G3: Genes|Genomes|Genetics Early Online, published on July 30, 2019 as doi:10.1534/g3.119.400347

Genome-Wide Association and Gene Co-Expression Network Analyses Reveal Complex Genetics of Resistance to Goss's Wilt of Maize

Amritpal Singh[*], Guangyong Li[†,‡], Alex B. Brohammer[*], Diego Jarquin[§], Candice N. Hirsch[*], James R. Alfano[†,‡] and Aaron J. Lorenz[*,**]

[*]Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108, USA
[†]Department of Plant Pathology, University of Nebraska, Lincoln, NE 68503-0722
[‡]Center for Plant Science Innovation, University of Nebraska, Lincoln, NE 68588-0660
[§]Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68503-0915

[**]Corresponding Author: Aaron J. Lorenz, Department of Agronomy and Plant Genetics, University of Minnesota, 1991 Upper Buford Circle, St. Paul, MN 55108, USA; Phone: 612-625-6754; Fax: 612-625-1268; Email: lore0149@umn.edu

Author Email Address: Amritpal Singh (singh575@umn.edu), Guangyong Li (gli3@unl.edu), Alex B. Brohammer (broha006@umn.edu), Diego Jarquin (diego.jarquin@gmail.com), Candice N. Hirsch (cnhirsch@umn.edu), James R. Alfano (jalfano2@unl.edu), Aaron J. Lorenz (lore0149@umn.edu)

Running title: Genetic architecture of resistance to Goss's wilt of maize

Key words: Maize (*Zea mays*), genome-wide association mapping, quantitative disease resistance, Goss's wilt, weighted gene co-expression network analysis, Clavibacter, quantitative trait loci, genotyping-by-sequencing, single nucleotide polymorphism

Word counts:
Total Manuscript: 9339; Summary: 353; Introduction: 1178; Results: 1916; Discussion: 1530; Experimental Procedures: 3160; Acknowledgements: 132; Supporting Information Titles: 222; Figure Legends: 391, Tables: 457

© The Author(s) 2013. Published by the Genetics Society of America.

Abstract

Goss's bacterial wilt and leaf blight is a disease of maize caused by the gram positive bacterium *Clavibacter michiganensis* subsp. *nebraskensis* (*Cmn*). First discovered in Nebraska, Goss's wilt has now spread to major maize growing states in the United States and three provinces in Canada. Previous studies conducted using elite maize inbred lines and their hybrids have shown that resistance to Goss's wilt is a quantitative trait. The objective of this study was to further our understanding of the genetic basis of resistance to Goss's wilt by using a combined approach of genome-wide association mapping and gene co-expression network analysis. Genome-wide association analysis was accomplished using a diversity panel consisting of 555 maize inbred lines and a set of 450 recombinant inbred lines (RILs) from three bi-parental mapping populations, providing the most comprehensive screening of Goss's wilt resistance to date. Three SNPs in the diversity panel and 10 SNPs in the combined dataset, including the diversity panel and RILs, were found to be significantly associated with Goss's wilt resistance. Each significant SNP explained 1% to 5% of the phenotypic variation for Goss's wilt (total of 8 - 11%). To augment the results of genome-wide association mapping and help identify candidate genes, a time course RNA sequencing experiment was conducted using resistant (N551) and susceptible (B14A) maize inbred lines. Gene co-expression network analysis of this time course experiment identified one module of 141 correlated genes that showed differential regulation in response to *Cmn* inoculations in both resistant and susceptible lines. SNPs inside and flanking these genes explained 13.3% of the phenotypic variation. Among 1,000 random samples of genes, only 8% of

samples explained more phenotypic variance for Goss's wilt resistance than those implicated by the co-expression network analysis. While a statistically significant enrichment was not observed ($P < 0.05$), these results suggest a possible role for these genes in quantitative resistance at the field level and warrant more research on combining gene co-expression network analysis with quantitative genetic analyses to dissect complex disease resistance traits. The results of the GWAS and co-expression analysis both support the complex nature of resistance to this important disease of maize.

Goss's bacterial wilt and leaf blight of maize, caused by a gram positive bacterium *Clavibacter michiganensis* subsp. *nebraskensis* (*Cmn*), can dramatically reduce grain yield of maize if infection becomes severe enough. Studies using artificial inoculations of *Cmn* conducted during the 1990s showed yield losses up to 44% (Carson and Wicks 1991), and yield losses of 50% or more have been reported by producers (Robertson 2012). Goss' wilt was first discovered on one farm in southcentral Nebraska in 1969 (Schuster *et al.* 1972). Over the next decade, Goss's wilt spread to 58 counties in Nebraska and 34 counties across seven states, becoming a major concern for maize producers (Wysong *et al.* 1981). The incorporation of host plant resistance into maize hybrids greatly reduced the incidence of Goss's wilt, with it being only sporadically observed in the far western Corn Belt throughout the late 1980s until the mid-2000s (Jackson *et al.* 2007). Goss's wilt suddenly re-emerged around 2006 as indicated by a large increase in the number samples diagnosed with Goss's wilt submitted to plant disease diagnostic clinics (Jackson *et al.* 2007). Since 2006, Goss's wilt has continued to spread throughout North American maize growing regions, having been reported in Nebraska, Iowa, Colorado, Missouri, Indiana, Illinois, Kansas, Minnesota, North Dakota, South Dakota, Wisconsin, Texas, and Louisiana. In Canada it has been confirmed in Alberta, Manitoba, and Ontario (Ruhl *et al.* 2009; Malvick *et al.* 2010; Korus *et al.* 2011; EPPO 2014; Friskop *et al.* 2014; Howard *et al.* 2015; Singh *et al.* 2015; Hosack *et al.* 2016). In recent years Goss's wilt has ranked among major diseases of maize in North America in terms of estimated yield losses, with estimates ranging from 38.5 million

bushels in 2012 to 215.9 million bushels in 2014 in major maize producing U.S. states and Canada (Mueller and Wise 2012, 2014; Mueller *et al.* 2016).

Identifying sources of resistance to Goss's wilt and breeding resistance into maize hybrids remains a viable strategy to reduce yield losses. A great degree of variation in the level of resistance to Goss's wilt exists in maize. Maize inbred lines have shown variable response to Goss's wilt varying from resistant, intermediate, to highly susceptible based on screenings in the 1970s and 1980s with a limited number of inbred lines (Schuster *et al.*, 1972; Calub *et al.*, 1974; Wysong *et al.*, 1981, Treat *et al.*, 1990). For example, B14 and its derived lines, such as A619, have been found to be generally susceptible; Oh43 was reported to be moderately susceptible; and Mo17 was reported to be resistant (Schuster *et al.*, 1972; Calub *et al.*, 1974). Studies using classical mating designs, including diallels and generation means analyses, have indicated that resistance to Goss's wilt is under polygenic control, with additive genetic variation being the predominant form of genetic variation (Gardner and Schuster 1974; Martin *et al.* 1975; Rocheford *et al.* 1989; Treat and Tracy 1990; Ngong-Nassah *et al.* 1992).

Efforts to identify molecular markers linked to resistance to help unravel the genetic architecture of Goss's wilt have been lacking until recently because the incidence of this disease was only sporadic during the time plant geneticists and breeders were developing and adopting molecular markers, QTL mapping, and marker-assisted selection (circa late 1980s to 2000s). The re-emergence of Goss's wilt has increased interest in identifying sources of resistance, molecular markers that could be used for selection, and genes controlling resistance. Using three bi-parental maize populations

connected by a common parent, (Singh *et al.* 2016) identified eleven QTL of small effect, half of which were population specific. Schaefer and Bernardo (2013) performed genome-wide association mapping for Goss's wilt and identified several SNPs associated with resistance, each explaining no more than 10% of the phenotypic variation. The diversity panel used by Schaefer and Bernardo (2013), however, was limited in size and diversity, being comprised of only elite maize inbred lines. This limited diversity, combined with relatively few molecular markers, lowered the mapping resolution achieved by this study. Most recently, Cooper *et al.* (2018) used the Intermated B73 X Mo17 mapping population and related introgression lines to map QTL controlling Goss's wilt resistance. These authors detected several of the previously mapped QTL, including a QTL on chromosome 1 that overlapped with a locus known to be important in resistance to multiple diseases of maize (Cooper *et al.* 2018).

Transcriptomic approaches, such as RNA-seq, can help identify resistance candidate genes by quantifying changes in gene expression levels after resistance and susceptible plants have been exposed to a pathogen. Several studies, however most of them involving fungal pathogens of maize, have investigated the response of maize to its pathogens using transcriptomics (Lanubile *et al.* 2014; Liu *et al.* 2015, 2016; Araujo *et al.* 2017). From these studies, several genes have been reported to be differentially expressed in resistant and susceptible lines in response to *Fusarium* and *Bipolaris* species with enrichment in functions including pathogen recognition and signaling, growth and development, plant hormone signal transduction, and defense (Liu *et al.*, 2016; Lanubile *et al.*, 2014; Liu *et al.*, 2015). Study of gene expression changes in near-isogenic lines for

disease resistance QTL to *Fusarium* further revealed two different mechanism of resistance provided by the QTL, where one QTL conferred resistance through the expression of defense related genes and the other imparted resistance by affecting auxin signaling and transport (Liu *et al.*, 2016). Systemic acquired resistance in maize to *Colletotrichum* was found to be controlled by genes involved in Salicylic acid pathway and chromatin modification (Araujo *et al.*, 2017).

Recently published transcription profiling of resistant and susceptible maize genotypes (Doehlemann *et al.* 2008; Meyer *et al.* 2017; Yu *et al.* 2018), expression QTL analysis (Christie *et al.* 2017), and meta-analysis of expression profiling studies (Wang *et al.* 2018) for *Cercospora zeina*, *C. zeae-maydis*, and *Ustilago maydis* pathogens also revealed complex defense response and molecular interaction with maize. Different approaches revealed changes in hormonal signaling such as jasmonic acid, and salicylic acid, as well as oxidative reduction in response to pathogens. Studies also pointed towards implication of genes involved in primary defense such as leucine rich repeat receptor like kinases, PAMPs, flavonoid and terpenoid biosynthesis genes, Calmodulin like proteins, and genes encoding pathogenesis related proteins.

Combing GWAS and transcriptomic analyses provides a powerful systems approach to characterize the genetic basis of resistance to this pathogen of maize. The objectives of this study were: 1) Characterize the genetic variation for Goss's wilt within a large panel of diverse inbred lines; 2) Discover genomic regions and candidate genes controlling variation for resistance to Goss's wilt; 3) Discover co-expressed modules of genes that show changes in gene expression patterns between a resistant and susceptible

maize line in response to inoculation with *Cmn*, and compare these genes to those implicated by GWAS; 4) Quantify the amount of variation in field resistance explained by differentially expressed gene modules.

Materials and Methods

GERMPLASM AND SELECTION OF DIVERSITY PANEL FOR GENOME-WIDE ASSOCIATION ANALYSIS

Two sets of germplasm were used for the GWAS: a diversity panel of 555 inbred lines and three bi-parental linkage mapping populations selected from the maize nested association mapping (NAM) population. The diversity panel of 555 maize inbred lines was selected from a larger set of 2,815 maize inbred lines genotyped by (Romay *et al.* 2013). This panel consisted of lines from Stiff Stalk Synthetic (SSS), non-SSS, tropical, popcorn, and sweet corn genetic backgrounds. Subpopulation classification (i.e., SSS, non-SSS, popcorn) of each line was according to Romay *et al*. (2013). The first year of Goss's wilt resistance evaluation included 400 inbred lines selected from the total set of 2,815 inbred lines. To arrive at this set of 400, the initial set was first reduced to 900 lines by retaining only those that reached mid-silk within four days of B73 according to mid-silk growing degree data provided by Romay *et al*. (2013). This was done to reduce confounding effects that variation in days to flowering could have on disease ratings. A *k*-means clustering analysis was then applied to SNP data to classify the 900 lines into 400 clusters, and one line from each of the 400 clusters was randomly selected. Seed of the 400 inbred lines was obtained from the North Central Regional Plant Introduction Station (NCRPIS) in Ames, IA and increased during the summer of 2014 in the breeding

192 nursery through self-pollination. Insufficient statistical power to detect associations in the
193 first year of this study prompted an increase in the panel size for evaluations in 2015. The
194 panel evaluated in 2015 consisted of 555 lines, which was formed by the addition of 155
195 inbred lines from the set of 900 for which adequate seed quantities were immediately
196 available. No specific criteria were applied to selection of these additional 155 lines
197 beyond immediate seed availability.

