, disease ratings were obtained for two replications of 300 lines of the diversity panel. In 2017, disease ratings were obtained for two replications of 223 lines of the diversity panel. The difference in the number of lines tested in 2016 and 2017 was a result of seed availability. An incomplete block design was implemented in the agricolae package (de Mendiburu, 2017) of R version 3.3.2 (R Core Development Team, 2016) using resistant (FR4326) and susceptible (CQ183 and CQ184A) check lines in each block. A total of four replications were evaluated across 2 yr.

Phenotyping

The *Cmn* strain *16Cmn001* was maintained in glycerol stocks stored at −80°C for use in the 2016 and 2017 field seasons. Inoculations were conducted as described in Cooper et al. (2018). Briefly, single colonies were grown in nutrient broth yeast extract on a shaker at room temperature for 2 to 3 d. The final bacterial cell concentration was adjusted to 10^7 colony forming units (CFU) mL^{-1} using a spectrophotometer (OD_{600} = 0.05), and the bacteria were suspended in 0.1M NaCl (Pataky, 1985). Inoculations were performed twice, 1 wk apart, between the V4 and V7 stages using a modified pinprick inoculation method to simulate mechanical damage (Blanco et al., 1977; Chang et al., 1977). Disease ratings were performed three to four times, approximately every 2 wk after initial inoculation using the methods described in Cooper et al. (2018). Inbreds were scored on a per-plot basis using a 0-to-100% scale with 5% intervals meaning there were 20 possible rating categories (Cooper et al., 2018). Ratings represented the total percentage of infected leaf area, with 0% representing no symptoms and 100% denoting complete plant death (Poland and Nelson, 2011). Using the formula $A_k = \sum_{i=1}^{N-1}(\frac{y_i + y_{i+1}}{2})(t_{i+1} - t_i)$, where y_i refers to individual disease scores, t_i equals time between ratings, and n represents the number of investigations, the area-under-disease-progress curve (AUDPC) was calculated for each plot to represent disease progression throughout the season (Wilcoxson et al., 1975). Days-to-anthesis notes were taken on a plot basis, and the date was recorded when 50% of the tassels were shedding pollen.

Phenotypic Data Analysis

A $\log_{10}$ transformation was performed on the raw AUDPC scores to normalize the data. For both the combined 2016 and 2017 data and separately within each year, mixed models were fitted in R/lme4 (Bates et al., 2015) and best linear unbiased predictors (BLUPs) were calculated. For the 2016 and 2017 combined dataset genotype, year, the genotype-by-year interaction, replication, and block were included in the model (Table 1). All factors were fit as random effects. Variance components of significant factors in BLUP calculations were obtained using R/lme4 (Bates et al., 2015). To account for a potentially large genotype-by-year interaction, BLUPs were also calculated for each year individually. For individual year analyses genotype, replication and block were included in the model. Raw phenotypic data and BLUPs

Table 1. Variance component estimates and standard errors for factors included in the combined and individual year Goodman maize diversity panel analysis. Variance component estimates are based on transformed response variables. All factors were significant at $\alpha = 0.05$.

	Variance	Standard error
2016–2017		
Genotype	0.041	0.201
Year	0.051	0.226
Genotype × year	0.009	0.093
Block [replication (year)]	0.010	0.102
Error	0.017	0.129
2016		
Genotype	0.040	0.200
Replication	1.0×10^{-4}	0.014
Block (replication)	0.002	0.041
Error	0.0122	0.109
2017		
Genotype	0.065	0.254
Replication	0.009	0.096
Block (replication)	0.021	0.143
Error	0.023	0.150

are available in Supplemental File S1. Heritability was calculated using variance component estimates from the same model fitted to the combined 2016 and 2017 data, using the PROC MIXED function in SAS software (version 9.4; SAS Institute, 2013), as described by Holland et al. (2003). The following equation was used to estimate heritability on an individual plot basis:

$$\widehat{h^2} = \frac{\hat{\sigma}^2_G}{\hat{\sigma}^2_G + \hat{\sigma}^2_{GY} + \hat{\sigma}^2_{\varepsilon'}}$$

where $\hat{\sigma}^2_G$ is the genotypic variance component, $\hat{\sigma}^2_{GY}$ is the genotype-by-year interaction, and $\hat{\sigma}^2_{\varepsilon}$ is the experimental error variance. Similarly, the following equation was used to estimate the heritability on a line-mean basis:

$$\widehat{h^2} = \frac{\hat{\sigma}^2_G}{\hat{\sigma}^2_G + \frac{\hat{\sigma}^2_{GY}}{e} + \frac{\hat{\sigma}^2_{\varepsilon}}{er}}$$

where e is the harmonic mean of the number of environments and r is the number of replications. Pearson's correlation coefficients were obtained using the PROC CORR function in SAS software (version 9.4; SAS Institute, 2013).

