ABSTRACT

ZILA, CHARLES THOMAS. Traditional and Genomic Methods for Improving Fusarium Ear Rot Resistance in Maize. (Under the direction of James B. Holland).

Fusarium ear rot of maize is a problem in maize-growing regions worldwide, impacting both food availability and quality. The deployment of maize hybrids possessing genetic resistance to the disease is an effective strategy for controlling Fusarium ear rot and reducing the prevalence of mycotoxins in livestock feed and human foodstuffs. However, the highly polygenic nature of resistance, in combination with a large environmental component, has made it a challenge to improve resistance levels of adapted maize breeding pools.

A key step in improving Fusarium ear rot resistance is the demonstration that resistance alleles from unadapted donor sources can be introgressed into adapted germplasm and confer resistance without having an undue effect on agronomic performance. In this study, I evaluated a set of $BC_4F_{3:4}$ and $BC_4F_{4:5}$ lines derived from backcrosses of the highly resistant donor source GE440 into the elite propriety stiff stalk line FR1064. These lines were evaluated for disease resistance across three years, and topcrosses to NC478 and LH283×LH287 were evaluated for both disease resistance and agronomic potential for three years. Two inbred lines, coded as NC301 and NC303, were identified as having superior disease resistance compared to FR1064 but are not statistically different for many agronomic traits including yield. Based on the improved disease resistance and favorable agronomic characters of NC301 and NC303, the USDA-ARS Plant Science Research Unit and North Carolina State University propose to release these two lines given their usefulness for improving disease resistance in elite temperate breeding pools.

Identifying and utilizing novel resistance allele variants is critical in making continued genetic gains in Fusarium ear rot resistance. Traditional QTL mapping approaches have had limited success in identifying useful allele variants as most Fusarium ear rot QTL have small effects and are not consistent between populations. Alternatively, genome-wide association studies (GWAS) offer finer-scale mapping combined with the ability to mine diverse breeding material for novel resistance alleles. I conducted a GWAS of Fusarium ear rot resistance on a commonly-used maize core diversity panel of 279 inbred lines using phenotypic data collected from North Carolina and Galicia, Spain. Although no significantly associated markers were identified in Galicia, three markers significantly associated with improved resistance were identified in the North Carolina data set, each with a relatively small additive effect. Targeted allele selection for these variants may be useful for some improvement of resistance levels in breeding programs. In addition, a large amount of additive background polygenic variance was observed in the analysis, indicating the potential usefulness of genomic selection in follow-up studies.

Within the last year, a tremendous amount of genotypic data has become available on the USDA-ARS national maize inbred seed bank, consisting of over 2,800 diverse inbred lines from public and private breeding programs worldwide. I used a set of 200,978 SNP markers to conduct a GWAS of Fusarium ear rot resistance in a sample of 1,689 inbred lines from this collection. Seven SNPs located on three chromosomes were identified as significantly associated with ear rot resistance. The three most significant SNPs were then included as fixed linear effects in a genomic best linear unbiased prediction (G-BLUP) model and compared to a traditional G-BLUP model. Although the three SNPs had small effects, their inclusion in the genomic prediction model significantly improved prediction accuracies

from 11% to 15%. These results suggest a relatively simple and straight-forward method for improving genomic selection models by accounting for known GWAS hits in addition to background polygenic variation.

Traditional and Genomic Methods for Improving Fusarium Ear Rot Resistance in Maize

by
Charles Thomas Zila, III

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Crop Science

Raleigh, North Carolina

2014

APPROVED BY:

_____	_____
James B. Holland Committee Chair	Peter J. Balint-Kurti
_____	_____
Gary A. Payne	Ross W. Whetten

DEDICATION

To my mom and dad, for providing encouragement and support of my education for almost three decades. Who would've thought that 24 years after Country Woods preschool I'd end up with a doctorate degree in North Carolina? Most importantly, to my amazing wife Katie: from Indiana to Raleigh (and back!), and every little detour in between, you've been there every step of the way. I can't imagine sharing this journey with anyone else.

BIOGRAPHY

Charlie Zila was born in 1986 to Chuck Zila, a farmer, and Jan Zila, a nurse, in La Porte, Indiana. Along with his younger brother Jim, Charlie was raised on a small corn and soybean farm in northwest Indiana. While in high school, he would meet his future wife, Katie. After graduating from high school in 2004, he attended Purdue University to complete a Bachelor of Science degree in Plant Breeding & Plant Genetics. During his time at Purdue, Charlie gained valuable plant breeding experience under the tutelage of small grains breeder Dr. Herb Ohm. He also gained industry breeding experience by working at Monsanto research stations in Lebanon, IN and Oxford, IN in the summers of 2007 and 2008, respectively.

Charlie and then-fiancée Katie moved to North Carolina in December of 2008. While Katie began her career as a 5^{th} grade elementary teacher, Charlie began a Master's degree with Dr. Jim Holland in corn genetics. He would complete his Master's degree in May 2011 and would continue on in the Holland Lab for his doctoral dissertation work. Continuing and expanding upon a breeding project from his Master's thesis, Charlie would become the local Fusarium ear rot expert, dealing with traditional breeding, association analyses, and genomic selection strategies to combat the disease.

Charlie finally married his now-wife in November 2011 after nearly nine years of dating. After defending his doctoral dissertation, they will relocate back to Indiana so Charlie can begin a soybean breeding program at the Pioneer station located in Windfall, IN. Once there, Katie will begin the arduous task of ensuring Charlie doesn't spend his entire salary on video games, trading card games, and Star Wars action figures.

ACKNOWLEDGMENTS

There are a countless number of people I can credit who have influenced my professional and personal life over the years. First, I want to thank my mom and dad for nurturing my curiosity and love of science and agriculture since I was a child. I also thank my brother Jim and my best friend Joe Pavich; you guys have been there for me since we were kids, and I can't think of any better partners in crime.

At Purdue, it should go without saying that sincere thanks goes to Dr. Herb Ohm. You fostered my interest in plant breeding, which would lay the foundation for making many other great professional relationships.

Here in North Carolina, greatest thanks go to my advisor, Dr. Jim Holland. Your knowledge and patience over the last five years made graduate school into a far less painful experience than what I had anticipated. Thanks also go to my other doctoral committee members, Drs. Gary Payne, Peter Balint-Kurti, and Ross Whetten. Many, many members of the Holland Lab have contributed help or insight to this dissertation over the last two and a half years, including, but not limited to: Drs. Shilpa Sood, Funda Ogut, Bode Olukolu, and Jason Cook; technicians Jason Brewer and Josie Bloom; temporary employees Steve Pigozzo, Brittany Scott, Colt Jackson, Carlos Vega, and Sarah Davidson-Dyer; and graduate students David Horne, Kristen Kump, Yang Bian, and Shang Xue. Visiting doctoral student Fernando Samayoa from the CSIC (Spain) made possible a great collaboration in this dissertation.

I also want to thank the many fellow graduate students both past and present that have been great friends to me while here at NCSU. Among them: Jill Recker, Oliver Ott, Dave

Eickholt, Stine Petersen, Justin Ma, Emma Flemmig, and Dr. Jose Santa Cruz. Above all else, I want to single out Dr. Jake Delheimer. Hanging out with you on the weekends made the rest of the week bearable.