198 The three bi-parental RIL populations selected from the NAM population were
199 B73 x Oh43, B73 x HP301, and B73 x P39. These same populations were used in an
200 earlier study to identify QTL for Goss's wilt through linkage mapping (Singh *et al.*,
201 2016). While the common parent, B73, is moderately resistant, the other three parents
202 were found to be comparatively susceptible to Goss's wilt in a preliminary screening.
203 Moreover, Oh43, HP301, and P39 are dent, popcorn, and sweetcorn types, respectively,
204 which allowed for the discovery of Goss's wilt resistance alleles from three distinct
205 genetic backgrounds. In 2012, 195 RILs from B73 x Oh43 population were screened for
206 Goss's wilt. In 2013, 172 RILs from B73 x Oh43 population, 141 RILs from B73 x
207 HP301 population, and 125 RILs from B73 x P39 population were evaluated. In 2014,
208 174 RILs from B73 x Oh43 population, 143 RILs from B73 x HP301 population, and 124
209 RILs from B73 x P39 population were phenotyped for Goss's wilt. The lines were
210 evaluated using a completely randomized field design with replicated checks as described
211 earlier (Singh *et al*. 2016).

GENOTYPIC DATA

ZeaGBSv1.0 dataset consisting of 681,257 GBS SNPs published by Romay *et al*. (2013) downloaded from panzea.org in 2013 was used to select the diversity panel. In 2014, ZeaGBSv2.7 version dataset of 955,690 GBS SNPs was made available for maize germplasm public germplasm including NAM RILs at panzea.org, which was then downloaded and used for the genome-wide association analysis. Briefly, the GBS data made available by Romay *et al*. (2013) was obtained using the *ApeKI* restriction enzyme as previously described (Elshire *et al.* 2011). A reference genome-based GBS pipeline in TASSEL 5.0 was used for SNP discovery with standard parameters as applied to maize Discovery Build was used to call the SNPs (Glaubitz *et al.* 2014). The imputed version of the ZeaGBSv2.7 dataset, imputed using Fast Inbred Line Library ImputatioN (FILLIN) method, was used (Swarts *et al.* 2014).

Markers with more than 80% missing data and a minor allele frequency (MAF) less than 0.05 were filtered out, leaving 342,237 SNPs for analysis. This filtered and imputed SNP dataset was used for principal component analysis (PCA), LD decay, and GWAS analyses.

DIVERSITY PANEL CHARACTERIZATION

Population structure within the diversity panel was visually assessed using PCA and ADMIXTURE, a software for model-based estimation of ancestry (Alexander *et al.* 2009). A cross validation procedure implemented in ADMIXTURE was used to initially choose the optimum number of subpopulations (K) by minimizing the cross validation error. Multiple runs of ADMIXTURE were conducted at different values of K ranging

from 3 to 20. Twenty replications were performed for each value of K. Cluster memberships for each replicate were aligned using the cluster matching software CLUMPP (Jakobsson and Rosenberg 2007). A plot of cross validation error versus K was examined, but an exact optimum could not be determined (Supplementary Figure 1). K was set to three based on visual inspection of a PCA plot of PC1 vs PC2 (Supplementary Figure 2a), subpopulation membership plots from ADMIXTURE differing in K (Supplementary Figure 2b), and subpopulation information from Romay *et al.* (2013). The average cluster membership across twenty replications of the ADMIXTURE analysis was used as a covariate in the GWAS model to account for population structure.

Decay of linkage disequilibrium (LD) was assessed as pairwise R^2 between the SNP markers within 10 kb windows using PLINK v1.07 available at http://zzz.bwh.harvard.edu/plink/ (Purcell *et al.* 2007). To assess the relationship among the lines of the diversity panel and visualize the clustering of inbred lines according to Goss's wilt resistance, a Hamming distance matrix was created using PLINK in which the distance was calculated as 1-IBS, where IBS is the identity-by-state coefficient. The distance matrix was used to create a neighbor joining tree using the method of (Saitou and Nei 1987) as implemented in *ape* package in R (Paradis *et al.* 2004).

GOSS'S WILT PHENOTYPING AND DISEASE NURSERY

In 2014, the diversity panel of 400 inbred lines was planted in a Goss's wilt nursery at the Agricultural Research and Development Center of the University of Nebraska in Mead, NE. Plots were arranged in a randomized complete block design with three replications. Susceptible line B14A, and two susceptible and two resistant

proprietary check inbred lines from Dow AgroSciences were included to assess disease development. In 2015, the diversity panel of 555 lines was planted at the same location using the same experimental design as in 2014. Inoculations with *Cmn* were carried out following the same procedure as described previously (Singh *et al.*, 2016). Briefly, wounds were created on plant leaves with motorized weed whippers and *Cmn* inoculum was sprayed within seconds of injuring the plants to ensure infection. Disease ratings were recorded 15, 30, and 45 days after inoculations (DAI). A disease rating scale of 1 to 9 on a whole plot basis was used in this study, where 1 represents complete resistance, 2 indicates disease spread less than approximately 5 cm from the point of inoculation, 3 represents limited spread but more than 5 cm from the point of inoculation, 4 indicates a large spread with lesions often extending to middle of the leaf, 5 indicates systemic infection and lesions on un-inoculated leaves, 6 indicates blight of un-inoculated leaves and wilting of plants, 7 indicates severe blight and wilt, 8 indicate severe blight and severe wilt with limited green tissue on leaves and stems of plants, and 9 represents a completely dead plot (Singh *et al.* 2016).

PHENOTYPIC ANALYSIS

Three visual ratings taken after inoculation at 15, 30, and 45 DAI were combined to calculate WMD scores. For calculation of WMD, the average of two consecutive ratings was taken and multiplied by the number of days between the two ratings. These values were summed and divided by the total number of days spanning the first and last rating (Balint-Kurti *et al.* 2010; Singh *et al.* 2016). Analysis of variance (ANOVA) was

conducted on WMD values using ASReml-R (Butler *et al.* 2009) by fitting the following model:

$$y_{ijk} = \mu + g_i + e_j + r_{k(j)} + ge_{ij} + \varepsilon_{ijk} \quad [1]$$

Where, y_{ijk} represents the WMD value, μ is the grand mean, g_i is the effect of genotype i, e_j is the effect for year j, $r_{k(j)}$ is the effect of the kth replication nested within the jth year, ge_{ij} is the interaction effect between inbred line and year, and ε_{ijk} is the residual. All effects except for the residual were treated as fixed effects. Best linear unbiased estimates (BLUEs) for WMD of the inbred lines were calculated. For estimating heritability, inbred line, inbred line-by-year interaction, and residual variances were estimated using ASReml-R by fitting these effects as IID random effects. Plot-based heritability was calculated as $H^2 = \frac{\sigma_G^2}{\sigma_G^2+\sigma_{GY}^2+\sigma_\varepsilon^2}$ where σ_G^2 is the variance among inbred lines, σ_{GY}^2 is the inbred-by-year interaction variance, and σ_ε^2 is the residual variance.

GENOME-WIDE ASSOCIATION MODEL

Genome-wide association mapping was performed using the following mixed linear model:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{w}m + \mathbf{Zu} + \mathbf{e} \quad [2]$$

where **y** is a vector of WMD BLUEs of the inbred lines; **β** is a vector of fixed effects, including an intercept and subpopulation (K = 3) effects; **X** is an incidence matrix relating **β** to **y** and contains subpopulation membership probabilities output from ADMIXTURE; *m* is the fixed SNP effect of the SNP being tested; **w** is a vector indicating the allelic state of each inbred line for the marker being tested; **u** is a vector of

298 random polygenic effects where $\mathbf{u} \sim MVN(0, \mathbf{G}\sigma_u^2)$ and **G** is a genomic relationship
299 matrix (GRM) calculated using the marker data; **Z** is a design matrix relating **u** to **y**; and
300 **e** is a vector of random residuals where $\mathbf{e} \sim MVN(0, \mathbf{I}\sigma_e^2)$. Model [2] was implemented
301 using the factored spectrally transformed linear mixed model (FaST-LMM) algorithm,
302 including calculation of the GRM (Lippert *et al.* 2011). Parameter -fileSim was added to
303 FaST-LMM command to obtain the GRM which uses realized relationship matrix from
304 SNPs to calculate the GRM.

305 The linear mixed model described was also applied to the combined dataset of
306 diversity panel (N = 555) and bi-parental populations (N = 450). The model was modified
307 to include a fixed environmental effect to account for the different environments in which
308 these sets of germplasm were evaluated. Also, subpopulation effects were extended to
309 include subpopulation effects for each of the three bi-parental populations. The
310 subpopulation effect incidence matrix, *X*, which included subpopulation memberships as
311 described above, was extended by adding three columns for each of the bi-parental
312 populations, where each column contained a 1 when the RIL in that row belonged to the
313 corresponding bi-parental population and a 0 in all other columns.

314 In order to declare SNPs as significantly associated with WMD in GWAS
315 analyses, a false-discovery rate (FDR) based on a q-value of 0.1 was used (Storey and
316 Tibshirani 2003). To calculate the percent variation explained by significant SNPs
317 (R^2_{SNP*}) after accounting for subpopulation effects, a multiple regression model was fit
318 including WMD BLUEs as the dependent variable, and subpopulation effects and those
319 of significant SNPs as independent variables (full model). A model including only

subpopulation effects as independent variables was also fit (reduced model), and R^2_{SNP*} was calculated as the difference between the full and reduced models.

The total variation in WMD BLUEs within the diversity panel that could be explained by all genotyped SNPs (polygenic background effects) was calculated using a variance component approach (Gusev *et al*., 2014):

$$\hat{g} = u + e \quad [3]$$

where $\hat{g}$ is the BLUE for inbred lines estimating using model [1], u is a random polygenic background effect where $u \sim MVN(0, G\sigma_u^2)$, and e is residual. The proportion of variation explained by all genotyped SNPs was calculated as $\sigma_u^2 \Big/ \sigma_{g+}^2 \sigma_e^2$.

HAPLOTYPE ANALYSIS

Once significant SNPs were identified by the GWAS, a haplotype analysis was performed in the regions of the genome surrounding the significant SNPs using the software Haploview (Barrett *et al.* 2005). The SNPs within 10 kb of each significant SNP were included initially to conduct the haplotype block analysis. The regions were extended beyond 10 kb for haplotype block analyses if the haplotype block length in a region was greater than 10 kb. Haplotype blocks were defined according to the four gamete rule (Wang *et al.* 2002). A cutoff of 1% was used to define the haplotype block boundaries, meaning that if a fourth two-SNP haplotype allele was observed at a frequency of greater than 1%, recombination was assumed to have occurred between the SNPs that formed the haplotype. Allele frequencies of haplotype alleles were examined

within each subpopulation of the diversity panel to determine the allele frequency differences among subpopulations.

PLANT MATERIALS AND INOCULATIONS FOR TRANSCRIPTOME PROFILING

Transcriptome profiles were obtained for B14A and N551 under both control and *Cmn* inoculated conditions. B14A was found to be susceptible and N551 was found to be resistant to *Cmn* in the field screening described above (Figure 1c). Mean field ratings of B14A and N551 were 3.4 and 1.0, respectively. B14A and N551 both belong to the stiff stalk heterotic group (Figure 1b) and are related to one another as B14 was a founder of the synthetic population from which N551 was derived (Russell 2006). The IBS, as calculated from SNPs, between N551 and B14A was 0.82 which was greater than the average IBS of 0.67 in the diversity panel.

To perform inoculations, seeds of the inbred lines were planted in plastic inserts and placed in a greenhouse for two weeks before inoculations. Inoculations were carried out when the plants were at the V2 stage (Abendroth *et al.* 2011). For inoculations, plants were transported to the lab and were inoculated using a vacuum infiltration method with the Welch 1400 Duo Seal Vacuum Pump. Inoculum was prepared from *Cmn* isolate 12038 from the Alfano Lab (University of Nebraska-Lincoln), which was tested and determined to be virulent on maize. The bacterial cells were suspended in 10mM $MgCl_2$ for measuring concentration and then were mixed into distilled water. The inoculum bacteria concentration was set to 1×10^8 colony forming units with a spectrophotometer. To increase the surface tension of the suspension, 0.005% Tween20 was added to the

362 inoculum. Control inoculations were done with water in place of the *Cmn* inoculum and
363 samples were also collected from these controls at each time point. The plants were
364 placed upside down into the inoculum and the vacuum was applied to each plant for three
365 minutes.

366 Leaf samples were collected at 0, 8, and 15 hours post inoculations (hpi). All
367 above ground leaves were cut with sterilized scissors and immediately placed in liquid
368 nitrogen after wrapping in aluminum foil. These time points were chosen to evaluate
369 genes that change expression patterns early in response to *Cmn* infection and contribute
370 to primary defense response. In a preliminary RT-PCR experiment designed to determine
371 optimal hpi, the pathogen responsive genes PR1 and PR5 were expressed within 12 to 15
372 hours. A previous study in Arabidopsis found that genes involved in early defense
373 signaling responded to elicitor flg22 as early as 30 minutes after treatment (Asai *et al.*
374 2002).