Association Mapping

Three independent association-mapping analyses were run: one for the combined 2016 and 2017 dataset and one for each individual year. The combined Goss's wilt BLUPs across 2016 and 2017 were used for the multi-year analysis, while the BLUPs calculated for each individual year were used for the single-year analyses. The genotype-by-sequencing (GBS) SNPs, the Illumina MaizeSNP50 BeadChip (55K), and the 4K SNP marker set available on the Panzea database (4K) were combined and used in this analysis for a total of 416,376 markers (Cook et al., 2012; Romay et al., 2013; Zhao et al., 2006) (genotypic data available at http://www.panzea.org) A compressed unified mixed linear model (Zhang et al., 2010) was implemented in the R package Genome Association and Prediction Integrated Tool (Lipka et al., 2012). Principal component analysis was conducted by Lipka et al. (2013) using 34,368 nonindustry SNPs from the Illumina MaizeSNP50 BeadChip 55K marker set. The industry-derived 55K SNPs have a severe ascertainment bias between B73 and Mo17; therefore, the genetic distance between B73 and Mo17 (and related lines) will be overestimated (Ganal et al., 2011). Thus, we used the unbiased set to calculate relatedness between the lines. To account for population substructure at an appropriate level, Bayesian information criterion (Schwarz, 1978)–based backward elimination was used to select between zero to three of these first three principal components to include as fixed-effect covariates in the model. A kinship matrix to account for relatedness between the inbreds was derived in GAPIT using the 34,368 nonindustry SNPs from the Illumina MaizeSNP50 BeadChip 55K (Lipka et al., 2013). The Benjamini and Hochberg (1995) procedure was used to control the false discovery rate (FDR) at 10%; thus, any SNPs with FDR-adjusted P-values of $\leq$0.10 were declared to be significantly associated with resistance to Goss's wilt. Manhattan and quantile–quantile plots were created in R/qqman (Turner, 2014) using the GWAS results from GAPIT.

Genomic Prediction

Ridge regression best linear unbiased prediction (RR-BLUP) was performed using the rrBLUP package (Endelman, 2011) in R version 3.3.2 (R Core Development Team, 2016) using the BLUPs calculated for the combined year analysis. In RR-BLUP, a mixed linear model is fit in which all markers are considered random effects (Meuwissen et al., 2001; Whittaker et al., 2000). Each marker contributes an additive effect equal to the genetic variance divided by the total number of markers. Additionally, zero covariance between markers is assumed. Using RR-BLUP, GEBVs are predicted by estimating marker effects from a training population and then multiplying the effects of each marker by the prediction genotype to approximate phenotypic breeding values (Whittaker et al., 2000).

The 55K marker dataset, filtered to include only markers with a minor allele frequency >0.05, was used for genomic prediction. We used this genotypic dataset because the full set of GBS SNPs required additional computation time but was not necessary to achieve high prediction accuracy as shown in previous studies (Cao et al., 2017; Gowda et al., 2015). A fivefold cross-validation scheme with 100 iterations was used to generate the training and prediction populations. For each iteration, the prediction accuracy was calculated using the formula $P = r(\text{GEBV} : \text{PEBV}) / \sqrt{h^2}$. This represents the average correlation between the GEBVs and observed phenotypically estimated breeding values (PEBVs) across the five folds divided by the square root of the plot heritability (h^2). Subsets consisting of 100, 1000, 5000, 10000, and 25000 markers from the 55K marker dataset were randomly generated, and prediction accuracies across 100 iterations of a fivefold cross-validation scheme for each marker set were averaged to assess the impact of marker density and genomic prediction accuracy. Dunnett's (1955) multiple comparisons were performed between each marker density and the full 55K marker set using R/multcomp (Hothorn et al., 2008), where the full 55K marker set was set to be the control.

A simple linear regression model was fitted using the lm function in base R version 3.3.2 (R Core Development Team, 2016) to plot the relationship between GEBVs and PEBVs with GEBV as the explanatory variable and PEBV as the response variable. The slope and intercept of this linear regression was used to estimate our prediction bias (Arruda et al., 2015; Zhang et al., 2014).

RESULTS AND DISCUSSION

Characterization of Germplasm

The Goodman maize diversity panel is composed of temperate, tropical, sweet, and popcorn lines and encompasses 75% of the allelic diversity of maize (Romay et al., 2013). The diversity panel had a wide range of disease

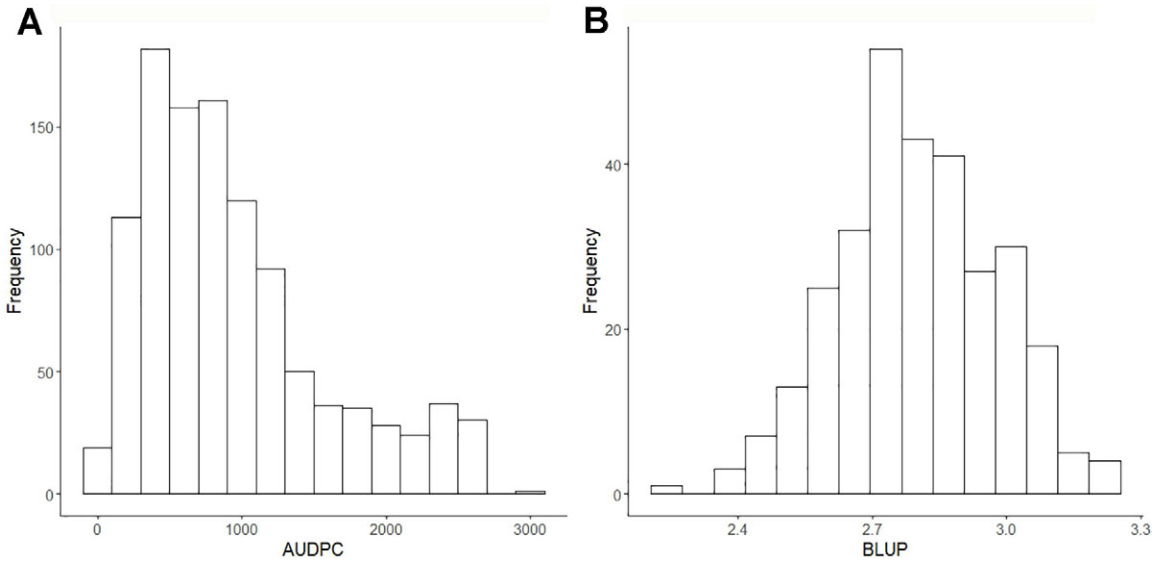

Fig. 1. Phenotypic distribution of Goss's wilt in Goodman maize diversity panel for the combined 2016 and 2017 dataset. (A) Distribution of the area under disease progress curve (AUDPC) values. These AUDPC vales were calculated using multiple visual diseased leaf area ratings taken during the growing season and reveal a tail of susceptibility within the diversity panel. (B) Best linear unbiased predictions (BLUPs). A $\log_{10}$ transformation of the data was performed to normalize the data before the genome-wide association study was conducted.

scores, from 0 to 100% diseased leaf area (Fig. 1). Severe Goss's wilt symptoms, including vascular wilt, foliar necrosis, stunting, lodging, and premature death were observed in susceptible lines. Moderately resistant lines displayed foliar lesions beyond inoculated leaves. Only highly resistant inbred lines lesions restricted to inoculated leaves.