Last, but greatest of all, I want to thank my beautiful, intelligent, and sometimes unbearably witty/sarcastic wife Katie. Your support for the last ten years is beyond words, and I can't wait to start the next chapter of our life together.

TABLE OF CONTENTS

LIST OF TABLES ix
LIST OF FIGURES xi
CHAPTER 1 : Literature Review 1
Understanding the Maize-*Fusarium verticillioides* Pathosystem and Breeding for Resistance 1
QTL and Association Mapping Studies in Maize 7
Genomic Selection Strategies in Maize 15
References 20
CHAPTER 2 : Proposal to Release Inbred Maize Lines NC301 and NC303 30
Abstract 31
Introduction 31
Materials and Methods 33
Breeding cross and line development 33
Field evaluation of inbred GEFR lines – experimental design 34
Field evaluation of topcross GEFR hybrids 35
Inoculation technique 37
Phenotypic trait measurements 38
Statistical methods 39
Characteristics 41
Discussion 43
References 45
CHAPTER 3 : A Genome-Wide Association Study Reveals Genes Associated with Fusarium Ear Rot Resistance in a Maize Core Diversity Panel 50
Abstract 53
Introduction 54
Materials & Methods 57
Genotypes and experimental design 57
Genotypic data 59
Statistical analyses 59
Estimation of least square means and heritabilities 59
Association analyses 62

Allele frequency analysis 63
Results 64
Line means and heritability 64
Association mapping of Fusarium ear rot resistance 65
Candidate genes colocalized with associated SNPs 66
Allele frequencies at candidate genes 67
Discussion 68
Heritability and false discovery rate estimation 68
Association analyses 71
Candidate genes 74
References 77
CHAPTER 4 : Integration of Genome-Wide Association and Genomic Selection Methods for Genetic Prediction of a Complex Disease Resistance Trait in Maize 90
Abstract 92
Introduction 93
Materials & Methods 97
Germplasm and experimental design 97
Inoculation and phenotyping methods 99
Genotypic data 100
Statistical analyses 101
Estimation of least square means 101
Association analyses 105
Cross validation of genome-wide prediction models 107
Results 109
Line means and heritability 109
Genome-wide association mapping of Fusarium ear rot resistance 110
Genes colocalized with associated SNPs 111
Comparison of genome-wide prediction models 112
Discussion 113
Heritability and genotypic correlation between experiments 113
Association mapping 114
Cross validation of genome-wide prediction models 118

References 121
APPENDICES 137
APPENDIX A : Supplemental Material for Chapter 3 138
APPENDIX B : Supplemental Material for Chapter 4 146
APPENDIX C : Protocols for Preparing Inoculum for Fusarium Ear Rot Experiments 154
Protocol for Preparing *Fusarium verticillioides* Cultures 154
Starter plates 154
Inoculation plates 156
Protocol for Preparing a Live Suspension of *Fusarium verticillioides* Inoculum 156
Preparing the conidia suspension 156
Standardization of conidia suspension 157

LIST OF TABLES

Table 2.1. Least square means for the twenty superior GEFR lines and checks for Fusarium ear rot resistance and fumonisin content as inbreds *per se* (averaged across ten environments) and for Fusarium ear rot resistance, fumonisin content, grain yield, and other agronomic traits as topcrosses to the tester NC478 (averaged across nine environments)..... 47

Table 2.2. Least square means for the twenty superior GEFR lines and checks for Fusarium ear rot resistance, fumonisin content, grain yield, and other agronomic traits as topcrosses to the tester LH283×LH287 (averaged across two environments for disease traits and three environments for agronomic traits)..... 48

Table 3.1. Genotypic covariance/variance/correlation matrix for Fusarium ear rot from the combined analysis of a maize diversity panel evaluated in five environments. The diagonal (bold) is an estimate of genetic variance ($\sigma G2$) plus the genotype by environment interaction ($\sigma GE2$) within each environment. Estimates of genetic variance (covariance between pairs of environments) are shown below the diagonal, and genetic correlations between inbred lines in each pair of environments are shown above the diagonal. 82

Table 3.2. Number of lines, number of groups, compression level, polygenic additive background genetic variance component, residual genotypic variance component, and proportion of total line mean variance explained by additive relationship matrix from the three mixed-linear model (MLM) analyses. 83

Table 3.3. Chromosome locations (AGP v2 coordinates), allele effect estimates, genes containing or adjacent to SNP, and other summary statistics for the three SNPs significantly associated with Fusarium ear rot resistance in the North Carolina analysis and the single SNP associated with resistance in the combined analysis. Statistics from environments in which the SNPs were not significantly associated with ear rot are also shown for comparison..... 84

Table 3.4. Allele frequencies of significantly associated SNPs in the five major maize subpopulations. 85

Table 4.1. Sample size (N), mean ear rot severity, genotypic variance component estimates ($\sigma G2$), average prediction error variance ($\sigma PPE2$) and heritability (HC) estimates for Fusarium ear rot resistance in the full inbred association panel, filtered association panel, across the topcross experiment, and within the B47 and PHZ51 topcrosses, respectively. . 126

Table 4.2. Number of lines, number of groups and compression level of the full 2480×2480 kinship matrix, and proportion of total line mean variance explained by additive relationship matrix from the four mixed-linear model (MLM) analyses..... 127

Table 4.3. Chromosome locations (AGP v2 coordinates), allele effect estimates, genes containing SNP, and other summary statistics for the seven SNPs significantly associated with Fusarium ear rot resistance from the two inbred association panel analyses. 128

Table 4.4. Allele frequencies of significantly associated SNPs in the five major maize heterotic groups..... 129

Table 4.5. Prediction accuracies (R^2) and regression coefficient estimates for the GBLUP and GBLUP + GWAS five-fold cross validation models. Values are reported as averages across the five-folds within each replicate. 130

Table A.1. Heritability estimates for Fusarium ear rot resistance, mean ear rot severity, heritability estimates for silking date, regression coefficients for silking date covariates, and significance level of regression coefficients. Estimates are reported for each environment individually, across years within the North Carolina and Galicia environments, and combined across all environments. 138

Table A.2. Climate data for the three North Carolina and two Galicia environments. Average daily minimum temperature, average daily maximum temperature, average daily overall temperature, and cumulative precipitation level are reported for two time intervals in each environment: planting date to the average silking date (date at which at least 50% of the plots within an environment had silked) and average silking date to 45 days post-silking..... 139

LIST OF FIGURES

Figure 2.1. Flowchart showing the screening process for the GEFR backcross lines. The inbred disease evaluation pipeline is shown on the left, the hybrid disease evaluation pipeline in the middle, and the hybrid yield evaluation pipeline on the right. Location abbreviations are as follows: Clayton = CLY; Kinston = KIN; Lewiston-Woodville = LEW; and Jackson Springs = SDH. 49

Figure 3.1. Results of the three GWAS showing significant associations (points above red FDR = 0.10 threshold lines) in the North Carolina (A), Galicia (B), and combined (C) analyses. The vertical axis indicates $-\log_{10}$ of *P*-value scores, and the horizontal axis indicates chromosomes and physical positions of SNPs. 86