375 Three biological replicates, with one plant representing a replicate, were included
376 for each line, treatment, and hpi combination, resulting in a total of 30 samples
377 (Supplementary Table 1). RNA was isolated using the Qiagen RNeasy mini kit (Qiagen,
378 Valencia, CA, USA) based on the manufacturer's protocol, and was purified using the
379 Qiagen RNA clean-up protocol according to the manufacturer instructions.

380 **LIBRARY PREPARATION AND SEQUENCING**

381 RNA samples were submitted to the University of Minnesota Genomics Center
382 for library preparation and sequencing. RNA sizing, quantification, and purity
383 assessments were done with an Agilent Bioanalyzer (Agilent, Santa Clara, CA). Standard

36 dual indexed TruSeq RNA libraries were created. Libraries were sequenced on the Illumina HiSeq 2500 instrument using v4 chemistry (Illumina, Inc., San Diego, CA). Single end 50 bp reads were generated from the sequencing runs.

SEQUENCE QUALITY CONTROL AND TRANSCRIPT ABUNDANCE ESTIMATES

Read quality was determined using FastQC version 0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). High contents of Illumina Universal and TruSeq adapters were detected and these adapters were removed with CutAdapt version 1.8.1 (Martin 2011). The quality cutoff value was set to 20 and a minimum processed read length value of 20 were used in CutAdapt while processing the reads. After adapter trimming, the reads were aligned to the B73 v4 reference genome using a splice-site aware aligner TopHat2 version 2.0.13 (Kim *et al.* 2013) with a minimum intron size (-i) of 5 and a maximum intron size (-I) of 60000. The B73 v4 genome assembly was downloaded from Gramene Release 33 (http://www.gramene.org). Read counting was performed with HTSeq version 0.5.3 (Anders *et al.*, 2015) with strand specific option (-s) set to no, the feature type (-t) set to gene, the mode for handling overlapping reads (-m) set to union, and a minimum alignment quality (-a) of 20.

GENE CO-EXPRESSION NETWORK ANALYSIS

The raw expression matrix of 39,324 genes was filtered to remove genes with consistently no counts across the samples. The filtered count matrix was read into the DESeq2 R package and log2(x+1) transformed (Love *et al.* 2014). A coefficient of variation (CV) filter was applied to filter out the genes with consistently low expression

across the samples in which the genes with CV less than 0.5 were removed. After applying the transformation and CV filter, the expression matrix with 9,303 genes comprisong 26 samples out of 30 sequenced, was processed with R package WGCNA to identify gene co-expression modules (Langfelder and Horvath 2008). The blockwiseConsensusModules function from the WGCNA package was used to identify modules with tree cut height 0.4 and all other options set to default. WGCNA is a systems biology approach which uses the global gene expression matrix for all the genes to group the genes that show similar expression pattern together into co-expression 'modules'.

Module eigen genes that were representative of the expression profile of all the genes inside each module were visualized by a heat map to find trends in expression profiles across the control and treated samples. The normalized and transformed expression matrix was centered and standardized to calculate the Z-scores and the expression profile of all the genes inside each module across the control and treated samples was plotted using parallel coordinate plots. A gene ontology (GO) analysis was performed using the AgriGO server to test if any of the modules were enriched in functions related to defense (Du *et al.* 2010).

It was hypothesized that the gene modules responding to inoculation with *Cmn* would explain a disproportionate amount of variation in WMD as compared to a bootstrapped set of random genes. GBS SNPs from the GWAS panel that fell inside genes of such module(s) and those that flanked within one kb (determined based on genome-wide average LD decay) were used to calculate a GRM, according to [3], and the

proportion of phenotypic variation in Goss's wilt resistance explained was calculated using a variance component approach (Gusev *et al*., 2014). To establish a baseline for determining whether the genes implicated by the gene module(s) explain a disproportionate amount of phenotypic variance, variance component approach was applied to 1,000 random sets of an equal number of genes as in the module of interest. It is important to note that the sampling unit is the gene, and all SNPs inside each sampled gene were used to calculate the GRM, meaning that the number of SNPs could vary across gene samples. To compare this approach to an approach where the number of SNPs was well held constant rather than genes, 1000 random sets of 83 SNPs were also sampled and used to generation a null distribution of variation explained using the variance component approach.

DATA AVAILABILITY STATEMENT

Phenotypic data used in this study is available at figshare GSA portal. FileS1 contains raw phenotypic data collected from 555 diversity panel lines, FileS2 contains Goss's wilt BLUEs for diversity panel inbred lines, FileS3 contains subpopulation memberships (population structure) of 555 diversity panel lines computed using ADMIXTURE, FileS4 is contains raw phenotypic data for biparental populations, FileS5 contains Goss's wilt BLUEs (1005 lines) used in Combined GWAS model, FileS6 contains membership assignments or population structure information used in Combined GWAS model, and FileS7 contains gene expression data used for WGCNA. Genotypic data used in this study was obtained from https://www.panzea.org/genotypes. Data is now

449 available at CyVerse Data Store. CyVerse data store file path is
450 /iplant/home/shared/panzea/genotypes/GBS/v27/.

451 **Results**

452 **SIGNIFICANT PHENOTYPIC VARIABILITY FOR RESISTANCE TO GOSS'S**
453 **WILT EXISTS AMONG INBRED LINES**

454 A diverse panel of 555 maize inbred lines was screened for Goss's wilt resistance
455 over two years, forming the largest and most diverse set of lines screened for resistance to
456 this gram positive bacterial disease to date. Variation in flowering time was restricted to
457 reduce the confounding effects of flowering time on resistance (see *Experimental*
458 *Procedures*). Resistance scores for each plot were taken at multiple time points, which
459 were then summarized by calculating the "weighted mean disease" (WMD). A skewed
460 distribution towards resistance was observed (Figure 1a), similar to what was observed in
461 an earlier evaluation of bi-parental mapping populations (Singh *et al*. 2016). Leaf blight
462 symptoms were common following inoculation, while only a few plots showed severe
463 wilting. In an analysis of variance of the diversity panel experiment, inbred line and
464 inbred line-by-year interaction effects were found to be significant ($P < 0.05$). Broad-
465 sense heritability on a plot basis was high ($H^2 = 0.75$), indicating a combination of
466 measurement precision and a large amount of genetic variation for Goss's wilt WMD in
467 this panel.
468
469

GENETICALLY SIMILAR LINES VARY IN THEIR LEVEL OF RESISTANCE TO GOSS'S WILT

After filtering on the basis of SNP quality and minor allele frequency, a total of 342,237 SNP calls on the diversity panel remained. Linkage disequilibrium between SNPs was found to rapidly decay to 0.2 within 1000 bp, suggesting a high amount of diversity in the chosen panel. Principal component analysis and ADMIXTURE analysis (K=3) using the SNP dataset revealed subpopulations within the diversity panel (Supplementary Figure 2). The first principal component (PC1) separated lines according to stiff stalk, non-stiff, and popcorn subpopulations. A few sweet corn and tropical lines included in the panel also clustered among themselves (Supplementary Figure 2a).

A neighbor joining tree of the diversity panel lines was created to visualize the subpopulations and distribution of Goss's wilt resistance by subpopulation. Color coding of lines by subpopulation clearly indicated that the lines clustered according to subpopulation, as expected (Figure 1b). No strong pattern in the distribution of lines according to Goss's wilt resistance score was observed (Figure 1c) as groups of highly related lines included both those that were resistant and susceptible. Although elite popcorn breeding lines are generally highly susceptible to Goss's wilt (Rocheford *et al.* 1985), the popcorn lines screened as part of this diversity panel ranged from resistant to highly susceptible, indicating a high degree of variation even within popcorn (Figure 1c). It was found that lines closely related to Mo17 were generally resistant, and most of the lines related to B14 were moderately to highly susceptible, in agreement with previous findings (Calub *et al.* 1974).

GOSS'S WILT RESISTANCE IS CONTROLLED BY MANY LOCI OF SMALL EFFECT

Association analysis within the diversity panel identified three SNPs on chromosome 5 that passed a FDR cutoff of 0.10, with two of the SNPs being within just 17 bp of each other (Table 1, Figure 2a). The amount of phenotypic variation for Goss's wilt resistance explained by individual SNPs within the diversity panel was 3% to 5% (Table 1). Together the three significant SNPs explained 8% of the phenotypic variation after accounting for population structure. In contrast, the polygenic genetic variance, modeled by the genomic relationship matrix (GRM) calculated using all SNPs, explained 64% of the phenotypic variation for Goss's wilt resistance within the diversity panel.

We subsequently conducted GWAS using a combined dataset that included both the diversity panel and three RIL populations. The phenotypic data for the three RIL populations was the same as that analyzed by Singh *et al*. (2016). Ten SNPs were found to be significant at a FDR of 0.10 in this analysis; these SNPs were on chromosomes 1, 2, and 5 (Table 1, Figure 2b). Each of the SNPs identified in the combined dataset explained a small amount of phenotypic variation for Goss's wilt resistance, ranging from 1% to 3%. Physical positions of the significant SNPs found in GWAS of the combined dataset were compared to the QTL intervals of QTL mapping conducted using only the bi-parental families with fewer SNPs by Singh *et al*. (2016). Significant SNPs in GWAS on chromosome 2 co-localized with QTL peak from joint-linkage mapping and on chromosome 5 co-localized with QTL peak of the B73 x HP301 bi-parental population reported by Singh *et al*. 2016 (Figure 3). However, certain QTL were not detected,

514 including QTL detected in both joint linkage mapping and within individual families (e.g.
515 chromosome 4, 9, and 10). Other QTL specific to one bi-parental family were not
516 detected in GWAS as well (e.g. chromosome 6, and 7).

517 Identification of candidate genes could contribute to an enhanced understanding
518 of the biology underlying resistance to Goss's wilt and provide a starting point for future
519 gene confirmation studies. A candidate gene analysis was performed by extending a
520 window around a given significant SNP until the average LD decayed to 0.20 on both
521 sides of the significant SNP. If the defined window around the significant SNPs
522 overlapped with the position of a gene, the gene was declared as a potential candidate.
523 Several protein coding genes were identified within these windows with notable
524 functions such as nucleic acid binding, zinc ion binding, electron transport, and a
525 glutamate receptor protein (Table 1). In each of the regions with significant SNPs, the
526 number of candidate genes ranged from zero to three (Table 1). One candidate gene on
527 chromosome 5 was found to have a function directly related to plant defense. This gene
528 codes for glutamate receptor protein and is a member of the glutamate receptor-like gene
529 family (*GRLs*) that has been reported to play a role in plant defense response (Forde and
530 Roberts 2014). Other candidate genes did not have an apparent direct role in defense to
531 pathogens (Table 1).

532 **HAPLOTYPE BLOCK ANALYSIS OF GOSS'S WILT RESISTANCE QTL**

533 A haplotype block analysis was conducted on each genomic region harboring
534 significant SNPs to assess local LD decay rate and distribution of haplotype alleles
535 among subpopulations. The haplotype blocks at the chromosome 1 and 5 regions were

536 smaller in size due to low LD in these regions, while chromosome 2 had larger haplotype
537 blocks (Supplementary Figure 3). Four of the five significant SNPs on chromosome 2
538 were located in block 1 (5 kb), while the fifth SNP (1,977,875 bp away) was excluded
539 from the haplotype block analysis in this region (Supplementary Figure 3b). The
540 chromosome 2 region was investigated further to look at the haplotype allele frequency in
541 the diversity panel as a whole as well as within individual subpopulations. Multiple
542 alleles were observed at each haplotype block within the region of interest on
543 chromosome 2. The allelic effects were small, as the WMD of the lines carrying each of
544 the haplotype allele ranged from 2.3 to 3.0 (Table 2). Allele A1 had the lowest mean
545 Goss's wilt WMD, and it was the most common allele in all subpopulations except for
546 popcorn in which it had a frequency of only 0.02. Allele A2 had a higher Goss's wilt
547 WMD and a high frequency among the popcorn lines (93.8% of lines). The allele with
548 the highest Goss's wilt WMD, A3, had a relatively high frequency in the stiff stalk (29%
549 of lines) and non-stiff stalk (20% of lines) subpopulations compared to the popcorn
550 subpopulation (4.2% of lines), indicating there may be an allele carried by many stiff
551 stalk and non-stiff stalk inbred lines that confers higher susceptibility than the allele
552 common in the popcorn subpopulation (Table 2).