Disease severity was compared between subpopulations within the diversity panel. Tropical and subtropical varieties displayed the highest levels of susceptibility to Goss's wilt (Fig. 2). This is in contrast to northern corn leaf blight, where tropical materials tend to harbor resistant alleles (Poland et al., 2011). Stiff and nonstiff stalk lines, while moderately resistant on average, both displayed long tails of outlying susceptibility. While trends did exist, all subpopulations exhibited a wide range of resistant and susceptible phenotypes. Overall, nonstiff stalk maize lines were the most resistant to Goss's wilt. However, because of the large number of outliers in all subpopulations, it would be unwise to rule out identifying resistance alleles from any group.

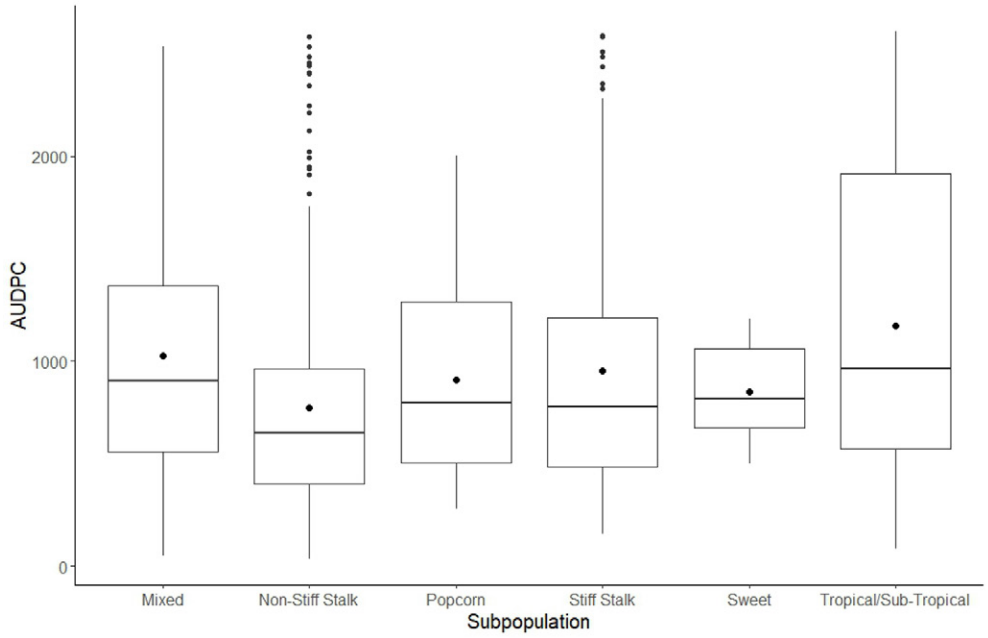

Fig. 2. Goss's wilt untransformed area under disease progress curve (AUDPC) values by subpopulation. The AUDPC values were calculated using the visual diseased leaf area scores from the combined 2016 and 2017 growing seasons.