Figure 3.2. LD heatmaps showing LD measure (r^2) calculated for each pairwise combination of SNPs in an approximately ±1 Mbp region surrounding each SNP significantly associated with ear rot resistance in the North Carolina analysis. (A) LD around chromosome 1 SNP. (B) LD around chromosome 5 SNP. (C) LD around chromosome 9 SNP. The significant SNP on each chromosome is highlighted by the perpendicular black lines within each heatmap. Colors indicate the magnitude of each pairwise r^2 measure (r^2=1 is red to r^2=0 is white)..... 88

Figure 4.1. Genetic relationships between the 1689 lines of the full inbred association panel visualized using a principal component analysis of the **K** matrix. The horizontal and vertical axes are the first and second principal components, respectively. The color gradient from blue to red of the points represents the relative mean Fusarium ear rot score of each line (blue is most resistant and red is most susceptible). Five major recognized heterotic group clusters are labeled in large gray font, and the 26 nested association mapping (NAM) population founders and Mo17 are labeled in small black font for reference. 131

Figure 4.2. Manhattan plots showing significant associations (points above the red FDR = 0.20 threshold lines) from the full inbred association panel (A) and filtered inbred association panel (B) GWAS analyses. The vertical axis indicates $-\log_{10}$ of *P*-value scores, and the horizontal axis indicates chromosomes and physical position of SNPs. 133

Figure 4.3. Manhattan plots showing significant associations (points above the red FDR = 0.20 threshold lines) from the B47 topcross (A) and PHZ51 topcross (B) GWAS analyses. The vertical axis indicates $-\log_{10}$ of *P*-value scores, and the horizontal axis indicates chromosomes and physical position of SNPs. 134

Figure 4.4. LD heatmaps showing LD measure (r^2) calculated for each pair-wise combination of SNPs in an approximately ±0.5 Mbp region surrounding each SNP significantly associated with ear rot resistance in the two inbred association panel analyses. (A) LD around the four SNPs chromosome 4 SNPs located in the 7.6 Mbp to 9.4 Mbp interval. (B) LD around chromosome 4 SNP at physical position 124.9 Mbp. (C) LD around chromosome 5 SNP. (D) LD around chromosome 9 SNP. The significant SNP(s) on each chromosome is highlighted by the perpendicular black lines within each heatmap. Colors indicate the magnitude of each pair-wise r^2 measure (r^2=1 is red to r^2=0 is white). 135

Figure A.1. (A) Example of a susceptible (100% severity) phenotype. (B) Example of a resistant (0% severity) phenotype. The arrow indicates the point of inoculation in the resistant ear. 140

Figure A.2. Scatter plot matrix illustrating the genotypic relationship of Fusarium ear rot resistance between environments. The model used to estimate variance components and genetic correlations in the combined analysis was used to predict least square means for each inbred line within each environment (treating line as a fixed effect instead of random). Means for each line on the natural log transformed scale are plotted against one another in each pair-wise combination of environments. 141

Figure A.3. Estimating the false discovery rate (FDR) for SNP marker association with Fusarium ear rot resistance in the North Carolina analysis. (A) A density histogram showing *p*-value distribution of 47,445 SNPs following GWAS. (B) The *q*-values plotted against their respective *p*-values. (C) The number of SNPs plotted against each of the respective *q*-value estimates. (D) The expected number of false positive SNPs versus the total number of significant SNPs given the *q*-values. 142

Figure A.4. Estimating the false discovery rate (FDR) for SNP marker association with Fusarium ear rot resistance in the Galicia analysis. (A) A density histogram showing *p*-value distribution of 47,445 SNPs following GWAS. (B) The *q*-values plotted against their respective *p*-values. (C) The number of SNPs plotted against each of the respective *q*-value estimates. (D) The expected number of false positive SNPs versus the total number of significant SNPs given the *q*-values. 143

Figure A.5. Estimating the false discovery rate (FDR) for SNP marker association with Fusarium ear rot resistance in the combined analysis. (A) A density histogram showing *p*-value distribution of 47,445 SNPs following GWAS. (B) The *q*-values plotted against their respective *p*-values. (C) The number of SNPs plotted against each of the respective *q*-value estimates. (D) The expected number of false positive SNPs versus the total number of significant SNPs given the *q*-values. 144

Figure B.1. Diagram showing the relationships between the full inbred line data set and the subsets used for GWAS and genomic prediction. The total number of phenotypic observations collected within the three years of field experiments is shown in the top three circles (2010 in red, 2011 in blue, and 2012 in green). The number of observations from each year that is present in each data subset (GWAS Set 1, GWAS Set 2, and GS) after filtering is shown in the three funnels. The type of analysis conducted on each subset is shown in the bottom row of text boxes. 146

Figure B.2. Estimating the false discovery rate (FDR) for SNP marker association with Fusarium ear rot resistance in the full inbred association panel analysis. (A) A density histogram showing raw *P*-value distribution of 200,978 SNPs following GWAS. (B) The FDR-adjusted *P*-values plotted against their respective raw *P*-values. (C) The number of SNPs plotted against each of the respective FDR-adjusted *P*-value estimates. (D) The expected number of false positive SNPs versus the total number of significant SNPs given the FDR-adjusted *P*-values. 148

Figure B.3. Estimating the false discovery rate (FDR) for SNP marker association with Fusarium ear rot resistance in the filtered inbred association panel analysis. (A) A density histogram showing raw *P*-value distribution of 200,978 SNPs following GWAS. (B) The FDR-adjusted *P*-values plotted against their respective raw *P*-values. (C) The number of SNPs plotted against each of the respective FDR-adjusted *P*-value estimates. (D) The expected number of false positive SNPs versus the total number of significant SNPs given the FDR-adjusted *P*-values. 149

Figure B.4. Estimating the false discovery rate (FDR) for SNP marker association with Fusarium ear rot resistance in the B47 topcross analysis. (A) A density histogram showing raw *P*-value distribution of 200,978 SNPs following GWAS. (B) The FDR-adjusted *P*-values plotted against their respective raw *P*-values. (C) The number of SNPs plotted against each of the respective FDR-adjusted *P*-value estimates. (D) The expected number of false positive SNPs versus the total number of significant SNPs given the FDR-adjusted *P*-values. 150

Figure B.5. Estimating the false discovery rate (FDR) for SNP marker association with Fusarium ear rot resistance in the PHZ51 topcross analysis. (A) A density histogram showing raw *P*-value distribution of 200,978 SNPs following GWAS. (B) The FDR-adjusted *P*-values plotted against their respective raw *P*-values. (C) The number of SNPs plotted against each of the respective FDR-adjusted *P*-value estimates. (D) The expected number of false positive SNPs versus the total number of significant SNPs given the FDR-adjusted *P*-values. 151

Figure C.1. Example of a PDA plate completely covered by mycelia of *Fusarium verticillioides*..... 159

Figure C.2. Diagram of the grid area of a hemocytometer as viewed under a microscope. Conidia within the five grid areas highlighted in red are those that should be counted when making a stock solution of *F. verticillioides* inoculum..... 160

CHAPTER 1: Literature Review

Understanding the Maize-*Fusarium verticillioides* Pathosystem and Breeding for Resistance