553 **A MODULE OF CO-EXPRESSED GENES SHOWED RESPONSIVENESS TO**
554 **INOCULATIONS WITH *CMN***

555 In addition to allelic variation, expression variation can contribute significantly to
556 trait variation. To identify genes with expression variation that may be contributing to
557 resistance to Goss's wilt, a time course RNA sequencing experiment of a relatively

558 resistant inbred line (N551) and susceptible inbred line (B14A) under control and
559 infected conditions was performed. These lines were chosen for this experiment because
560 they showed different levels of resistance to Goss's wilt, and they are highly related to
561 one another, reducing confounding genetic background effects (Figures 1b and 1c).
562 Weighted gene co-expression network analysis identified eight modules of co-expressed
563 genes in the time course experiment. The number of genes included in each module
564 ranged from 30 in module eight to 1,557 in module 1, with 7,373 genes not clustering
565 with any other genes. A heatmap of the eigen genes of the modules provides a high-level
566 overview of each module, allowing identification of trends in gene expression changes of
567 possible biological importance. We did not find any modules showing changes in gene
568 expression in response to infection that were specific to a genotype (Figure 4a). However,
569 module two showed changes in gene expression in both lines in response to *Cmn* (Figure
570 4a). In both N551 and B14A, the 141 genes that clustered into module two showed a
571 decrease in expression at eight hpi, and an even further decrease in expression at 15 hpi
572 (Figure 4a). Further evaluation of genes inside the second module from parallel
573 coordinate plots revealed that most genes were down-regulated in B14A and N551 at
574 eight and 15 hpi with *Cmn* (Figure 4b). Module one was genotype specific module and
575 rest of the modules appear to indicate variation at the sample level.

576 A gene ontology (GO) enrichment analysis revealed that the genes of the second
577 module were not enriched for defense functions. Most of the significant Biological
578 Process related GO terms pointed towards regulation of metabolism (Supplementary
579 Table 2). Checking the functional annotations of genes inside the second module revealed

several protein coding genes including protease inhibitor, wound induced proteins, Thiazole biosynthetic enzyme, aquaporin, DRE-binding protein, etc. (Supplementary Table 3).

We wanted to test whether SNPs within and flanking (up to one kb on either side) the genes in module two explained a disproportionate amount of the variance in Goss's wilt resistance observed in the diversity panel to help determine if these genes are important contributors to genetic variance for quantitative disease resistance. The physical positions of the SNPs used in the GWAS were aligned with the physical positions of the genes that clustered into module two. Twenty genes in module two had at least one SNP within or flanking them, with the remaining 121 genes having no SNPs within or flanking them. The total number of SNPs found was 83. These 83 SNPs were used to calculate a GRM, which was then fit into a linear mixed model to estimate the variance in WMD explained by these SNPs (Yang *et al.* 2010; Gusev *et al.* 2014). The module two SNPs explained 13.3% of the phenotypic variation for WMD, which is higher than the total percent variation explained by the significant GWAS SNPs. To determine if the variance explained by the SNPs of genes included in module two was greater than expected by a random set of genes, additional GRMs were calculated using SNPs inside and flanking 1,000 random samples of 20 genes. The number of SNPs in the 1,000 random samples ranged from 0 to 199 with a mean of 39 SNPs. The percent variation explained by the SNPs among the 1,000 random gene samples ranged from 0.0% to 64.8%. Only 8.6% of the 1,000 random samples explained more variation than the module two SNPs (Figure 4c), providing suggestive, but not strong nor statistically

significant evidence for a role of these genes in Goss's wilt resistance at the field level. A similar analysis was performed where 1000 random sets of SNPs were sampled in which the number of SNPs inside and flanking genes (83) was held constant rather than the number of genes. The variation explained by these 1000 sets of SNPs ranged from 5% to 20%, with 9.4% of the random samples explaining more variation than the 83 SNPs implicated by the gene expression analysis. These results led to the same conclusion as the results obtained where the gene was considered as the sampling unit.

Discussion

Genome-wide association mapping, made possible by the widest screening of Goss's wilt resistance reported to-date, and gene co-expression network analysis were used to increase our knowledge of the genetic basis of resistance to Goss's wilt of maize. Goss's wilt is an important disease of maize that has not been studied extensively using modern technologies of molecular genetics as compared to other maize diseases such as NCLB, SCLB, and gray leaf spot (Kump *et al.* 2011; Poland *et al.* 2011; Benson *et al.* 2015), only three published studies have used molecular markers to study the genetic basis of Goss's wilt resistance (Schaefer and Bernardo 2013; Singh *et al.* 2016; Cooper *et al.* 2018). Both GWAS and gene co-expression network analysis approaches pointed towards a complex nature of resistance to Goss's wilt, which is in agreement with the complex nature of plant signaling and defense processes as studied using model organisms (Jones and Dangl 2006; Bonardi and Dangl 2012) and hypothesized to be the case for Goss's wilt based on initial genetic studies (Martin *et al.* 1975; Rocheford *et al.* 1989; Treat *et al.* 1990).

A great degree of variation for Goss's wilt WMD was found within subpopulations, with relatively little variation between subpopulations in most cases. Our expectation was that inbred lines within certain subpopulations would be more resistant on average as compared to inbred lines from other subpopulations. This was, however, not observed in this study as resistant and susceptible lines were distributed among all the groups (Figure 1c). The popcorn subpopulation was more susceptible overall, but relatively resistant popcorn inbred lines were still found. Lines related to inbred line B14 tended to be susceptible whereas the lines related to Mo17 were found to be generally resistant, as has been reported in previous studies (Schuster *et al.* 1972; Calub *et al.* 1974).

The GWAS and calculated proportion of variance explained by all genotyped SNPs points towards a highly polygenic genetic architecture underlying Goss's wilt resistance. Similar to our previous QTL mapping study (Singh *et al*. 2016), each of the significant SNPs associated with Goss's wilt resistance in the GWAS explained a small amount of the phenotypic variation. For other important leaf diseases of maize including southern corn leaf blight, northern corn leaf blight, and gray leaf spot for which small effect SNPs have been associated with resistance (Kump *et al.*, 2011; Poland *et al.*, 2011; Benson *et al.*, 2015). Thirty-two QTL with small additive effects that together explained 80% and 93% of the phenotypic and genotypic variation, respectively, for southern corn leaf blight were reported in a GWAS of maize NAM population (Kump *et al.*, 2011). Similarly, 29 QTL were identified for northern corn leaf blight in the NAM population that explained 77% and 96% of the phenotypic and genetic variance, respectively, for

northern corn leaf blight, with each QTL being of small effect (Poland *et al.*, 2011). The studies using the NAM population explained more total phenotypic variation because of the nested linkage mapping design of the NAM population combined with a much larger population size. Nevertheless, the polygenic variance accounted for 64% of the phenotypic variation for Goss's wilt resistance in the diversity panel and thus a genomic selection approach may be more effective in breeding for Goss's wilt resistance compared to selecting for individual QTL. A similar conclusion was reported in a *Fusarium* ear rot study in maize in which the GRM explained nearly 50% of the variation for *Fusarium* ear rot and only 1.3 - 3% of the variation was explained by individual significant SNPs (Zila *et al.* 2013).

Combined GWAS (diversity panel and bi-parental populations together) increased the power of GWAS and more genomic regions were associated with Goss's wilt resistance as compared to GWAS of the diversity panel only. However, there were some QTL regions that were detected in joint-linkage and linkage mapping of individual bi-parental populations conducted by Singh *et al*. (2016) that were not detected by combined GWAS. One possible cause of not detecting certain QTL in combined GWAS approach could be different predominant linkage phases between SNPs and causal polymorphisms between bi-parental populations and the diversity panel, possibly causing effects to be cancelled out. Also, because the number of markers was much larger for GWAS, the higher statistical threshold of GWAS needed to keep the experiment-wise false positive error rate in check likely resulted failure to declare significance for smaller effect QTL.

667 In addition to the variation at the allelic level, differential regulation of genes in
668 response to pathogen may contribute to the phenotypic variation. Furthermore, the
669 number of host genes regulated by disease resistance is a basic systems biology question
670 that has been explored multiple times with different plant pathosystems. In a gene-co-
671 expression network analysis of response of Citrus to bacterium *Candidatus liberibacter*
672 spp., 3,507 genes were suggested to play a role in defense (Zheng and Zhao 2013).
673 Similarly, gene co-expression network analysis of genes regulated by immune and
674 hypersensitive resistance responses to *Blumeria graminis* f. sp *tritici* in diploid wheat
675 *Triticum urartu* indicated that 3,900 and 4,100 genes may be involved in the above two
676 types of resistance responses, respectively (Zhang *et al.* 2016). In the present study, we
677 identified one module that corresponded to response of maize inbred lines to *Cmn* was
678 identified consisting of 141 co-related genes. Most of the genes in this module showed
679 decreased expression in susceptible and resistant inbred lines at at 8 and 15 hpi. This
680 result corroborates with previous findings from molecular and co-expression network
681 studies have shown that plant immunity can be controlled by negative regulation of a
682 certain set of genes (Sato *et al.* 2010; Segonzac *et al.* 2014). Particularly, specific
683 subunits of protein Ser/Thr phosphatases have been involved in negative regulation of
684 defense signaling at different steps (Segonzac *et al.* 2014). Negative regulation of certain
685 genes within a plant immune system are required for effective functioning by preventing
686 over activation of the immune system which may cause auto immune responses such as
687 cell death, thus reducing the plant fitness (Sato *et al.* 2010).

688 A search for an overlap between genes implicated by the WGCNA and variation
689 explained in Goss's resistance at the field level was performed to determine the extent to
690 which genes in module two are important to variation in resistance among inbred lines.
691 None of the eight candidate genes identified in the GWAS analysis overlapped with the
692 genes inside of module two. Several reasons could explain this lack of strong
693 correspondence between GWAS and WGCNA. The power of the GWAS was low such
694 that it could not be expected to detect all the phenotypic variation due to possible low
695 frequencies of the causal alleles and rapid LD decay in this panel. Secondly, failure of
696 WGCNA implicated genes to explain different amount of variation than random samples
697 of genes for Goss's wilt resistance could be because the variation in gene expression of
698 these genes is not related to polymorphisms within the genes, but rather to
699 polymorphisms in cis- or trans-acting transcription factors. Thirdly, it is possible that the
700 gene expression differences between the two chosen inbred lines, B14A and N551, are
701 not representative of the gene expression patterns in the diversity panel as a whole.
702 Fourthly, the genes contributing to quantitative disease resistance in the field, 15 days or
703 more after inoculation may not be the same, or have little overlap with, genes responding
704 to infection within hours during the seedling stages. Finally, while a mixture of five
705 isolates was used for field inoculations of the GWAS, only one isolate was used to
706 inoculate B14A and N551 in the lab.
707 Using a variance component approach, we found that the genes implicated in the
708 WGCNA did not explain a disproportionate amount of the variance in Goss's wilt
709 resistance in the diversity panel using a significance level of $P < 0.05$. However, only 8%

of 1,000 random genes samples explained more phenotypic variance than those genes implicated by the WGCNA, suggesting a possible role of these genes. More research is needed to precisely determine the role of these genes in quantitative Goss's resistance at the field level. More generally, these results suggest a possible useful role for combining WGCNA with quantitative genetic analyses to determine genes underlying variation for complex disease resistance traits in crops. More research to refine these methods to attain this overall goal is warranted.

Goss's wilt and leaf blight of maize is an economically important disease, yet very few studies have been performed with the aim of elucidating the genetic basis of host resistance using modern genomic and transcriptomic techniques. This study explored an extensive amount of genetic variation for Goss's wilt resistance and identified SNPs associated with resistance using a GWAS approach. The amount of variation attributed to the background polygenic effect through a genomic relationship matrix was eight-fold higher than that explained by statistically significant QTL alone. These results suggest the genetic architecture of this disease is highly polygenic, which is consistent with the findings of our previous linkage mapping study and those using classical mating designs. Several biological processes maybe involved in quantitative resistance to Goss's wilt such which were revealed by GO analysis and functional annotations of module two genes. Results from this study provide important information about the number of QTL, effect size of QTL, and the candidate genes underlying Goss's wilt resistance, and provides a critical step towards further elucidating the genetic mechanisms of Goss's wilt resistance. Due to presence of several small effect QTL that together contribute to

resistance to Goss's wilt, a single QTL likely cannot be targeted for incorporating resistance into maize germplasm through marker-assisted selection. Rather, a genome-wide selection approach for population improvement maybe a better strategy for imparting resistance to Goss's wilt into maize breeding programs.

ACKNOWLEDGEMENTS

We are thankful to Dow AgroSciences for funding this research. ABB was supported by the DuPont Pioneer Bill Kuhn Honorary Fellowship and the MnDRIVE Global Food Ventures Graduate Student Fellowship. The Minnesota Supercomputing Institute (MSI) at the University of Minnesota and Holland Computing Center at the University of Nebraska provided computational resources that contributed to this research. We appreciate the help received from Nonoy Bandillo, Dnyaneshwar C. Kadam, Jon Luetchens, M.D. Liakat Ali, and Collin Lamkey for planting and inoculating the Goss's wilt experiments. We are also thankful to Jackson lab members: Tamra Jackson, Jae Brungardt, Brad Tharnish, and Cody Kaarstad for providing inoculum. The U.S. National Plant Germplasm System and North Central Regional Plant Introduction Station provided seeds of germplasm used in this study, for which we are very grateful.