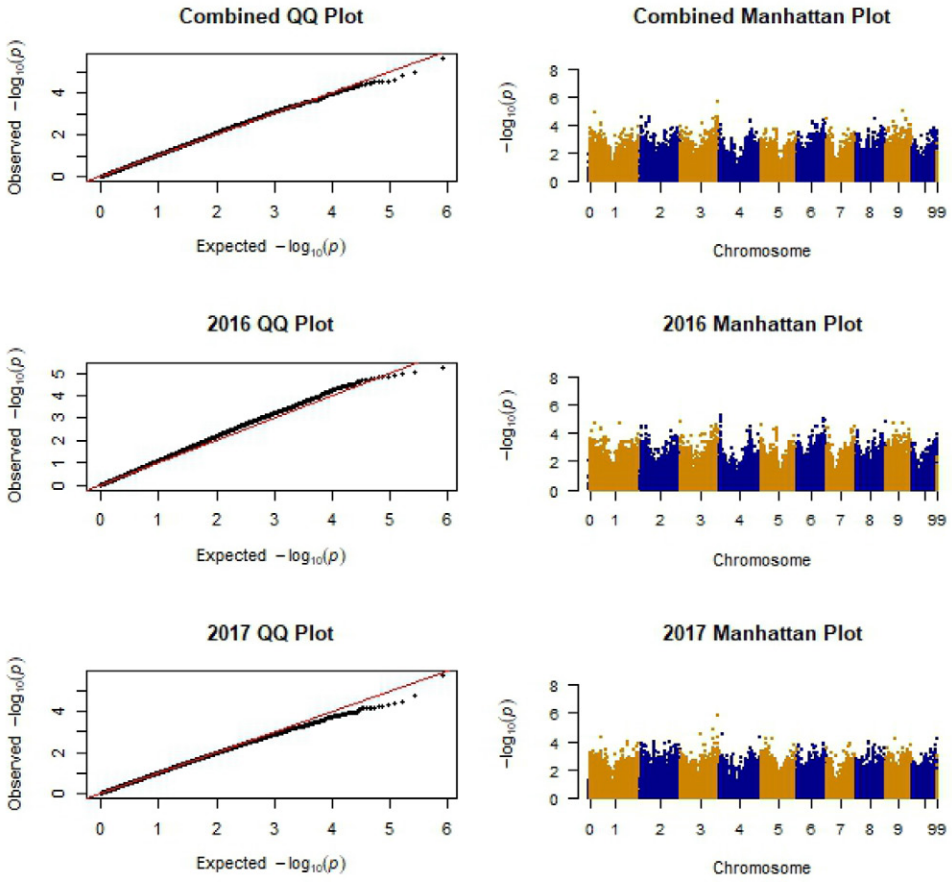

Fig. 3. Quantile–quantile and Manhattan plots for the 2016–2017 combined dataset and the individual year analyses. A false discovery rate of 10% was used to identify significant single nucleotide polymorphisms (SNPs). No significant quantitative trait loci (QTL) were detected.

Disease severity was greater in the 2016 growing season than in the 2017 season. To assess reproducibility, Pearson's correlation coefficients were calculated between replications within and across years. Replications within 2016 had a correlation coefficient of 0.782 ($P < 0.0001$), and replications within 2017 had a correlation coefficient of 0.796 ($P < 0.0001$). The correlation coefficient between the 2016 and 2017 BLUPs was 0.584 ($P < 0.0001$). Hot and dry weather conditions during inoculations in 2017 may account for some of the differences observed between years.

Heritability was calculated using the method described by Holland et al. (2003). For the 2016 and 2017 combined dataset, a plot heritability of 0.638 (standard error of 0.034) on an individual plot basis and a heritability of 0.782 (standard error of 0.028) on a line-mean basis were estimated. Previous studies have reported heritabilities for Goss's wilt ranging from 0.24 to 0.35 and 0.53 to 0.62, respectively (Cooper et al., 2018; Singh et al., 2016). The highest heritability estimates have been observed in crosses between lines with diverse phenotypes (Ngong-Nassah et al., 1992). The high heritabilities calculated in this study indicate that a large proportion of overall variation for Goss's wilt within the diversity panel can be explained by genetic differences among individuals.

Association Mapping

The mixed model fitted to the combined 2016 and 2017 data revealed a large genotype-by-year interaction, which demonstrates that environmental conditions play a role in disease resistance (Table 1). Therefore, we performed association mapping on both the combined and the individual year datasets. Despite the heritable nature of the trait and diversity of the population, no significant SNPs were detected in the combined or individual year datasets at 10% FDR (Fig. 3). Additionally, no significant associations were found when analyses were conducted by subpopulation. The lack of statistically significant associations could be due to many factors, including the allelic diversity present in the panel, the range of resistance and susceptibility within each subpopulation, low minor allele frequencies, and a lack of linkage disequilibrium between tested markers and major-effect genes. While QTL may be present within the genome, their effect size may be too small to detect through conventional GWAS methods. Increased population size and marker density may allow for the detection of significant associations.

Maturity has been shown to play a role in Goss's wilt development. Two-week-old seedlings are the most susceptible, and plants are increasingly resistant as they mature (Calub et al., 1974). The diversity panel encompasses a range of germplasm with widely

Table 2. Average predictive ability, standard deviation, prediction accuracy, and Dunnett's comparison to full 55K marker set of prediction accuracy for 100 iterations of genomic prediction using five-fold cross-validation in ridge regression best linear unbiased prediction under variable marker densities.

No. of markers	r†	Standard error	$r / \sqrt{h^2}$ ‡	Pr(>|t|)¶
100	0.359	0.006	0.449	<0.001***
1000	0.495	0.003	0.619	<0.001***
5000	0.538	0.002	0.673	<0.001***
10,000	0.546	0.001	0.673	0.182
25,000	0.551	6.9×10^{-4}	0.690	0.921
51,471	0.553	4.4×10^{-5}	0.693	NA

*** Significant at the 0.001 probability level.

† *r*, predictive ability as a Pearson's correlation coefficient.

‡ Prediction accuracy.

¶ Test to detect whether the number of markers significantly differs from full marker set. *P*-values are adjusted using the Dunnett procedure, with 51,741 markers set as the control.

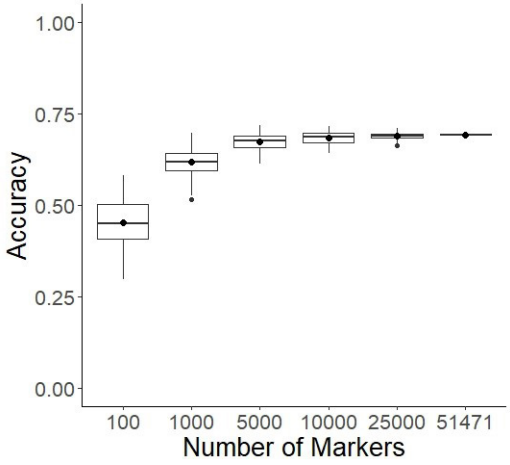

Fig. 4. Effect of marker density on predictive ability for 100 iterations of genomic prediction for data obtained from ridge regression best liner unbiased predictions (RR-BLUP) using five-fold cross-validation. Datasets including 10,000; 25,000; and 51,471 markers all achieved comparable prediction accuracies.

varying flowering times, which may have confounded our attempts to accurately rate disease severity. A Pearson correlation coefficient of −0.223 ($P < 0.0001$) was observed between days to anthesis and Goss's wilt severity, indicating later flowering lines were more resistant than earlier flowering lines. However, when days to anthesis was included as a covariate in our BLUP calculations, no significant associations were identified in the GWAS.