Fusarium ear rot is the most common ear disease of maize (*Zea mays* L. ssp *mays*) in the United States and is endemic in all maize-producing countries (White, 1999; Mesterházy et al., 2012). The disease is predominately caused by the fungus *Fusarium verticillioides* (Sacc.) Nirenberg (formerly *F. moniliforme* Sheldon), but the related species *F. proliferatum* (Matsushima) Nirenberg and *F. subglutinans* (Wollenweb. & Reinking) Nelson, Tousoun, and Marasas are also capable of producing Fusarium ear rot symptoms (White, 1999). Fusarium ear rot symptoms are most severe in environments like the southern United States and lowland tropics that experience hot and dry weather during and after flowering (Miller, 1994; White, 1999). *F. verticillioides* is a heterothallic fungus (teleomorph *Giberella moniliformis* Wineland), requiring two mating types (*MAT-1* and *MAT-2*) for sexual reproduction (Kerényi et al., 2004). Conidia of *F. verticillioides* can overwinter in field debris, and the fungus may enter the ear through airborne conidia growing down the silk channel, through wounds caused by insects such as corn earworm (*Helicoverpa zea*), European corn borer (*Ostrinia nubilalis*), and thrips (*Frankliniella* spp.), or through systemic spread up the stalk into the ear shank and developing ear (Koehler, 1942; Miller, 1994; White, 1999; Parsons and Munkvold, 2010). Symptoms of Fusarium ear rot can present in many ways, including white or pink fungal growth on the ear, "starbursting" of infected kernels (white streaks radiating from the silk scar to the base of the kernel), and complete

destruction of the kernels in severe cases (White, 1999; Eller et al., 2008). Severe infections can result in yield losses due to decreased grain yield and grain test weight. In addition to the hemibiotrophic pathogen state of *F. verticillioides* that produces visible symptoms, the fungus commonly occurs in a symptomless endophytic state during most stages of the maize growth cycle (Bacon et al., 2008), and it is present as an endophyte in most maize seed lots (Warren and Kommedahl, 1973; Leslie et al., 1990; Yates et al., 1997). In symptomless plants, hyphae of *F. verticillioides* grow intercellularly and absorb nutrients in the apoplasm, causing minimum injury to host plants (Bacon et al., 2008). Disease occurs when a combination of abiotic and biotic stresses causes the fungus to adopt a pathogenic state in a susceptible host (Bacon and Hinton, 1996; Bacon et al., 2008). In the compatible pathogenic reaction, fungal growth occurs both intercelluarly and intracellularly (Bacon and Hinton, 1996), and dissolution of the walls of pericarp cells accompanies the development of observable disease symptoms (Duncan and Howard, 2010).

Taxonomic analyses across all known *Fusarium* species based on morphological characters and genetic analyses group the various species into one of seven clades (I-VII), and clades are subdivided into sections based on synapomorphic character states (Watanabe et al., 2011). *F. verticillioides, F. proliferatum,* and *F. subglutinans* all belong to the *Liseola* section of Clade V, where the closest related species are the *F. oxysporum* species complex in section *Elegans* of the same clade. By comparison, the *F. graminearum* species complex, another important group of pathogens of maize as well of wheat, belongs to section *Discolor* of Clade VII. Genome comparison of *F. verticillioides* to *F. oxysporum* and *F. graminearum* revealed that *F. verticillioides* shares 90% of its genome with *F. oxysporum* but only 79%

with *F. graminearum* (Ma et al., 2010). Phylogenetic analyses suggest that the *Liseola* section is actually comprised of as many as 50 species (referred to as the *Giberella fujikuroi* species complex), but rapid evolutionary divergence within this section complicates phylogenetic relationships among the species complex (Kvas et al., 2009; Watanabe et al., 2011). Currently, there is limited understanding of the diversity among *F. verticillioides* isolates in the United States and abroad. However, some studies demonstrate significant variation in both pathogenic aggressiveness and mycotoxin production of *F. verticillioides* isolates sampled within maize growing regions worldwide (Leslie et al., 1992; Danielsen and Jensen, 1998; Proctor et al., 2006).

Fusarium verticillioides is capable of producing the mycotoxin fumonisin. Fumonisin is predominately produced during the pathogenic phase of the fungus, but the mycotoxin has also been detected in endophytic associations with maize (Bacon et al., 2001, 2008). Although several different fumonisin structures have been identified, fumonisin B_1 is the predominant structure produced by *F. verticillioides* (ApSimon, 1994; Proctor et al., 2006). Fumonisins have structures similar to sphingolipids and have been demonstrated to disrupt sphingolipid metabolism in animal species (ApSimon, 1994; Riley et al., 1994). Consumption of maize grain contaminated by fumonisins has been associated with several different diseases in animals and humans, including (but not limited to): equine leukoencephalomalacia (Kellerman et al., 1990), porcine pulmonary edema (Harrison et al., 1990), liver cancer in rats (Gelderblom et al., 1988), neural tube defects in mice (Voss et al., 2006), human esophageal cancer (Rheeder et al., 1992), and growth retardation in human children (Kimanya et al., 2010; Shirima et al., 2013). Currently, the European Union has set

regulations limiting fumonisin contamination in foods for human consumption to 1 ppm (ppm = 1 $\mu g\ g^{-1}$) (Kyprianou, 2007), and the Food and Drug Administration recommends no more than between 2 and 4 ppm of fumonisin contamination in milled maize products for human consumption (Center for Food Safety and Applied Nutrition, 2001). However, fumonisin can often still be found in many milled and processed maize products in the United States and elsewhere (Sydenham et al., 1991; Escobar et al., 2013; Shephard et al., 2013). With the onset of global climate change, fumonisin contamination (in addition to contamination by other mycotoxins such as aflatoxin, deoxynivalenol, and ergot) in livestock feed and human foodstuffs is expected to be a growing problem in the coming decades in both industrial and subsistence societies (Dwivedi et al., 2013).

Although levels of Fusarium ear rot and fumonisin contamination can be reduced indirectly by controlling insect vectors through transgenic pest resistance (Munkvold et al., 1997, 1999; Pray et al., 2013), an effective and complementary method to control the disease is through the deployment of naturally occurring host genetic resistance. Fusarium ear rot resistance is under polygenic control and strongly influenced by environmental effects; no fully immune genotypes have been discovered (King and Scott, 1981; Nankam and Pataky, 1996; Clements et al., 2004). The complexity of resistance has impeded modern breeding efforts such that most commercial maize hybrids possess lower levels of resistance than desired (Bush et al., 2004). In spite of the complexity of genetic resistance and the low heritability of resistance measured on an individual plot basis, resistance on the basis of family means from well-replicated studies is moderately to highly heritable (Robertson et al., 2006; Eller et al., 2008; Bolduan et al., 2009; Löffler et al., 2011; Hung and Holland, 2012).

These same studies report high genetic correlations between Fusarium ear rot symptoms and fumonisin contamination, indicating that visual selection for ear rot resistance can be effective in reducing susceptibility to fumonisin contamination among both inbred populations as well as testcrosses (provided well-replicated studies are performed). Hung and Holland (2012) also reported a significant genetic correlation ($r \geq 0.78$) between inbred resistance *per se* and testcross resistance. The high genetic correlation, in combination with a lower estimate of genetic variance for resistance in testcrosses compared to inbreds, suggested that the most efficient way to improve hybrid ear rot resistance is to screen inbred parents for resistance before costly testcross evaluations. These conclusions are in contrast to those of Löffler et al. (2011), who placed more emphasis on screening for resistance in testcross generations instead of inbred parents. However, both studies suggest that the best strategy for improving the resistance level of commercial maize hybrids is some combination of inbred parent screening in the nursery in addition to testcross evaluations.