Literature Cited

Abendroth, L. J., R. W. Elmore, M. J. Boyer, and S. K. Marlay, 2011 Corn Growth and Development.

Alexander, D. H., J. Novembre, and K. Lange, 2009 Fast model-based estimation of ancestry in unrelated individuals. Genome Res.

Araujo, A. C. G., C. G. de Freitas, D. O. Nolasco, G. da R. Fernandes, L. V. Cota *et al.*, 2017 Comparative transcriptomic analysis indicates genes associated with local and systemic resistance to Colletotrichum graminicola in maize. Sci. Rep. 7: 2483.

Asai, T., G. Tena, J. Plotnikova, M. R. Willmann, W.-L. Chiu *et al.*, 2002 MAP kinase signalling cascade in Arabidopsis innate immunity. Nature 415: 977–983.

Balint-Kurti, P. J., J. Yang, G. V. Esbroeck, J. Jung, and M. E. Smith, 2010 Use of a Maize Advanced Intercross Line for Mapping of QTL for Northern Leaf Blight Resistance and Multiple Disease Resistance. Crop Sci. 50: 458–466.

Barrett, J. C., B. Fry, J. Maller, and M. J. Daly, 2005 Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263–265.

Benson, J. M., J. A. Poland, B. M. Benson, E. L. Stromberg, and R. J. Nelson, 2015 Resistance to Gray Leaf Spot of Maize: Genetic Architecture and Mechanisms Elucidated through Nested Association Mapping and Near-Isogenic Line Analysis. PLoS Genet. 11.

Bonardi, V., and J. L. Dangl, 2012 How complex are intracellular immune receptor signaling complexes? Front. Plant Sci. 3: 237.

Butler, D., B. Cullis, A. Gilmour, and B. Gogel, 2009 *{ASReml}-R reference manual*.

Calub, A., W. A. Compton, C. O. Gardner, and M. L. Schuster, 1974 Reaction of 113 corn inbreds to leaf freckles and wilt of corn. Plant Dis. Report. 58: 956–960.

775 Carson, M. L., and Z. W. Wicks, 1991 Relationship between Leaf Freckles and Wilt
776 Severity and Yield Losses in Closely Related Maize Hybrids. Phytopathology 81:
777 95–98.
778 Christie, N., A. A. Myburg, F. Joubert, S. L. Murray, M. Carstens *et al.*, 2017 Systems
779 genetics reveals a transcriptional network associated with susceptibility in the
780 maize–grey leaf spot pathosystem. Plant J. 89: 746–763.
781 Cooper, J. S., P. J. Balint-Kurti, and T. M. Jamann, 2018 Identification of Quantitative
782 Trait Loci for Goss's Wilt of Maize. Crop Sci. 58: 1192–1200.
783 Doehlemann, G., R. Wahl, R. J. Horst, L. M. Voll, B. Usadel *et al.*, 2008 Reprogramming
784 a maize plant: transcriptional and metabolic changes induced by the fungal
785 biotroph Ustilago maydis. Plant J. 56: 181–195.
786 Du, Z., X. Zhou, Y. Ling, Z. Zhang, and Z. Su, 2010 agriGO: a GO analysis toolkit for
787 the agricultural community. Nucleic Acids Res. 38: W64-70.
788 Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto *et al.*, 2011 A Robust,
789 Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species.
790 PLOS ONE 6: e19379.
791 EPPO, 2014 PQR database. Paris, France: European and Mediterranean Plant Protection
792 Organization.
793 Forde, B. G., and M. R. Roberts, 2014 Glutamate receptor-like channels in plants: a role
794 as amino acid sensors in plant defence? F1000Prime Rep. 6: 37.

795 Friskop, A., K. Kinzer, M. McConnell, Z. Liu, K. Korus *et al.*, 2014 First Report of
796 Goss's Bacterial Leaf Blight and Wilt of Corn Caused by Clavibacter
797 michiganensis subsp. nebraskensis in North Dakota. Plant Dis. 98: 1739–1739.
798 Gardner, C. O., and M. L. Schuster, 1974 Genetic studies of susceptibility to bacterial
799 leaf freckles and wilt, Corynebacterium nebraskense. Maize Genet. Coop. Newsl.
800 47: 155–157.
801 Glaubitz, J. C., T. M. Casstevens, F. Lu, J. Harriman, R. J. Elshire *et al.*, 2014 TASSEL-
802 GBS: a high capacity genotyping by sequencing analysis pipeline. PloS One 9:
803 e90346.
804 Gusev, A., S. H. Lee, G. Trynka, H. Finucane, B. J. Vilhjálmsson *et al.*, 2014 Partitioning
805 Heritability of Regulatory and Cell-Type-Specific Variants across 11 Common
806 Diseases. Am. J. Hum. Genet. 95: 535–552.
807 Hosack, P. K., L. E. Sweets, G. L. Miller, M. J. Calcutt, and B. E. Arenz, 2016 First
808 Report of Goss's Bacterial Wilt and Leaf Blight on Corn Caused by Clavibacter
809 michiganensis subsp. nebraskensis in Missouri. Plant Dis.
810 Howard, R. J., M. W. Harding, J. Lynn, L. M. Kawchuk, and N. M. Rasmussen, 2015
811 First Report of Goss's Bacterial Wilt and Leaf Blight on Corn Caused by
812 Clavibacter michiganensis subsp. nebraskensis in Alberta, Canada. Plant Dis. 99:
813 1034.
814 Jackson, T. A., R. M. Harveson, and A. K. Vidaver, 2007 Goss's bacterial wilt and leaf
815 blight of corn. NebGuide Coop Ext Serv IANR Univ Nebr Linc. G1675.

816 Jakobsson, M., and N. A. Rosenberg, 2007 CLUMPP: a cluster matching and
817 permutation program for dealing with label switching and multimodality in
818 analysis of population structure. Bioinformatics 23: 1801–1806.
819 Jones, J. D. G., and J. L. Dangl, 2006 The plant immune system. Nature 444: 323–329.
820 Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley *et al.*, 2013 TopHat2: accurate
821 alignment of transcriptomes in the presence of insertions, deletions and gene
822 fusions. Genome Biol. 14: R36.
823 Korus, K. A., A. D. Timmerman, R. D. French-Monar, and T. A. Jackson, 2011 First
824 Report of Goss's Bacterial Wilt and Leaf Blight (Clavibacter michiganensis subsp
825 nebraskensis) of Corn in Texas. Plant Dis. 95: 73–73.
826 Kump, K. L., P. J. Bradbury, R. J. Wisser, E. S. Buckler, A. R. Belcher *et al.*, 2011
827 Genome-wide association study of quantitative resistance to southern leaf blight
828 in the maize nested association mapping population. Nat. Genet. 43: 163-U120.
829 Langfelder, P., and S. Horvath, 2008 WGCNA: an R package for weighted correlation
830 network analysis. BMC Bioinformatics 9: 559.
831 Lanubile, A., A. Ferrarini, V. Maschietto, M. Delledonne, A. Marocco *et al.*, 2014
832 Functional genomic analysis of constitutive and inducible defense responses to
833 Fusarium verticillioides infection in maize genotypes with contrasting ear rot
834 resistance. BMC Genomics 15: 710.
835 Lippert, C., J. Listgarten, Y. Liu, C. M. Kadie, R. I. Davidson *et al.*, 2011 FaST linear
836 mixed models for genome-wide association studies. Nat. Methods 8: 833–835.

837 Liu, M., J. Gao, F. Yin, G. Gong, C. Qin *et al.*, 2015 Transcriptome Analysis of Maize
838 Leaf Systemic Symptom Infected by Bipolaris zeicola. PLOS ONE 10: e0119858.
839 Liu, Y., Y. Guo, C. Ma, D. Zhang, C. Wang *et al.*, 2016 Transcriptome analysis of maize
840 resistance to Fusarium graminearum. BMC Genomics 17: 477.
841 Love, M. I., W. Huber, and S. Anders, 2014 Moderated estimation of fold change and
842 dispersion for RNA-seq data with DESeq2. Genome Biol. 15: 550.
843 Malvick, D., R. Syverson, D. Mollov, and C. A. Ishimaru, 2010 Goss's Bacterial Blight
844 and Wilt of Corn Caused by Clavibacter michiganensis subsp nebraskensis
845 Occurs in Minnesota. Plant Dis. 94: 1064–1064.
846 Martin, M., 2011 Cutadapt removes adapter sequences from high-throughput sequencing
847 reads. EMBnet.journal 17: 10–12.
848 Martin, P. R., C. O. Gardner, A. G. Calub, and M. L. Schuster, 1975 Inheritance
849 ofsusceptibility and tolerance to leaf freckles and wilt (Corynebacterium
850 nebraskense) of corn. Maize Genet. Coop. Newsl. 49: 137–138.
851 Meyer, J., D. K. Berger, S. A. Christensen, and S. L. Murray, 2017 RNA-Seq analysis of
852 resistant and susceptible sub-tropical maize lines reveals a role for kauralexins in
853 resistance to grey leaf spot disease, caused by Cercospora zeina. BMC Plant Biol.
854 17: 197.
855 Mueller, D., and K. Wise, 2012 Corn disease loss estimates from the United States and
856 Ontario, Canada. Purdue Ext. Publ. BP-96-12—W.
857 Mueller, D., and K. Wise, 2014 Corn disease loss estimates from the United States and
858 Ontario, Canada. Purdue Ext. Publ. BP-96-14—W.

Mueller, D., K. Wise, A. Sisson, T. Allen, G. 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. 211–222.

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.

Paradis, E., J. Claude, and K. Strimmer, 2004 APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20: 289–290.

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. U. S. A. 108: 6893–6898.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira *et al.*, 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81: 559–575.

Robertson, A. E., 2012 Goss's wilt: a 2012 recap and looking ahead to 2013. Proc 2013 Wis. Crop Manag. Conf. 52.

Rocheford, T. R., C. A. Gardner, and A. K. Vidaver, 1985 Chromosomal location of additional genes for resistance to Corynebacterium (Clavibacter) michiganense ssp. nebraskense. Maize Genet. Coop. Newsl. 60: 73–74.

Rocheford, T. R., C. O. Gardner, and A. K. Vidaver, 1989 Genetic studies of resistance inmaize (Zea mays L.) to Goss's bacterial wilt and blight (Clavibacter michiganensis subsp. nebraskensis). J Hered 80: 351–356.

881 Romay, M. C., M. J. Millard, J. C. Glaubitz, J. A. Peiffer, K. L. Swarts *et al.*, 2013
882 Comprehensive genotyping of the USA national maize inbred seed bank. Genome
883 Biol. 14: R55.
884 Ruhl, G., K. Wise, T. Creswell, A. Leonberger, and C. Speers, 2009 First Report of
885 Goss's Bacterial Wilt and Leaf Blight on Corn Caused by Clavibacter
886 michiganensis subsp nebraskensis in Indiana. Plant Dis. 93: 841–841.
887 Russell, W. K., 2006 Registration of N551 and N552 Parental Inbred Lines of Maize.
888 Crop Sci. 46: 1014–1015.
889 Saitou, N., and M. Nei, 1987 The neighbor-joining method: a new method for
890 reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406–425.
891 Sato, M., K. Tsuda, L. Wang, J. Coller, Y. Watanabe *et al.*, 2010 Network Modeling
892 Reveals Prevalent Negative Regulatory Relationships between Signaling Sectors
893 in Arabidopsis Immune Signaling. PLOS Pathog. 6: e1001011.
894 Schaefer, C. M., and R. Bernardo, 2013 Genomewide Association Mapping of Flowering
895 Time, Kernel Composition, and Disease Resistance in Historical Minnesota
896 Maize Inbreds. Crop Sci. 53: 2518–2529.
897 Schuster, M. L., B. Hoff, M. Mandel, and I. Lazar, 1972 Leaf freckles and wilt, a new
898 corn disease. 27th Ann Corn Sorghum Res. Congr Chic. Ill Dec12-14 176–191.
899 Segonzac, C., A. P. Macho, M. Sanmartín, V. Ntoukakis, J. J. Sánchez-Serrano *et al.*,
900 2014 Negative control of BAK1 by protein phosphatase 2A during plant innate
901 immunity. EMBO J. 33: 2069–2079.