Genomic Prediction

Genomic prediction was proposed as a selection method for quantitative traits with putatively complex genetic architectures (Meuwissen et al., 2001), and we tested whether we could accurately predict Goss's wilt phenotypes using the diversity panel dataset despite finding no significant associations. Whole-genome prediction is capable of capturing small-effect loci that fall below the threshold for detecting significant marker effects (Heffner et al., 2009). Using RR-BLUP, a direct correlation of 0.553 ($\pm 4.4 \times 10^{-5}$) was obtained for Goss's wilt using the 55K marker set (Table 2), translating to a prediction accuracy of 0.69. In comparison, the largest published QTL for Goss's wilt resistance explains <10% of phenotypic variance (Cooper et al., 2018; Schaefer and Bernardo, 2013; Singh et al., 2016). The lack of significant associations but high prediction accuracy further supports that Goss's wilt resistance is a highly polygenic trait.

Several factors suggest that GS is a preferred option for genetic improvement for Goss's wilt. The genetic architecture of resistance to Goss's wilt is complex, and whole-genome prediction models have shown higher prediction accuracy than MAS approaches (Lorenz et al., 2011). Additionally, performing precise and high-throughput Goss's wilt inoculations poses a significant challenge in obtaining accurate phenotyping data. The nonmotile nature of the pathogen requires individual wounding and application of inoculum to each plant. Furthermore, senescence of the inoculated leaves can impair disease ratings later in the season. These factors and the high prediction accuracies displayed in our prediction model provide a convincing argument for the use of genomic prediction for breeding for Goss's wilt resistance.

Different maturation rates between temperate and tropical inbreds may have influenced disease severity, biasing our prediction model. To test for this, days to anthesis was included as a covariate in our BLUP calculations, and prediction accuracy was compared between models with and without days to anthesis. Although maturity was significantly associated with Goss's wilt severity, including it as a covariate did not significantly alter our 0.69 prediction accuracy. Therefore, it was excluded from our final results.

Ridge regression best linear unbiased prediction has been shown to function well under reduced marker densities (Habier et al., 2007). To evaluate the optimal marker density for GS of Goss's wilt, genomic prediction was performed on 100 random samples of 100, 1000, 5000, 10,000, and 25,000 SNPs from the 55K marker set. At 5000 (~10% of total marker dataset) markers, the gain in prediction accuracy due to increased marker coverage begins to lessen (Fig. 4). At 10,000 (~20% of total dataset) and 25,000 (~50% of total dataset) SNPs, prediction accuracy is no longer significantly different then predictions made using the full 51,471 SNP marker set (Table 2). There is a clear point of diminishing returns when increasing marker density for GS no longer provides substantial additional prediction accuracy. This point appears to be between 5000 and 10,000 SNPs. This indicates that exhaustive genotyping is not necessary for genomic prediction of Goss's wilt resistance. These results are consistent with previous studies, which found that for maize lethal necrosis only 1000 of 14,000 (Gowda et al., 2015) and for tar spot complex of maize 1000 of 20,000 (Cao et al., 2017) SNPs were necessary to achieve comparable accuracies to the full SNP analysis. Both of these reports are

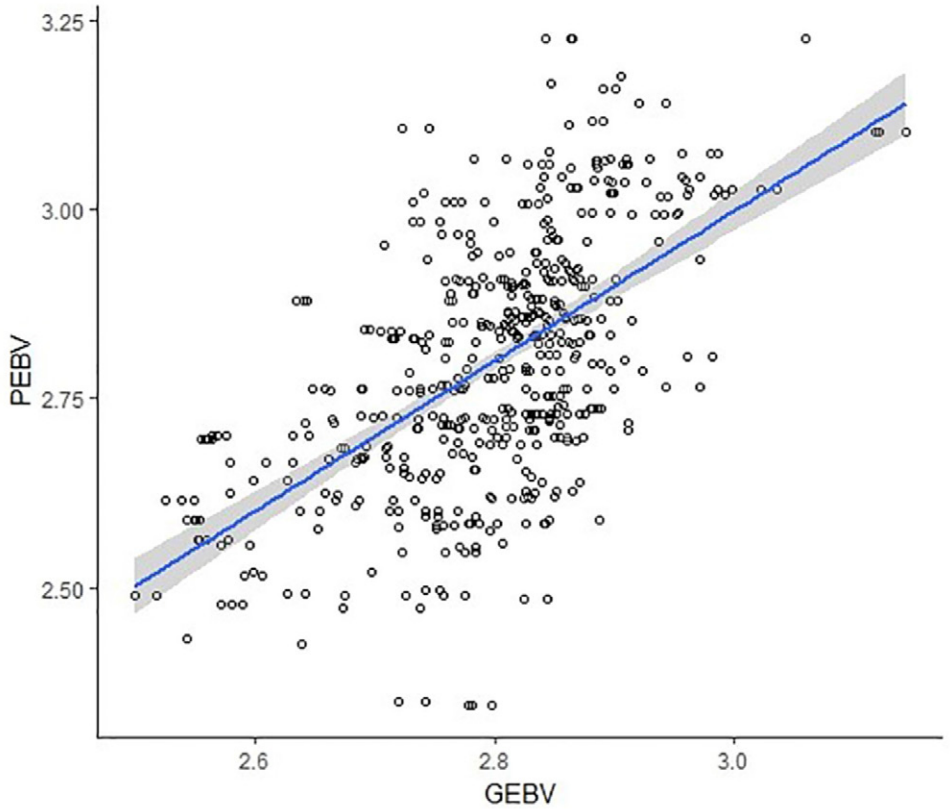

Fig. 5. Relationship between phenotypically estimated breeding values (PEBVs) and genomic-estimated breeding values (GEBVs) obtained from ridge regression best linear unbiased prediction (RR-BLUP) using five-fold cross-validation. The linear nature of the model y = 0.99x + 0.02 indicates an unbiased prediction model.

for marker densities required for diverse maize populations. We expect similar rates of linkage disequilibrium decay in our population, and resistance is primarily quantitative for all three diseases.