Although no fully immune genotypes have been identified, inbred lines expressing high levels of Fusarium ear rot resistance have been described (Clements et al., 2004; Eller et al., 2008). Unfortunately, most known resistant inbred lines come from unadapted or exotic germplasm pools, such as GE440 (Clements et al., 2004). Derived from the open-pollinated variety Hastings Prolific, GE440 possesses many undesirable agronomic characteristics, such as white kernel color, poor root lodging resistance, and poor yield. Robertson-Hoyt et al. (2007) examined the relationship between disease resistance and agronomic performance in a $BC_1F_{1:2}$ population (hereafter referred to as the "GEFR" population) using GE440 as the resistant donor parent and the susceptible proprietary line FR1064 as the recurrent parent.

The authors reported that disease resistance traits were not strongly correlated with agronomic traits, suggesting that it is possible to select for improved resistance without having an undue effect on agronomic performance. Additional backcrosses and improvement by visual selection were made in the GEFR population by Eller et al. (2010), and Zila (2011) identified superior $BC_4F_{3:5}$ and $BC_4F_{4:6}$ GEFR lines with improved resistance compared to FR1064 but with similar agronomic characteristics on both an inbred *per se* and testcross basis.

A primary objective of this doctoral dissertation is the continued evaluation of the GEFR population by augmenting experiments described in Zila (2011). Combined yield trial and disease trial data from Zila (2011) and from additional experiments carried out as part of this dissertation project were used to identify those GEFR line(s) consistently displaying superior Fusarium ear rot resistance combined with good agronomic performance. I recommend the release of at least two superior lines for use as an adapted source of Fusarium ear rot resistance in elite Stiff Stalk breeding populations.

Despite the fact that GE440 is very unadapted compared to elite temperate germplasm, it was chosen as the resistance donor for the GEFR population because similar levels of resistance have not been observed in more adapted germplasm (Clements et al., 2004). Modified backcrossing procedures using unadapted donor sources can take a large number of generations to recover lines with similar agronomics to the recurrent parent, and given the quantitative nature of resistance, recovered lines may possess a lower level of resistance than is observed in the donor parent (Eller et al., 2010; Zila, 2011). The identification of novel sources of resistance in more adapted germplasm will facilitate more

effective improvement of elite breeding pools for ear rot resistance. Significant genetic variation for resistance has been observed among the major subpopulations of maize (Santiago et al., 2013); however, the genetic basis for this variation within and among subpopulations is largely uncharacterized. Furthermore, the highly polygenic nature of resistance makes it difficult to precisely identify specific DNA sequence variations that confer or are associated with improved resistance.

QTL and Association Mapping Studies in Maize

The first major strategy to dissect the genetic architecture of Fusarium ear rot resistance was the use of biparental populations and linkage-based quantitative trait locus (QTL) mapping. To date, however, no large effect Fusarium resistance QTL have been identified through QTL mapping, and results are often not consistent across populations. Robertson-Hoyt (2006) identified six resistance QTL regions associated with 47% of the observed phenotypic variation in the GEFR population and five regions associated with 31% of the observed phenotypic variation in a NC300×B104 recombinant inbred line (RIL) population. Three loci with significant effects on resistance were common between the two populations, and the authors also reported some overlap with QTL described by Pérez-Brito et al. (2001). In a biparental population derived from the Chinese lines 87-1×Zong 3, Ding et al. (2008) identified six QTL regions associated with ear rot resistance within different testing environments, and none of them colocalized with regions identified by either Pérez-Brito et al. (2001) or Robertson-Hoyt et al. (2006). Furthermore, only two of the six QTL regions (both located in bin 3.04) were consistently identified across all testing environments,

with the largest QTL on average associated with 17% of the observed phenotypic variation across environments. This estimate of the effect of the QTL on variation is expected to be upwardly biased due to the size of the population sample used (Beavis, 1998; Melchinger et al., 1998). In a population derived from the resistant line BT-1 crossed to the susceptible line N6, Li et al. (2011) reported four QTL regions collectively associated with 22% of the observed phenotypic variation in resistance, and none colocalized with any other QTL region described in any of the aforementioned studies. A follow-up study by Chen et al. (2012) using a different mapping population with the same resistant parent (BT-1), but different susceptible parent (Xi502), reported three QTL associated with resistance, of which only one colocalized with a QTL region reported by Li et al. (2011).

Resolving small effect QTL to causal genes is exceedingly difficult in biparental populations (Holland, 2007). The recombination events captured in a RIL population are typically insufficient for fine mapping (Korte and Farlow, 2013). Association mapping offers higher mapping resolution and a greater sampling of alleles compared to QTL mapping by exploiting historical and evolutionary recombination events at the population level (Yu and Buckler, 2006). Two broad strategies exist for association mapping: candidate-gene association mapping, which limits the search space to predetermined candidate genes involved in pathways that may be suspected to control the trait of interest; and genome-wide association mapping (GWAS), which surveys the entire genome to detect novel marker-trait associations (Zhu et al., 2008). Thornsberry et al. (2001) first introduced association mapping to plant species by using the candidate-gene approach to explore the association between *Dwarf8* polymorphisms with flowering time and plant height in maize. The candidate-gene

approach was used for two reasons: first, the authors were interested specifically in the *Dwarf8* gene; secondly, and more importantly, limits of genomic technology at the time made it infeasible to genotype individuals in the association panel with enough genome-wide markers to conduct GWAS. Low linkage disequilibrium (LD) species, such as maize, require a large number of genetic markers from across the entire genome in order to detect marker-trait associations. Although candidate-gene association mapping is a reasonable option when the cost of genotyping is high, it requires extensive knowledge of the biochemical pathway(s) contributing to the trait of interest (Remington and Purugganan, 2003). For most complex traits, including Fusarium ear rot resistance, there exists little to no information about the pathways controlling those traits.

The availability of a reference genome sequence for a species allows for easier resequencing of individuals and the development of more cost effective genotyping platforms. The assembly of an annotated reference sequence for the *Arabidopsis thaliana* accession Columbia allowed for the resequencing of other Arabidopsis accessions and the development of an Affymetrix single nucleotide polymorphism (SNP) array containing over 250,000 markers (Arabidopsis Genome Initiative, 2000; Clark et al., 2007; Kim et al., 2007). Atwell et al. (2010) conducted the first GWAS in Arabidopsis that utilized the cost-effective genotyping of the Affymetrix SNP array. In spite of the limited size of the association panel in the study (107 lines), the GWAS detected many loci known *a priori* to be associated with flowering time, the hypersensitive response, and other phenotypes. Although candidate genes tended to be overrepresented by significant SNP hits, novel genic regions were also identified as associated with the various traits of interest. Given ample marker coverage, the results of

this study highlighted the power of GWAS in detecting regions of the genome not previously known to contain genes associated with traits of interest.