902 Singh, A., A. P. Andersen, T. A. Jackson-Zeims, and A. J. Lorenz, 2016 Mapping
903 quantitative trait loci for resistance to Goss's bacterial wilt and leaf blight in
904 North American Maize by joint linkage analysis. Crop Sci. 56: 1–8.
905 Singh, R., C. Hollier, T. Burks, and R. Frazier, 2015 First Report of Goss's Wilt of Corn
906 Caused by Clavibacter michiganensis subsp. nebraskensis in Louisiana. Plant Dis.
907 99: 1268–1268.
908 Storey, J. D., and R. Tibshirani, 2003 Statistical significance for genomewide studies.
909 Proc. Natl. Acad. Sci. 100: 9440–9445.
910 Swarts, K., H. Li, R. Navarro, J. Alberto, D. An *et al.*, 2014 Novel Methods to Optimize
911 Genotypic Imputation for Low-Coverage, Next-Generation Sequence Data in
912 Crop Plants. Plant Genome 7.
913 Treat, C. L., and W. F. Tracy, 1990 Inheritance of resistance to Goss's wilt in sweet corn.
914 J Am Soc Hortic Sci 114: 672–674.
915 Treat, C. L., W. F. Tracy, P. N. Drolsom, and J. G. Coors, 1990 Inheritance of Resistance
916 to Gosss Wilt in Maize. Crop Sci. 30: 893–896.
917 Wang, N., J. M. Akey, K. Zhang, R. Chakraborty, and L. Jin, 2002 Distribution of
918 Recombination Crossovers and the Origin of Haplotype Blocks: The Interplay of
919 Population History, Recombination, and Mutation. Am. J. Hum. Genet. 71: 1227–
920 1234.
921 Wang, J., Y. Zhang, J. Du, X. Pan, L. Ma *et al.*, 2018 Combined analysis of genome-wide
922 expression profiling of maize (Zea mays L.) leaves infected with Ustilago maydis.
923 Genome 61: 505–513.

Wysong, D. S., B. Doupnik, and L. Lane, 1981 Goss's wilt and corn lethal necrosis - can they become a major problem? Proc. Annu. Corn Sorghum Ind. Res. Conf. 36: 104–152.

Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders *et al.*, 2010 Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42: 565–569.

Yu, Y., J. Shi, X. Li, J. Liu, Q. Geng *et al.*, 2018 Transcriptome analysis reveals the molecular mechanisms of the defense response to gray leaf spot disease in maize. BMC Genomics 19: 742.

Zhang, J., H. Zheng, Y. Li, H. Li, X. Liu *et al.*, 2016 Coexpression network analysis of the genes regulated by two types of resistance responses to powdery mildew in wheat. Sci. Rep. 6: srep23805.

Zheng, Z.-L., and Y. Zhao, 2013 Transcriptome comparison and gene coexpression network analysis provide a systems view of citrus response to 'CandidatusLiberibacter asiaticus' infection. BMC Genomics 14: 27.

Zila, C. T., L. F. Samayoa, R. Santiago, A. Butrón, and J. B. Holland, 2013 A genome-wide association study reveals genes associated with fusarium ear rot resistance in a maize core diversity panel. G3 Bethesda Md 3: 2095–2104.

FIGURE LEGENDS

Figure 1. Genetic relationship among the inbred lines of the diversity panel and distribution of Goss's wilt among the inbred lines. (a) Histogram showing the distribution of weighted mean disease scores for inbred lines included in the diversity panel. Vertical dotted lines display the check lines B73 and B14A; (b) Neighbor joining tree of 555 lines of the diversity panel created from a distance matrix calculated using the GBS SNP data. Labels are colored coded by the six subpopulation groups within the diversity panel; (c) The same neighbor joining tree as in (b) but labels are color coded according to Goss's weighted mean disease values. Red color indicates that a line is susceptible and green color represents resistant lines. Specific groups are zoomed to indicate trends in Goss's wilt resistance distribution. For example, while most of the B14-related lines were susceptible while Mo17-related lines were relatively resistant. N551 is present in the B14 zoomed window along with B14A.

Figure 2. Manhattan plots showing results from genome-wide association mapping. (a) GWAS results on the diversity panel (N =555) displayed in a Manhattan plot where the y-axis is the negative $\log_{10}$ of p-values for the SNPs from model [2]. Associations that passed the false-discovery rate of 0.10 are colored green; (b) genome-wide association mapping results on the combined dataset (N =1005) displayed in a Manhattan plot where the y-axis is the $-\log_{10}(P)$ for the SNPs from model [2]. Associations that passed the false-discovery rate of 0.10 are colored green.

Figure 3. Comparison of physical positions of QTL detected in bi-parental linkage mapping and GWAS. Bi-parental linkage mapping was conducted by Singh *et al.* (2016), and significant SNPs from GWAS in the combined dataset co-located on chromosomes 1, 2, and 5. The x-axis of each plot represents the physical position of each SNP, and the y-axis displays the negative $\log_{10}$ of p-values for each SNP included in the GWAS. Gray colored solid points represent all SNPs used in GWAS. Significant SNPs in the GWAS are indicated by dark green dots, and 2-LOD support interval of QTL detected by Singh *et al.* (2016) are shown by the red or blue windows.

Figure 4. Expression patterns of eight modules identified using weighted gene co-expression network analysis (WGCNA) across different samples. (a) Heatmap of eigen genes of eight modules. An eigen gene is representative of the gene expression pattern of genes inside that module. Labels of the columns are inbred line_treatment_hours_rep; (b) Normalized expression of genes inside module two obtained from WGCNA across all the samples. This module showed changes in gene expression in both susceptible line B14A as well as resistant line N551 in response to *Cmn*; (c) Distribution of phenotypic variance explained by SNPs inside and flanking the 1,000 random samples of genes. As a comparison, phenotypic variation explained by SNPs inside and flanking the genes of module two is depicted by red vertical dotted line.

SUPPORTING INFORMATION

Supplementary Figure 1. Plot of different number of subpopulations (K) from 1 to 20 versus cross validation error (CV error). The cross validation procedure as implemented in ADMIXTURE was used to initially find the optimum number of subpopulations (K) by minimizing the cross validation error. Multiple runs of ADMIXTURE were conducted at different values of K ranging from 1 to 20.

Supplementary Figure 2. Population structure within the diversity panel assessed using principal component analysis (a) and ADMIXTURE (b). (a) PC1 and PC2 explained 6% and 5% variance respectively. (b) Each vertical bar corresponds to an inbred line, where lines are arranged horizontally according to *a priori* information on subpopulation. Vertical bars are colored according to subpopulation membership.

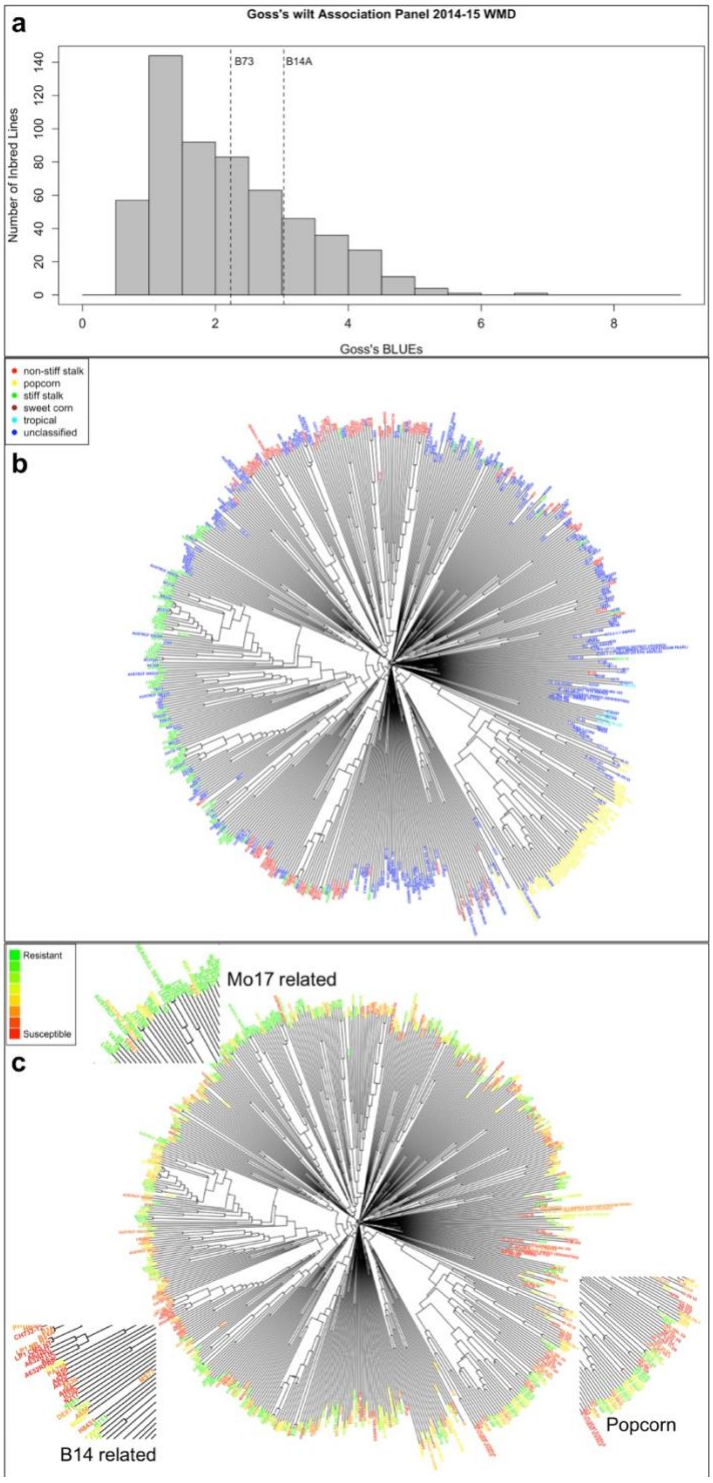
Supplementary Figure 3. Haplotype block analysis showing regions with significant SNPs on Chromosome 1, 2 and 5. Haplotype block analysis was conducted on each genomic region with significant SNPs to assess local LD decay rate. Haplotype analysis was performed with software Haploview. Standard color scheme was used in which bright red color indicates D' =1 and LOD $\geq$ 2, blue color indicates D' =1 and LOD < 2, white indicates D' <1 and LOD < 2, and shades of pink/red indicates D' <1 and LOD $\geq$ 2. D prime values are shown in boxes, empty boxes indicate D prime equals to 1. (a) Chromosome 1 region, had seven haplotype blocks with block 2 having maximum length of 4 kilo base pairs (kb). (b) Largest haplotype blocks were found on chromosome 2 region with maximum length of 79 kb. (c) Chromosome 5 region had 2 haplotype blocks with largest block length of 12 kb.

Supplementary Table 1. Treatments applied to the samples for the RNA-seq experiment.

Supplementary Table 2. Significant Gene Ontology terms found using AgriGO server for the genes of second module.

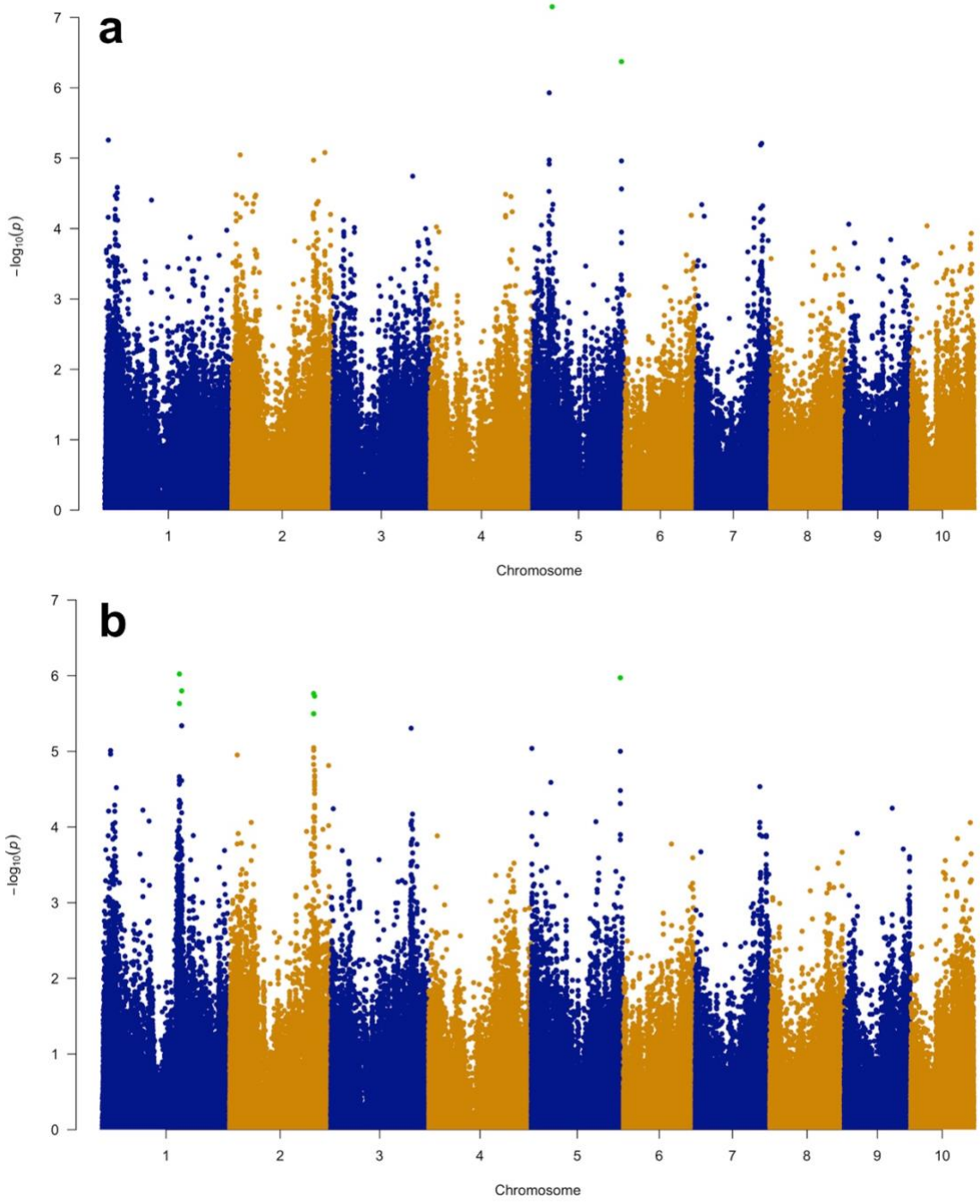
Supplementary Table 3. Gene annotations for genes of second module. All gene annotations for Zea mays were downloaded from gramene.org (Gramene FTP site) and compared to the list of genes of second module.