The bias of our predictions was plotted as a simple linear regression model with PEBV as the response variable and GEBV as the explanatory variable. The results of this fitted model suggest a strong linear relationship between these variables (Fig. 5). Values were plotted evenly along the fitted regression line of $\hat{Y} = 0.99x + 0.02$, with a narrow confidence interval near the middle of the distribution and slightly more variable values near each end (Fig. 4). An intercept of zero and a slope of one indicate an unbiased prediction model (Arruda et al., 2015). The confidence interval of our intercept was [−0.31, 0.35] and the 95% confidence interval for our slope was [0.88, 1.1]. These results indicate that our model was both accurate and unbiased.

Moderate to high prediction accuracies were obtained for Goss's wilt when comparing our results to previous studies. Goss's wilt achieved similar accuracies to northern leaf blight (0.64–0.71), Gibberella ear and stalk rot (0.65–0.70), and tar spot complex (0.55–0.74) (Cao et al., 2017; Riedelsheimer et al., 2013; Technow et al., 2013). Prediction accuracy for Goss's wilt was higher than southern leaf blight (0.50–0.52), maize lethal necrosis (0.36–0.56), and gray leaf spot (0.22–0.25) (Bian and Holland, 2017; Gowda et al., 2015). Our results indicate that genomic prediction accuracy is strong for Goss's wilt and may provide a viable approach to identify resistant materials in diverse germplasm and develop resistant materials.

CONCLUSIONS

Genomic selection is an emerging method used by breeders to circumvent long and costly phenotyping and can be effective when other methods of marker-assisted breeding fail (Heffner et al., 2010; Meuwissen et al., 2001). In addition, GS has merit when dealing with emerging diseases under quarantine or other federal regulation (Poland and Rutkoski, 2016). In such instances, GS allows for the development of resistant varieties without the potential release of a pathogen to limited geographical ranges. In this study, we were able to achieve a prediction accuracy of 0.69 for Goss's wilt in the Goodman diversity panel. Therefore, given the difficulty of phenotyping, the lack of large-effect QTL for Goss's wilt, and limited geographical range of the pathogen, GS may provide a promising alternative. This is the first report of the application of genomic prediction models to a bacterial disease of maize, and we demonstrate a high prediction accuracy for this highly polygenic trait as evidenced by a high heritability but no significant associations in the GWAS. In conclusion, where GWAS for Goss's wilt proved unsuccessful at identifying targets for MAS,

GS has emerged as successful alternative for increasing Goss's wilt resistance in maize.

Supplemental Material

Supplemental File S1. This file includes raw phenotypic values and BLUPs for combined and individual years.

Author Contributions

JC and TJ performed inoculations, disease ratings, data analysis, and drafted the manuscript. BR, ES, and AL assisted with the genomic prediction and GWAS analyses.

Conflict Of Interest

There are no conflicts of interest.

ACKNOWLEDGMENTS

We would like to acknowledge Kelsey Mehl for contribution of the pathogen strain used. Additionally, we would like to acknowledge USDA Hatch Project ILLU-802-928 for financial support.

REFERENCES

Arruda, M.P., P.J. Brown, A.E. Lipka, A.M. Krill, C. Thurber, and F.L. Kolb. 2015. Genomic selection for predicting Fusarium head blight resistance in a wheat breeding program. Plant Genome 8. doi:10.3835/plantgenome2015.01.0003

Bates, D., M. Maechler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67:1–48. doi:10.18637/jss.v067.i01

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. Roy Stat Soc B Met 57:289–300.

Bian, Y., and J.B. Holland. 2017. Enhancing genomic prediction with genome-wide association studies in multiparental maize populations. Heredity 118:585–593. doi:10.1038/hdy.2017.4

Blanco, M.H., M.G. Johnson, T.R. Colbert, and M.S. Zuber. 1977. An inoculation technique for Stewart's wilt disease of corn. Plant Dis. Rep. 61:413–416.

Calub, A.G., M.L. Schuster, C.O. Gardner, and W.A. Compton. 1974. Effect of plant age and inoculum concentration on leaf freckles and wilt of corn. Crop Sci. 14:398–401. doi:10.2135/cropsci1974.0011183X001400030017x

Cao, S., A. Loladze, Y. Yuan, Y. Wu, A. Zhang, J. Chen, et al. 2017. Genome-wide analysis of tar spot complex resistance in maize using genotyping-by-sequencing SNPs and whole-genome prediction. Plant Genome 10. doi:10.3835/plantgenome2016.10.0099

Carson, M., and Z. Wicks. 1991. Relationship between leaf freckles and wilt severity and yield losses in closely related maize hybrids. Phytopathology. 81. doi:10.1094/Phyto-81-95

Chang, C.M., A.L. Hooker, and S.M. Lim. 1977. An inoculation technique for determining Stewart's bacterial leaf blight reaction in corn. Plant Dis. Rep. 61:1077–1079.

Chung, C.L., J.M. Longfellow, E.K. Walsh, Z. Kerdieh, G. Van Esbroeck, P. Balint-Kurti, et al. 2010. Resistance loci affecting distinct stages of fungal pathogenesis: Use of introgression lines for QTL mapping and characterization in the maize–*Setosphaeria turcica* pathosystem. BMC Plant Biol. 10:103. doi:10.1186/1471-2229-10-103

Cook, J.P., M.D. McMullen, J.B. Holland, F. Tian, P. Bradbury, J. Ross-Ibarra, et al. 2012. Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol. 158:824–834. doi:10.1104/pp.111.185033

Cooper, J., P. Balint-Kurti, and T.M. Jamann. 2018. Identification of quantitative trait loci for Goss's wilt of maize. Crop Sci. 58:1192-1200. doi:10.2135/cropsci2017.10.0618

de Mendiburu, F. 2017. agricolae: Statistical procedures for agricultural research. R Foundation for Statistical Computing. https://CRAN.R-project.org/package=agricolae (accessed 1 Sept. 2017).