Within the cereal grains, the first reference sequence assembled was that of rice (*Oryza sativa* L.; International Rice Genome Sequencing Project, 2005). As a species, genome-wide rates of LD decay are much lower for rice ($r^2 \leq 0.28$ within 167 kbp; Huang et al., 2010) than compared to Arabidopsis ($r^2 \leq 0.20$ within 10 kbp; Kim et al., 2007). Huang et al. (2010) theorized that the large LD blocks in rice could be the result of its small effective population size and high rate of inbreeding. To further complicate matters, strong population structure is present in rice due to the isolation of the *O. sativa* L. ssp. *indica* and *O. sativa* L. ssp. *japonica* subspecies. Utilizing an approach first explored in human genetics (International HapMap Consortium, 2005), Huang et al. (2010) constructed a high-density rice haplotype map (HapMap) based on 517 landraces that identified nearly 3.6 million SNPs. In the same study, the rice HapMap enabled a GWAS of 14 agronomic traits. After controlling for population structure using a compressed mixed linear model (Yu et al., 2006; Zhang et al., 2010), 37 significantly associated SNPs were identified from across the 14 traits. For grain color, width, and quality, some of the SNPs associated with those traits had relatively large effects and were located in well-documented causal genes. However, for most other traits, significantly associated SNPs had relatively small effects and were located in novel loci. A follow-up study by Huang et al. (2012) utilizing the same methodologies but with a larger association panel of 950 rice landraces identified 32 additional novel loci associated with flowering time ten grain-related traits, suggesting that larger sample size increased the power to detect marker-trait associations in the GWAS.

Sorghum (*Sorghum bicolor* L. Moench) is the closest relative of maize within the cultivated cereal grasses (Swigoňová et al., 2004). Compared to rice, sorghum has a slightly lower rate of inbreeding in addition to a larger genome size (730 Mbp in sorghum versus 389 Mbp in rice; International Rice Genome Sequencing Project, 2005; Paterson et al., 2009) but a similar average rate of LD decay across the genome ($r^2 \leq 0.10$ within 150 kbp; Morris et al., 2013). In contrast, the total genome size of maize, a highly outcrossing species, is approximately 2300 Mbp (Schnable et al., 2009), and r^2 estimates of LD decay to less than 0.10 within 10 kbp in diverse samples (Remington et al., 2001). Paterson et al. (2009) used whole-genome shotgun sequencing to assemble the reference genome sequence for sorghum based on the inbred line BTx623. In addition to the larger genome size of sorghum compared to rice, Paterson et al. (2009) reported that 62% of the sorghum genome is comprised of highly repetitive heterochromatin whereas only 15% of the rice genome is heterochromatin. Genomic complexity arising from a high frequency of repetitive regions can make marker discovery a challenging prospect (Gore et al., 2007). Elshire et al. (2011) first proposed a genotype-by-sequencing (GBS) technique to address issues with marker development in large, complex genomes. Demonstrated in maize and barley (*Hordeum vulgare* L.), GBS efficiently reduces genome complexity by using methylation-sensitive restriction enzymes to enrich for genic regions of the genome over repetitive regions (Elshire et al., 2011). By applying GBS technology to sorghum, Morris et al. (2013) was able to generate roughly 265,000 SNPs on a large association panel of 971 accessions. Many of the sorghum accessions in the association panel were introgression lines that carry common early maturity and dwarfing loci, thereby providing a positive control in a GWAS of plant height. All

significant SNP loci associated with plant height located to three well-characterized dwarfing genes (*dw1*, *dw2*, and *dw3*), validating the ability of GWAS to detect large effect candidate genes in a broad association panel. Since the previous analysis did not reflect the typical GWAS scenario, a GWAS analysis of inflorescence architecture, a trait not under strong selection in the association panel, was conducted. Thirteen genes with sequence homology to known genes associated with inflorescence architecture in Arabidopsis, rice, or maize were identified, but several other previously unknown genic regions were also associated with inflorescence architecture. The authors concluded that the GWAS results showed promise in improving the introgression of novel alleles from unadapted sorghum germplasm into elite breeding pools.

The reference genome sequence for maize based on the inbred line B73 became available around the same time as the first sorghum genome sequence (Schnable et al., 2009). The availability of the reference genome quickly enabled the development of the first maize HapMap based on variation discovered among 26 diverse inbred lines (Gore et al., 2009). It also enabled the development of the maize Illumina Infinium SNP genotyping array, which contains almost 50,000 SNP markers (Ganal et al., 2011). This SNP microarray (also referred to as the 50k SNP array) has been used to genotype many diverse populations (Schaefer and Bernardo, 2013; Li et al., 2013; Bouchet et al., 2013; Xue et al., 2013), including 279 inbred lines of a commonly used maize core diversity panel (Flint-Garcia et al., 2005; Cook et al., 2012). The maize diversity panel captures much of the genetic diversity present in public breeding programs worldwide. The large number of markers that the 50k SNP array has provided for the diversity panel has enabled GWAS for several complex traits in maize

including kernel composition traits (Cook et al., 2012) and the hypersensitive response (Olukolu et al., 2013). Olukolu et al. (2013) identified SNPs associated with the hypersensitive defense located in or adjacent to five genes not previously known *a priori*. All five genes had predicted functions involved with the programmed cell death pathway. One goal of this dissertation was to use the genotypic data utilized by Olukolu et al. (2013) to conduct a GWAS of Fusarium ear rot resistance in the maize core diversity panel.

GWAS has some limitations, however. Significant population structure among the association panel can cause spurious associations, and insufficient sample size can limit the detection of rare low frequency alleles (Yu et al., 2006). Different mixed model approaches have been described to properly account for population structure, including: modeling structure through linear covariates generated by a program such as STRUCTURE (**Q** method); principal component analysis of genome-wide markers (**P** method); and the use of a numerator genomic or kinship relationship matrix (**G** and **K** methods, respectively), which is similar to some genomic selection models (Yu et al., 2006; Zhang et al., 2010). Since **Q** is modeled through linear covariates, it can be combined with **K** to allow for higher-level population structure control (**Q**+**K**). However, the use of **Q** alone can be insufficient in accounting for fine-scale structure. Population structure in the original study of association between *Dwarf8* and maize flowering time was modeled only with **Q** (Thornsberry et al., 2001). Re-analysis with the **Q**+**K** model revealed that all associations reported by Thornsberry et al. (2001) were either spurious or grossly overestimated (Larsson et al., 2013). Larsson et al. (2013) concluded that the previous associations of *Dwarf8* with flowering time were artifacts of population structure, and that natural variation in *Dwarf8* is

unlikely to be contributing to natural phenotypic variation observed in flowering time. Even when properly controlling for population structure, true associations can still be missed. Causative alleles that are perfectly associated with population structure are indistinguishable from false positives (Larsson et al., 2013). In addition, rare causative alleles can go undetected if there are is insufficient sample of lines carrying the rare variants in the association panel. In a GWAS of flowering time in a large panel of 2,815 diverse inbred accessions, Romay et al. (2013) examined two known flowering time genes – *Vgt1* and *ZmCCT*. Although both have large additive effects on flowering time, *Vgt1* was represented by more than 80 significantly associated SNPs while *ZmCCT* was represented by one SNP. Investigation of the allele frequencies at these two loci revealed that variation at the *Vgt1* locus was widespread across all major germplasm pools, while variation at the *ZmCCT* locus was almost entirely restricted to tropical material (which has comparatively higher hapolotype diversity than temperate material).