1034 Figure 1.



1035

1036 Figure 2.

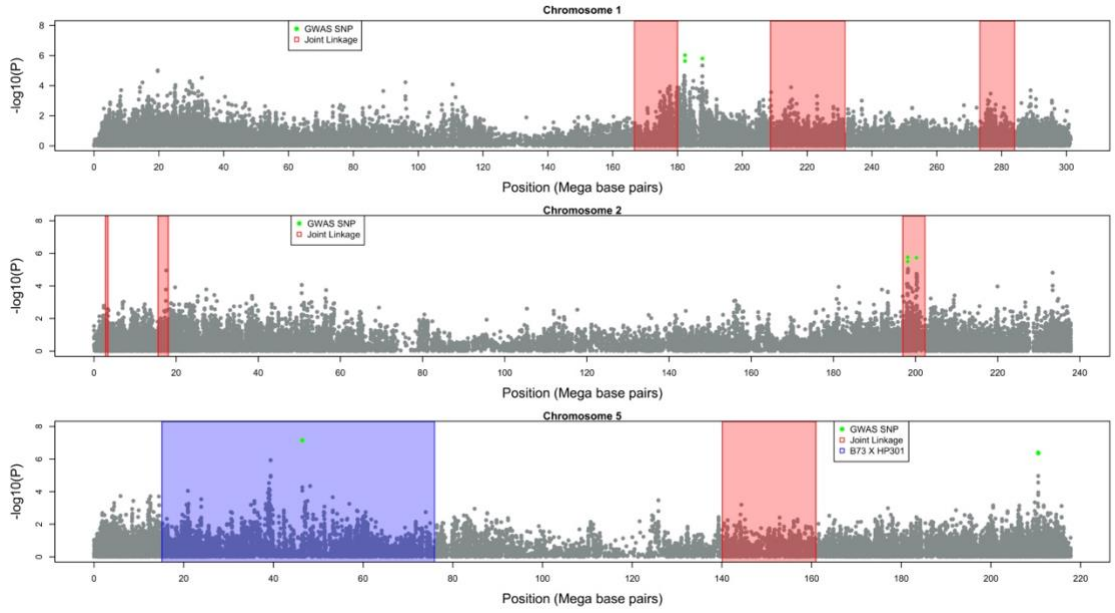


1037

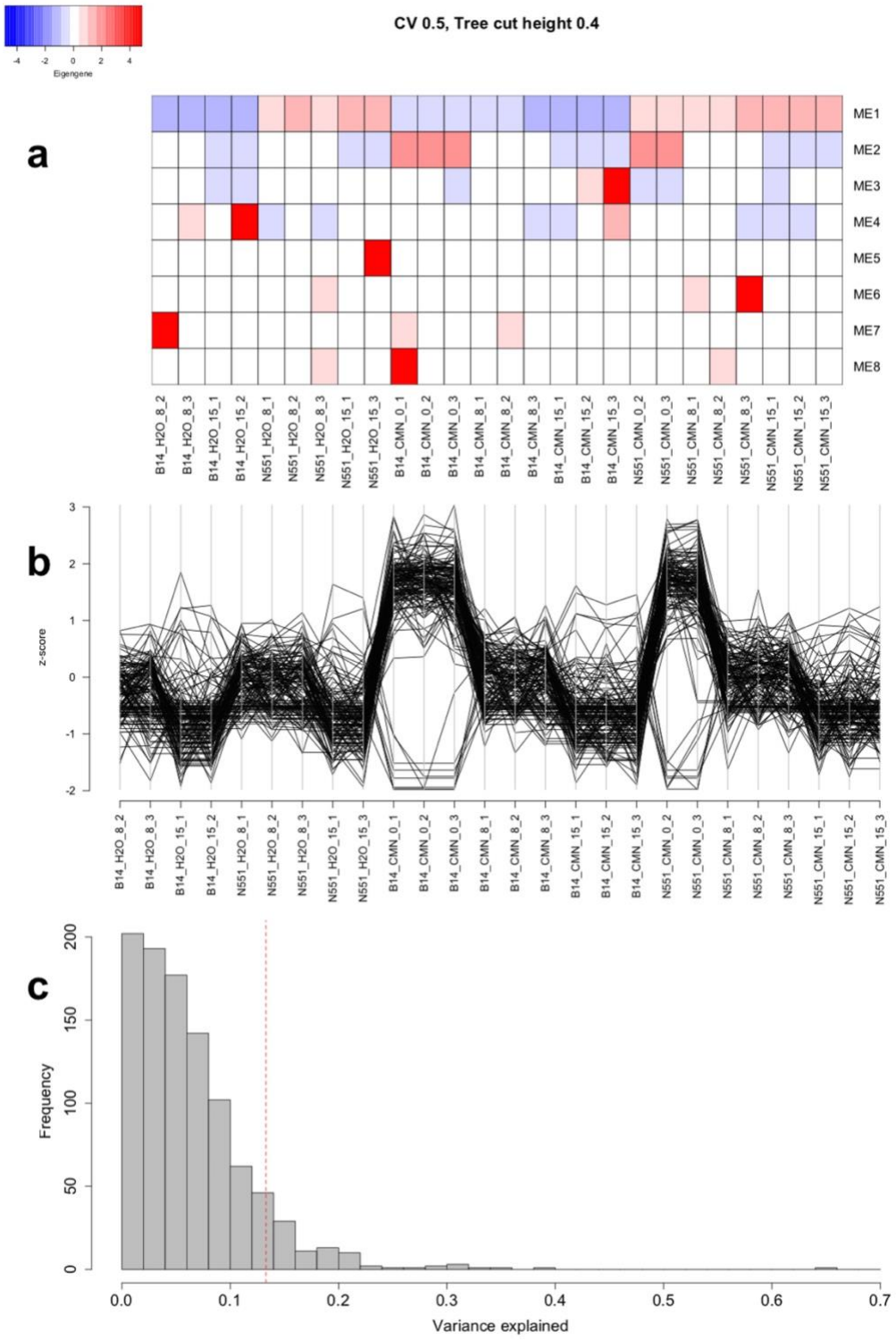
1038

1039

Figure 3.

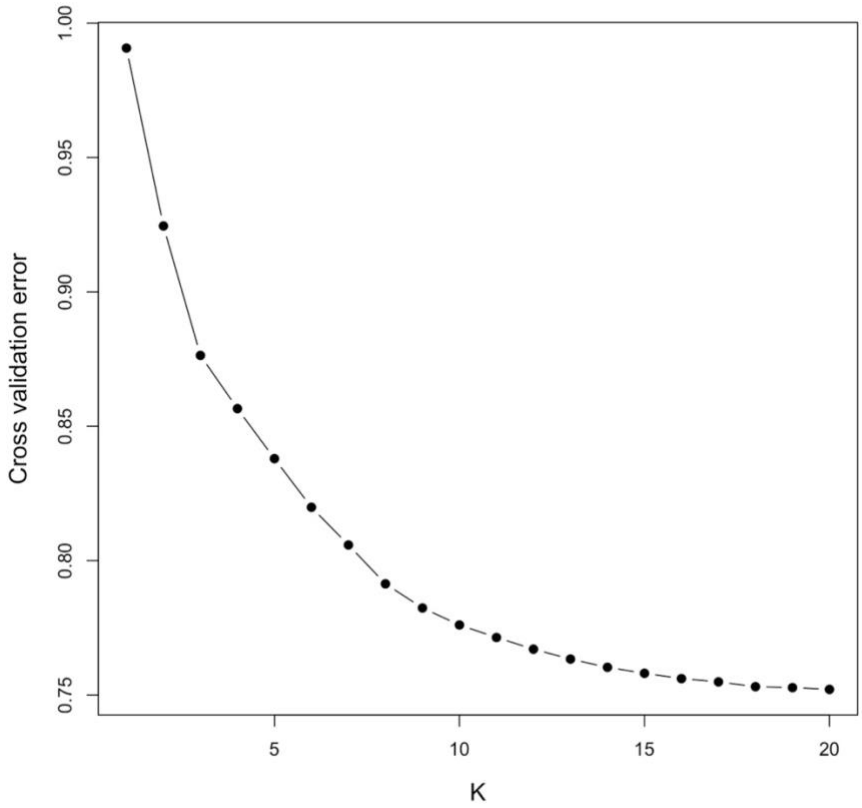

Chromosome 1
GWAS SNP
Joint Linkage
-log10(P)
Position (Mega base pairs)
Chromosome 2
GWAS SNP
Joint Linkage
-log10(P)
Position (Mega base pairs)
Chromosome 5
GWAS SNP
Joint Linkage
B73 X HP301
-log10(P)
Position (Mega base pairs)