Dunnett, C.W. 1955. A multiple comparison procedure for comparing several treatments with a control. J. Am. Stat. Assoc. 50:1096–1121. doi:10.1080/01621459.1955.10501294

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

Flint-Garcia, S.A., A.C. Thuillet, J. Yu, G. Pressoir, S.M. Romero, S.E. Mitchell, et al. 2005. Maize association population: A high-resolution platform for quantitative trait locus dissection. Plant J. 44:1054–1064. doi:10.1111/j.1365-313X.2005.02591.x

Ganal, M.W., G. Durstewitz, A. Polley, A. Berard, E.S. Buckler, A. Charcosset, et al. 2011. A large maize (*Zea mays* L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS ONE 6:e28334. doi:10.1371/journal.pone.0028334

Goddard, M.E., and B.J. Hayes. 2007. Genomic selection. J. Anim. Breed. Genet. 124:323–330. doi:10.1111/j.1439-0388.2007.00702.x

Gowda, M., B. Das, D. Makumbi, R. Babu, K. Semagn, G. Mahuku, et al. 2015. Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. Theor. Appl. Genet. 128:1957–1968. doi:10.1007/s00122-015-2559-0

Habier, D., R.L. Fernando, and J.C. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389–2397. doi:10.1534/genetics.107.081190

Heffner, E.L., A.J. Lorenz, J.L. Jannink, and M.E. Sorrells. 2010. Plant breeding with genomic selection: Gain per unit time and cost. Crop Sci. 50:1681–1690. doi:10.2135/cropsci2009.11.0662

Heffner, E.L., M.E. Sorrells, and J.L. Jannink. 2009. Genomic selection for crop improvement. Crop Sci. 49:1–12. doi:10.2135/cropsci2008.08.0512

Holland, J.B., W.E. Nyquist and C.T. Cervantes-Martinez. 2003. Estimating and interpreting heritability for plant breeding. In: J. Janick, editor, Plant breeding reviews. John Wiley & Sons, Inc., Hoboken, NJ. p. 9–112.

Hothorn, T., F. Bretz, and P. Westfall. 2008. Simultaneous inference in general parametric models. Biom. J. 50:346–363. doi:10.1002/bimj.200810425

Howard, R.J., M.W. Harding, J. Lynn, L.M. Kawchuk, and N.M. Rasmussen. 2015. First report of Goss's bacterial wilt and leaf blight on corn caused by *Clavibacter michiganensis* subsp *nebraskensis* in Alberta, Canada. Plant Dis. 99:1034–1035. doi:10.1094/PDIS-11-14-1117-PDN

Jackson, T.A., R.M. Harveson and A.K. Vidaver. 2007. Reemergence of Goss's wilt and blight of corn to the central high plains. Plant Health Progress. doi:10.1094/PHP-2007-0919-01-BR

Jamann, T.M., J.A. Poland, J.M. Kolkman, L.G. Smith, and R.J. Nelson. 2014. Unraveling genomic complexity at a quantitative disease resistance locus in maize. Genetics 198:333–344. doi:10.1534/genetics.114.167486

Lipka, A.E., M.A. Gore, M. Magallanes-Lundback, A. Mesberg, H. Lin, T. Tiede, et al. 2013. Genome-wide association study and pathway-level analysis of tocochromanol levels in maize grain. G3: Genes, Genomes, Genet. 3:1287–1299. doi:10.1534/g3.113.006148

Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, et al. 2012. GAPIT: Genome association and prediction integrated tool. Bioinformatics 28:2397–2399. doi:10.1093/bioinformatics/bts444

Lorenz, A.J., S.M. Chao, F.G. Asoro, E.L. Heffner, T. Hayashi, H. Iwata, et al. 2011. Genomic selection in plant breeding: Knowledge and prospects. Adv. Agron. 110:77–123. doi:10.1016/B978-0-12-385531-2.00002-5

Lorenz, A.J., K.P. Smith, and J.L. Jannink. 2012. Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. Crop Sci. 52:1609–1621. doi:10.2135/cropsci2011.09.0503

Mallowa, S.O., G.Y. Mbofung, S.K. Eggenberger and R.L. Den. 2016. Infection of maize by *Clavibacter michiganensis* subsp. *nebraskensis* does not require severe wounding. Plant Dis. 100:724–731. doi:10.1094/PDIS-08-15-0923-RE

Mehl, K.M., J.D. Weems, K.A. Ames, and C.A. Bradley. 2015. Evaluation of foliar-applied copper hydroxide and citric acid for control of Goss's wilt and leaf blight of corn. Can. J. Plant Pathol. 37:160–164. doi:10.1080/07060661.2015.1012741

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

Mueller, D. S., K. A. Wise, A. J. Sisson, T. W. Allen, G. C. Bergstrom, et al. 2016. Corn yield loss estimates due to diseases in the United States and Ontario, Canada from 2012 to 2015. Plant Health Prog. 17:211–222. doi:10.1094/PHP-RS-16-0030

Myles, S., J. Peiffer, P.J. Brown, E.S. Ersoz, Z. Zhang, D.E. Costich, et al. 2009. Association mapping: Critical considerations shift from genotyping to experimental design. Plant Cell 21:2194–2202. doi:10.1105/tpc.109.068437