Since the development of the 50k SNP array, the GBS approach described by Elshire et al. (2011) and Glaubitz et al. (in press) has been used to generate 681,257 SNP markers on a large set of 2,815 maize accessions from the germplasm collection maintained at the USDA-ARS North Central Regional Plant Introduction Station (NCRPIS) in Ames, IA (Romay et al., 2013). The NCRPIS collection (also referred to as the Ames collection) represents almost a century of public and private breeding efforts from across the globe. Accessions in this collection include the aforementioned maize core diversity panel, other public breeding lines, private industry lines with expired Plant Variety Protection Act

(exPVPA) certificates, and landraces. The application of the NCRPIS collection genotypic data from Romay et al. (2013) to this dissertation will be discussed in the next section.

Genomic Selection Strategies in Maize

The massive number of genetic markers enabled by SNP genotyping arrays (Ganal et al., 2011) and GBS technologies (Elshire et al., 2011; Glaubitz et al., in press) has enabled genomic selection (GS) strategies in maize. First proposed in animal breeding by Meuwissen et al. (2001), GS utilizes all available genomic information to estimate breeding values for individuals in population. Unlike linkage-based QTL mapping or GWAS, GS does not involve identifying a subset of markers with highest associations with a trait but instead uses a large set of random genome-wide markers to predict breeding values (Bernardo, 2014). Marker-assisted selection (MAS), which relies on precise QTL mapping, is often ill-suited to efficiently select for favorable small-effect QTL in crop species (Heffner et al., 2009). Genomic selection offers an attractive alternative for the improvement of additive polygenic traits in maize, including Fusarium ear rot resistance. Several studies have demonstrated the relatively higher accuracy of prediction from GS models compared to QTL models for maize grain yield, a highly polygenic trait with a low to moderate heritability (Frisch et al., 2010; Albrecht et al., 2011; Riedelsheimer et al., 2012; Zhao et al., 2012; Technow et al., 2012). Compared to phenotypic selection and MAS strategies, GS has also been predicted to improve potential selection accuracies for resistance to Fusarium head blight (causal pathogen *Fusarium graminearum* sensu lato) in wheat (Lorenz et al., 2012) and Northern

corn leaf blight (causal pathogen *Exserohilum turcicum* (Pass.) Leonard and Suggs) in maize (Technow et al., 2013).

Different methods have been proposed to conduct GS, including best linear unbiased prediction (BLUP; Henderson, 1984), ridge regression BLUP (RR-BLUP; Piepho, 2009), and Bayesian (Sorensen and Gianola, 2002) approaches. Before the advent of molecular marker technology, the BLUP method relied on a numerator relationship matrix estimated from pedigrees (**A**) to characterize the covariance between relatives (Henderson, 1977). In modern breeding, the genomic similarity between relatives can be estimated by using a large number of genome-wide SNP markers. Genomic information can be incorporated into a BLUP model two ways: RR-BLUP and genomic BLUP (G-BLUP). Described in detail by Meuwissen et al. (2001) and Habier et al. (2007), RR-BLUP includes marker information by fitting all SNPs as random effects in the prediction model. The typical RR-BLUP model can be written as

$$y = 1\mu + \mathbf{X}g + e$$

where y is a vector of phenotypes, μ is the mean, **X** is a matrix that contains SNP genotypes coded as gene content (0, 1, or 2 copies of the minor allele), g is the effect of each SNP, and e is the residual vector (Bernardo and Yu, 2007). The proportion of genetic variance explained by each SNP is assumed to be equal, and breeding values for individuals in a population are estimated by summing allele effects across all SNPs. Instead of fitting all SNPs simultaneously as random effects, VanRaden (2008) proposed replacing the **A** matrix with a realized genomic relationship matrix (**G**) which is calculated from observed allele

frequencies in marker data as opposed to classical pedigrees. Following Henderson's (1984) mixed model notation, the G-BLUP model can be written as

$$y = \mathbf{X}b + \mathbf{Z}u + e$$

where y is a vector of phenotypes, **X** is an incidence matrix relating the fixed effects to each individual, b is a vector of fixed effects, **Z** is an incidence matrix allocating observations to genetic values, u is a vector containing the additive genetic effects for an individual, and e is the residual vector. Intuitively, the sum $\mathbf{Z}u$ is the vector of breeding values for all individuals. The variance of u is modeled as

$$\mathrm{Var}(u) = \mathbf{G}\sigma_G^2$$

where **G** is the realized genomic relationship matrix and σ_G^2 is the genetic variance (Yu et al., 2006). As discussed by Clark and van der Werf (2013), G-BLUP has three properties that make it more desirable to use than RR-BLUP. First, the dimensions of the genetic effects in the mixed model equations are reduced from $m \times m$ (where m is the number of markers) in RR-BLUP to $n \times n$ (where n is the number of individuals) in G-BLUP, which makes solving the mixed model equations more computationally efficient. Second, prediction accuracies of estimated breeding values from G-BLUP can be computed in the same way as the traditional pedigree-based BLUP method. Lastly, G-BLUP information can be combined with pedigree information in a single step method. Furthermore, Habier et al. (2007) demonstrated that in most cases RR-BLUP and G-BLUP are equivalent models, making G-BLUP the more attractive option given its more efficient computation compared to RR-BLUP.

Neither G-BLUP nor RR-BLUP methodologies explicitly model the effects of known major genes or QTL versus additive polygenic background QTL (Bernardo, 2014), but

several nonlinear Bayesian methods (such as Bayes A, Bayes B, Bayes L, and Bayes R) have been described that allow genetic variance estimates to differ among SNP loci (Meuwissen et al., 2001; Habier et al., 2010; Gianola, 2013). Bayesian prediction models allow for different estimates of genetic variance across all SNP loci according to a prior assumption, and although all Bayesian methods use the same sampling model, the various methods differ only in the prior assumption of how marker variances are distributed across the genome (Gianola, 2013). Although some simulation studies suggest that Bayesian techniques have the potential to produce more accurate breeding value predictions than RR-BLUP or G-BLUP (Habier et al., 2007; Clark et al., 2011), studies with real data show no significant difference among Bayesian, RR-BLUP, and G-BLUP methodologies (Hayes et al., 2009; Moser et al., 2009). Moreover, Bayesian methods are often computationally intensive compared to RR-BLUP and GBLUP, and results can be difficult to biologically interpret (Moser et al., 2009; Gianola, 2013).