1054 Figure 4.



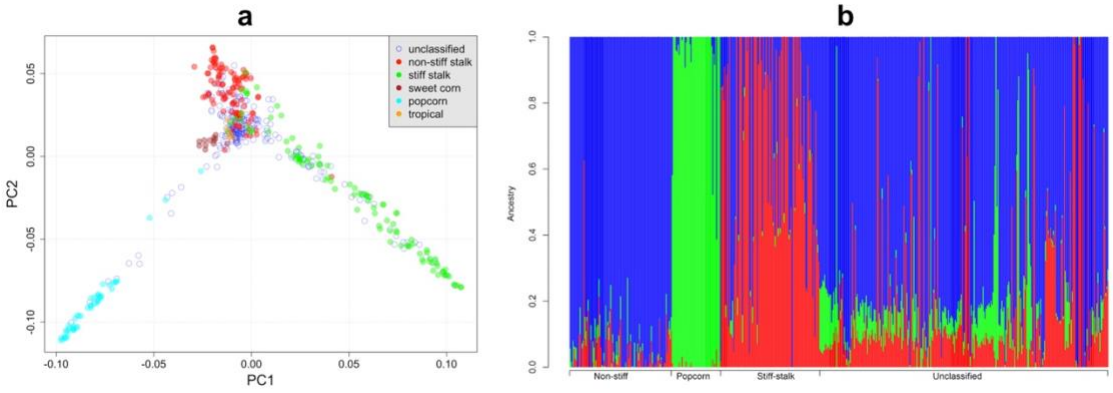
1055

1056 Supplementary Figure 1.



1057

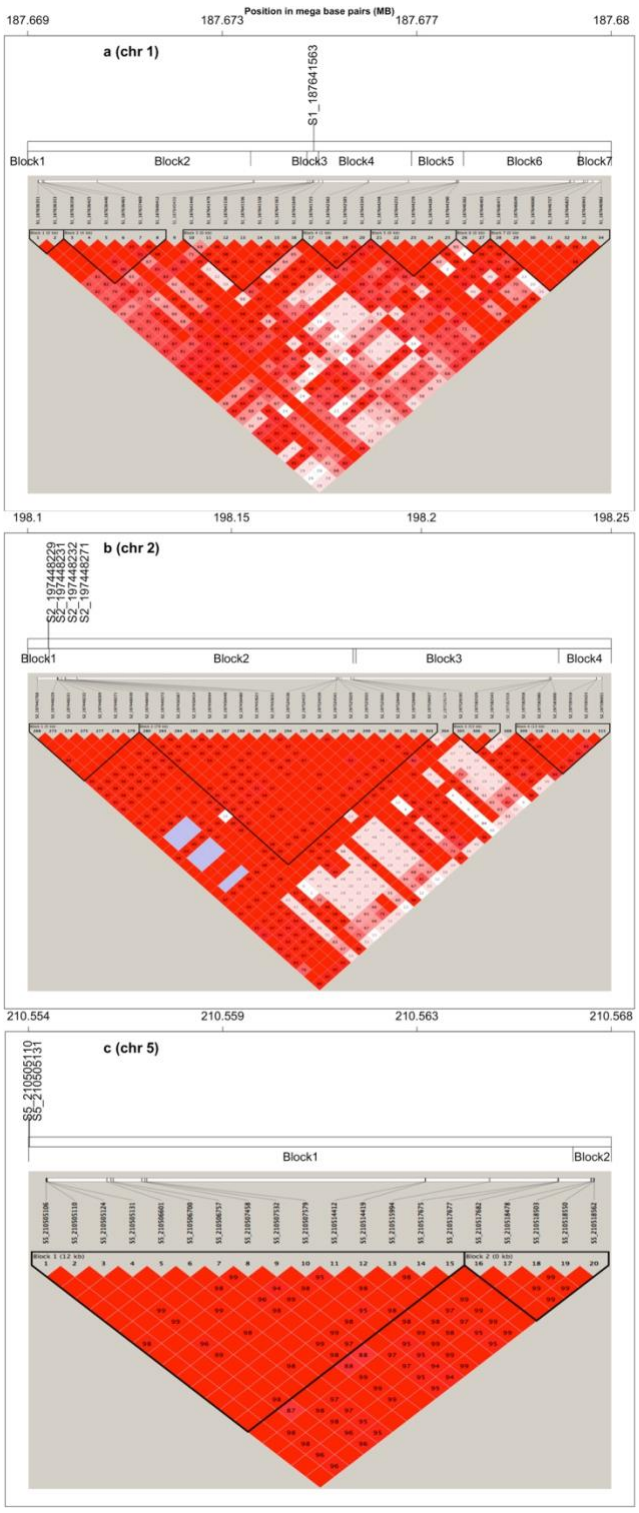
1058 Supplementary Figure 2.



1059

1060

Supplementary Figure 3.



TABLES

Table 1. Information about significant SNPs associated with Goss's wilt resistance in the diversity panel and combined dataset.

Chrom†	Physical Positon of SNP (bp)‡	*P*-value§	q-value¶	R^2#	SNP effect¥	Candidate Gene£	Gene Function‡‡
Diversity panel							
5	46,455,199	7.01×10^{-8}	0.02	0.05	0.28	GRMZM2G057459	Glutamate receptor
5	210,554,445	4.25×10^{-7}	0.10	0.03	0.26	GRMZM2G368206	PHD finger domain Zinc ion binding
5	210,554,466	4.25×10^{-7}	0.10	-	0.26		
Combined dataset (diversity panel and bi-parental families)							
1	182,307,976	9.49×10^{-7}	0.09	0.02	0.23	NA	
1	182,307,992	2.34×10^{-6}	0.08	0.02	0.22	NA	
1	187,675,076	1.59×10^{-6}	0.09	0.01	0.20	GRMZM2G132704	Nucleotide/RNA binding
						GRMZM2G132607	Ribokinase activity
						GRMZM2G132623	Constituent of ribosome
2	198,101,869	1.72×10^{-6}	0.09	0.01	0.20	GRMZM2G048582	Response to Nitrogen
						GRMZM2G048551	Zn ion binding
						GRMZM2G512469	Unknown
2	198,101,827	3.18×10^{-6}	0.09	0.01	0.20		
2	198,101,829	3.18×10^{-6}	0.09	-	0.20		
2	198,101,830	3.18×10^{-6}	0.09	-	0.20		
2	200,227,875	1.86×10^{-6}	0.09	0.01	0.21		
5	210,554,445	1.07×10^{-6}	0.09	0.03	0.19	GRMZM2G368206	Protein binding, zinc ion binding
5	210,554,466	1.07×10^{-6}	0.09	-	0.19		

† Chromosome

‡ Physical position of the SNPs in base pairs

§ P-value of the SNPs associated with Goss's wilt from GWAS

¶ False discovery rate or q-value of the SNPs, calculated from p-value

Variance explained by each SNP (R^2)

¥ Additive effect of the SNP from GWAS

£ Potential candidate genes in the region of significant SNPs

‡‡ Annotated function of the potential candidate gene

Table 2. Haplotype allele frequency in the diversity panel as a whole, and within individual subpopulations (stiff stalk, non-stiff stalk, popcorn, and unclassified). Haplotype blocks 1-4 are those at the chromosome 2 QTL region detected in the combined analysis of the diversity panel and RILs as shown in supplementary figure 2. Mean weighted mean disease (WMD) of the lines carrying each haplotype allele are presented.

Block†	**Haplotype Allele‡**	**Allele No.§**	**Haplotype Allele Frequency¶**					**Mean WMD¥**
			Panel	Stiff	Non-stiff	Popcorn	Unclassified	
BLOCK1 (5 kb)	ATGG	A1	0.65	0.66	0.80	0.02	0.72	2.38
	AGGT	A2	0.17	0.05	0.01	0.94	0.10	2.73
	CGAG	A3	0.18	0.29	0.20	0.04	0.18	2.89
BLOCK2 (79 kb)	CCTCAGAAACCACGGCGGA	B1	0.38	0.47	0.44	0.00	0.47	2.35
	TTCTGGGATATAAAGCCTA	B2	0.09	0.01	0.01	0.47	0.07	2.85
	TTCTGAGATCCGCGCCGTA	B3	0.16	0.28	0.18	0.04	0.17	2.92
	CCTCAGATACCACGGCGTC	B4	0.11	0.06	0.22	0.00	0.14	2.47
	CCTCAGATACCACGGTGTC	B5	0.10	0.12	0.16	0.00	0.10	2.28
	TTTCGGGATCCGCGGCGTA	B6	0.08	0.05	0.00	0.49	0.06	2.57
BLOCK3 (53 kb)	GCC	C1	0.46	0.51	0.30	0.48	0.50	2.44
	ACC	C2	0.18	0.28	0.18	0.04	0.17	2.97
	GTT	C3	0.36	0.21	0.52	0.48	0.33	2.44
BLOCK4 (13 kb)	CGGACG	D1	0.57	0.81	0.53	0.04	0.63	2.53
	TATACG	D2	0.11	0.08	0.14	0.11	0.10	2.64
	TATACC	D3	0.23	0.04	0.31	0.37	0.22	2.56
	TATGGG	D4	0.10	0.07	0.03	0.48	0.05	2.50

† Haplotype block defined using four gamete rule implemented in software Haploview

‡ An allele of a multiallelic haplotype

§ Designation for multiple alleles of a haplotype

¶ Frequency of each allele of a haplotype in the whole panel, stiff stalk, non-stiff stalk, popcorn, and unclassified sub-populations of the diversity panel

¥ Mean WMD of the lines carrying each haplotype alleles

SUPPLEMENTARY TABLES

Supplementary Table 1.

Line	Treatment	Hours post inoculation	Included/not included in WGCNA
B14A	Water	8	No
B14A	Water	8	Yes
B14A	Water	8	Yes
B14A	Water	15	Yes
B14A	Water	15	Yes
B14A	Water	15	No
B14A	CMN	0	Yes
B14A	CMN	0	Yes
B14A	CMN	0	Yes
B14A	CMN	8	Yes
B14A	CMN	8	Yes
B14A	CMN	8	Yes
B14A	CMN	15	Yes
B14A	CMN	15	Yes
B14A	CMN	15	Yes
N551	Water	8	Yes
N551	Water	8	Yes
N551	Water	8	Yes
N551	Water	15	Yes
N551	Water	15	No
N551	Water	15	Yes
N551	CMN	0	No
N551	CMN	0	Yes
N551	CMN	0	Yes
N551	CMN	8	Yes
N551	CMN	8	Yes
N551	CMN	8	Yes
N551	CMN	15	Yes
N551	CMN	15	Yes
N551	CMN	15	Yes

1095 Supplementary Table 2.

GO term	Description	p-value	FDR
GO:0032774	RNA biosynthetic process	0.00032	0.012
GO:0051252	regulation of RNA metabolic process	0.00022	0.012
GO:0006355	regulation of transcription, DNA-dependent	0.00021	0.012
GO:0006351	transcription, DNA-dependent	0.00032	0.012
GO:0031326	regulation of cellular biosynthetic process	0.0017	0.024
GO:0045449	regulation of transcription	0.0015	0.024
GO:0019219	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0.0016	0.024
GO:0010468	regulation of gene expression	0.0018	0.024
GO:0009889	regulation of biosynthetic process	0.0017	0.024
GO:0051171	regulation of nitrogen compound metabolic process	0.0017	0.024
GO:0010556	regulation of macromolecule biosynthetic process	0.0017	0.024
GO:0031323	regulation of cellular metabolic process	0.002	0.025
GO:0016070	RNA metabolic process	0.0022	0.025
GO:0080090	regulation of primary metabolic process	0.0025	0.026
GO:0060255	regulation of macromolecule metabolic process	0.0027	0.027
GO:0019222	regulation of metabolic process	0.003	0.028
GO:0006350	transcription	0.0032	0.028
GO:0006807	nitrogen compound metabolic process	0.0053	0.045
GO:0003700	transcription factor activity	3.80E-05	0.0016
GO:0030528	transcription regulator activity	0.00099	0.02

1096

1097

1098

1099

1100

1101

1102 Supplementary Table 3.

Gene name	Description
GRMZM2G368698	LOC100286362
GRMZM5G853245	nucleic acid binding protein
GRMZM5G883417	acyl-desaturase
GRMZM2G109201	LOC100281158
GRMZM5G892774	hypothetical protein LOC100384339
GRMZM2G170969	protease inhibitor/seed storage/LTP family protein
GRMZM2G122846	G-box-binding factor 1 (LOC100284009), mRNA
GRMZM2G038152	hypothetical protein LOC100384701
GRMZM2G305446	Aquaporin TIP3-1 (Tonoplast intrinsic protein 3-1)(ZmTIP3-1)(ZmTIP3;1)
GRMZM2G380377	DRE-binding protein 4
GRMZM2G103013	hypothetical protein LOC100277850
AC217293.3_FG007	hypothetical protein LOC100304375
GRMZM2G419994	wound induced protein
GRMZM2G133331	Putative uncharacterized protein
GRMZM2G069082	hypothetical protein LOC100272507
GRMZM2G474755	polcalcin Jun o 2
GRMZM2G156365	hypothetical protein LOC100280212
GRMZM2G042756	dehydration-responsive element-binding protein 1D
GRMZM2G389070	hypothetical protein LOC100277691
GRMZM2G390436	RING-H2 finger protein ATL5A
GRMZM2G006964	gibberellin 2-beta-dioxygenase 7
GRMZM2G098696	heat shock factor protein 7
GRMZM2G149024	transferase, transferring glycosyl groups
AF546187.1_FG011	Cytochrome P450 CYP81N4
GRMZM2G080191	hypothetical protein LOC100383760
GRMZM2G146163	chaperone protein dnaJ 20
GRMZM2G105364	GAST1 protein
GRMZM2G026910	LOC100284567
AF546188.1_FG005	Zein-alpha A20 Precursor (19 kDa zein A20)
GRMZM2G355572	LOC100280483
GRMZM2G129399	copper ion binding protein
GRMZM2G020054	hypothetical protein LOC100278463 (LOC100278463), mRNA
GRMZM2G092813	acyl-activating enzyme 11

GRMZM2G430646	hypothetical protein LOC100193824
GRMZM2G046618	Putative uncharacterized protein
GRMZM2G096412	hypothetical protein LOC100193798
GRMZM2G143747	hypothetical protein LOC100381926
GRMZM2G172450	hypothetical protein LOC100193017
GRMZM2G116603	F-box domain containing protein
GRMZM2G069126	sbCBF6
GRMZM2G074097	Thiazole biosynthetic enzyme 1-2, chloroplastic Precursor
GRMZM2G148908	hypothetical protein LOC100217218
GRMZM2G086123	hypothetical protein LOC100191796
GRMZM2G431030	hypothetical protein LOC100304009
GRMZM2G059964	cortical cell-delineating protein
GRMZM2G179049	NAC domain-containing protein 18
GRMZM2G382345	protein binding protein
GRMZM2G425728	early light-inducible protein ELIP (LOC541918), mRNA
GRMZM2G010474	hypothetical protein LOC100381709
GRMZM2G100543	hypothetical protein LOC100384219
GRMZM2G322917	hypothetical protein LOC100277625