Ngong-Nassah, E.N., M.L. Carson, and Z.W. Wicks. 1992. Inheritance of resistance to leaf freckles and wilt caused by *Clavibacter michiganense* subsp. *nebraskense* in early maturing maize inbred lines. Phytopathology 82:142–146. doi:10.1094/Phyto-82-142

Pataky, J.K. 1985. Relationships among reactions of sweet corn hybrids to Goss wilt, Stewart's bacterial wilt, and northern corn leaf-blight. Plant Dis. 69:845–848. doi:10.1094/PD-69-845

Poland, J.A., P.J. Bradbury, E.S. Buckler, and R.J. Nelson. 2011. Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proc. Natl. Acad. Sci. USA 108:6893–6898. doi:10.1073/pnas.1010894108

Poland, J.A., and R.J. Nelson. 2011. In the eye of the beholder: The effect of rater variability and different rating scales on QTL mapping. Phytopathology 101:290–298. doi:10.1094/PHYTO-03-10-0087

Poland, J., and J. Rutkoski. 2016. Advances and challenges in genomic selection for disease resistance. Annu. Rev. Phytopathol. 54:79–98. doi:10.1146/annurev-phyto-080615-100056

R Core Development Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Riedelsheimer, C., J.B. Endelman, M. Stange, M.E. Sorrells, J.L. Jannink, and A.E. Melchinger. 2013. Genomic predictability of interconnected biparental maize populations. Genetics 194:493–503. doi:10.1534/genetics.113.150227

Romay, M.C., M.J. Millard, J.C. Glaubitz, J.A. Peiffer, K.L. Swarts, T.M. Casstevens, et al. 2013. Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. 14:R55. doi:10.1186/gb-2013-14-6-r55

Rutkoski, J., J. Benson, Y. Jia, G. Brown-Guedira, J.L. Jannink, and M. Sorrells. 2012. Evaluation of genomic prediction methods for Fusarium head blight resistance in wheat. Plant Genome 5:51–61. doi:10.3835/plantgenome2012.02.0001

Rutkoski, J.E., E.L. Heffner, and M.E. Sorrells. 2011. Genomic selection for durable stem rust resistance in wheat. Euphytica 179:161–173. doi:10.1007/s10681-010-0301-1

SAS Institute. 2013. SAS system for Windows. v. 9.4. SAS Inst. Inc., Cary, NC.

Schaefer, C.M., and R. Bernardo. 2013. Genome-wide association mapping of flowering time, kernel composition, and disease resistance in historical Minnesota maize inbreds. Crop Sci. 53:2518–2529. doi:10.2135/cropsci2013.02.0121

Schuster, M.L. 1975. Leaf freckles and wilt of corn incited by *Corynebacterium nebraskense* Schuster, Ho, Mandel, Lazar, 1972. Research Bulletin. University of Nebraska-Lincoln, Institute of Agriculture and Natural Resources, Agricultural Experiment Station, Lincoln.

Schwarz, G. 1978. Estimating the dimension of a model. Ann. Stat. 6:461–464. doi:10.1214/aos/1176344136

Singh, A., A.P. Andersen, T.A. Jackson-Ziems, and A.J. Lorenz. 2016. Mapping quantitative trait loci for resistance to Goss's bacterial wilt and leaf blight in North American maize by joint linkage analysis. Crop Sci. 56:2306–2313. doi:10.2135/cropsci2015.09.0543

Singh, R., C. Hollier, T. Burks, and R. Frazier. 2015. First report of Goss's wilt of corn caused by *Clavibacter michiganensis* subsp. *nebraskensis* in Louisiana. Plant Dis. 99:1268–1268. doi:10.1094/PDIS-08-14-0807-PDN

Technow, F., A. Burger, and A.E. Melchinger. 2013. Genomic prediction of northern corn leaf blight resistance in maize with combined or separated training sets for heterotic groups. G3: Genes, Genomes, Genet. 3:197–203. doi:10.1534/g3.112.004630

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

Vidaver, A.K., and M. Mandel. 1974. *Corynebacterium nebraskense*, a new, orange-pigmented phytopathogenic species. Int. J. Syst. Bacteriol. 24:482–485. doi:10.1099/00207713-24-4-482

Whittaker, J.C., R. Thompson, and M.C. Denham. 2000. Marker-assisted selection using ridge regression. Genet. Res. 75:249–252. doi:10.1017/S0016672399004462

Wilcoxson, R.D., B. Skovmand, and A.A. Atif. 1975. Evaluation of wheat cultivars for the ability to retard development of stem rust. Ann. Appl. Biol. 80:275–281. doi:10.1111/j.1744-7348.1975.tb01633.x

Wisser, R.J., P.J. Balint-Kurti, and R.J. Nelson. 2006. The genetic architecture of disease resistance in maize: A synthesis of published studies. Phytopathology 96:120–129. doi:10.1094/PHYTO-96-0120

Wong, C.K., and R. Bernardo. 2008. Genomewide selection in oil palm: Increasing selection gain per unit time and cost with small populations. Theor. Appl. Genet. 116:815–824. doi:10.1007/s00122-008-0715-5

Zhang, Z., E. Ersoz, C.Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, et al. 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42:355–360. doi:10.1038/ng.546

Zhang, Z., U. Ober, M. Erbe, H. Zhang, N. Gao, J.L. He, et al. 2014. Improving the accuracy of whole genome prediction for complex traits using the results of genome wide association studies. PLoS ONE 9:e93017 doi:10.1371/journal.pone.0093017

Zhao, W., P. Canaran, R. Jurkuta, T. Fulton, J. Glaubitz, E. Buckler, et al. 2006. Panzea: A database and resource for molecular and functional diversity in the maize genome. Nucleic Acids Res. 34:D752–D757. doi:10.1093/nar/gkj011