Recently, Bernardo (2014) has proposed a simpler alternative to Bayesian techniques to account for the effects of major genes by explicitly modeling known major genes or large-effect QTL as fixed effects in GS models in order to improve prediction accuracies. In general, Bernardo (2014) observed that by including between one and three known large-effect QTL (each with $R^2 \geq 10\%$) in prediction models of testcross grain yield in a simulated recurrent selection population, prediction accuracies significantly improved in almost all cases compared to traditional GS modeling. However, one problem with this approach is that it requires *a priori* knowledge of QTL affecting the trait of interest to be able to explicitly model them. As previously described, QTL associated with Fusarium ear rot resistance are

often have very small effects and are difficult to validate. In spite of this, conducting a GWAS within a breeding population before running GS models may be able to provide enough information on larger-effect QTL to be able to appropriately model them in the prediction models. The final goal of this thesis was two-fold. First, I used the available marker data set from Romay et al. (2013) to conduct a GWAS of Fusarium ear rot resistance within the USDA North Central Regional Plant Introduction Station (NCRPIS) collection of inbred lines and compared those results to those of the aforementioned maize core inbred diversity panel. Secondly, SNPs significantly associated with ear rot resistance in the GWAS were modeled as known large-effect QTL in genomic prediction models of Fusarium ear rot resistance in the NCRPIS collection. Five-fold cross validation was used to compare prediction accuracies from these models to traditional GS models where only additive polygenic background QTL were accounted for.

References

Albrecht, T., V. Wimmer, H.-J. Auinger, M. Erbe, C. Knaak, et al. 2011. Genome-based prediction of testcross values in maize. Theor. Appl. Genet. 123: 339–350.

ApSimon, J.W. 1994. The biosynthetic diversity of secondary metabolites. p. 3–18. *In* Miller, J.D., Trenholm, H.L. (eds.), Mycotoxins in Grain: Compounds Other Than Aflatoxin. 1st ed. Eagan Press, St. Paul, MN.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408: 796–815.

Atwell, S., Y.S. Huang, B.J. Vilhjálmsson, G. Willems, M. Horton, et al. 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature 465: 627–631.

Bacon, C.W., A.E. Glenn, and I.E. Yates. 2008. *Fusarium verticillioides*: managing the endophytic association with maize for reduced fumonisins accumulation. Toxin Rev. 27: 411–446.

Bacon, C.W., and D.M. Hinton. 1996. Symptomless endophytic colonization of maize by *Fusarium moniliforme*. Can. J. Bot. 74: 1195–1202.

Bacon, C.W., I.E. Yates, D.M. Hinton, and F. Meredith. 2001. Biological control of *Fusarium moniliforme* in maize. Environ. Health Perspect. 109 Suppl 2: 325–332.

Beavis, W.D. 1998. QTL analyses: power, precision, and accuracy. p. 145–162. *In* Paterson, A.H. (ed.), Molecular Dissection of Complex Traits. CRC Press, Boca Raton, FL.

Bernardo, R. 2014. Genomewide selection when major genes are known. Crop Sci. 54: 76–88.

Bernardo, R., and J. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. Crop Sci. 47: 1082–1090.

Bolduan, C., T. Miedaner, W. Schipprack, B.S. Dhillon, and A.E. Melchinger. 2009. Genetic variation for resistance to ear rots and mycotoxins contamination in early European maize inbred lines. Crop Sci. 49: 2019–2028.

Bouchet, S., B. Servin, P. Bertin, D. Madur, V. Combes, et al. 2013. Adaptation of maize to temperate climates: mid-density genome-wide association genetics and diversity patterns reveal key genomic regions, with a major contribution of the *Vgt2* (*ZCN8*) locus. PLoS ONE 8: e71377.

Bush, B.J., M.L. Carson, M.A. Cubeta, W.M. Hagler, and G.A. Payne. 2004. Infection and fumonisin production by *Fusarium verticillioides* in developing maize kernels. Phytopathology 94: 88–93.

Center for Food Safety and Applied Nutrition, C. for V.M. 2001. Chemical Contaminants and Pesticides - Guidance for Industry: Fumonisin Levels in Human Foods and Animal Feeds; Final Guidance. Available at http://www.fda.gov/Food/GuidanceComplianceRegulatoryInformation/GuidanceDocuments/ChemicalContaminantsandPesticides/ucm109231.htm (verified 31 October 2013).

Chen, J., J. Ding, H. Li, Z. Li, X. Sun, et al. 2012. Detection and verification of quantitative trait loci for resistance to Fusarium ear rot in maize. Mol. Breed. 30: 1649–1656.

Clark, S.A., J.M. Hickey, and J.H. van der Werf. 2011. Different models of genetic variation and their effect on genomic evaluation. Genet. Sel. Evol. 43: 18.

Clark, R.M., G. Schweikert, C. Toomajian, S. Ossowski, G. Zeller, et al. 2007. Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. Science 317: 338–342.

Clark, S.A., and J. van der Werf. 2013. Genomic best linear unbiased prediction (gBLUP) for the estimation of genomic breeding values. p. 321–330. *In* Gondro, C., van der Werf, J., Hayes, B. (eds.), Genome-Wide Association Studies and Genomic Prediction. Methods in Molecular Biology. Humana Press.

Clements, M.J., C.M. Maragos, J.K. Pataky, and D.G. White. 2004. Sources of resistance to fumonisin accumulation in grain and Fusarium ear and kernel rot of corn. Phytopathology 94: 251–260.

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

Danielsen, M., and F. Jensen. 1998. Genetic characteristics of *Fusarium verticillioides* isolates from maize in Costa Rica. Plant Pathol. 47: 615–622.

Ding, J.-Q., X.-M. Wang, S. Chander, J.-B. Yan, and J.-S. Li. 2008. QTL mapping of resistance to Fusarium ear rot using a RIL population in maize. Mol. Breed. 22: 395–403.

Duncan, K.E., and R.J. Howard. 2010. Biology of maize kernel infection by *Fusarium verticillioides*. Mol. Plant-Microbe Interact. 23: 6–16.

Dwivedi, S., K. Sahrawat, H. Upadhyaya, and R. Ortiz. 2013. Food, nutrition and agrobiodiversity under global climate change. Adv. Agron. 120: 1–128.

Eller, M.S., J.B. Holland, and G.A. Payne. 2008. Breeding for improved resistance to fumonisin contamination in maize. Toxin Rev. 27: 371–389.

Eller, M.S., G.A. Payne, and J.B. Holland. 2010. Selection for reduced Fusarium ear rot and fumonisin content in advanced backcross maize lines and their topcross hybrids. Crop Sci. 50: 2249–2260.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6: e19379.

Escobar, J., S. Lorán, I. Giménez, E. Ferruz, M. Herrera, et al. 2013. Occurrence and exposure assessment of *Fusarium* mycotoxins in maize germ, refined corn oil and margarine. Food Chem. Toxicol. 62: 514–520.

Flint-Garcia, S.A., A. Thuillet, J. Yu, G. Pressoir, S.M. Romero, et al. 2005. Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J. 44: 1054–1064.

Frisch, M., A. Thiemann, J. Fu, T.A. Schrag, S. Scholten, et al. 2010. Transcriptome-based distance measures for grouping of germplasm and prediction of hybrid performance in maize. Theor. Appl. Genet. 120: 441–450.

Ganal, M.W., G. Durstewitz, A. Polley, A. Bérard, E.S. Buckler, 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.

Gelderblom, W.C., K. Jaskiewicz, W.F. Marasas, P.G. Thiel, R.M. Horak, et al. 1988. Fumonisins--novel mycotoxins with cancer-promoting activity produced by *Fusarium moniliforme*. Appl. Environ. Microbiol. 54: 1806–1811.

Gianola, D. 2013. Priors in whole-genome regression: the Bayesian alphabet returns. Genetics: genetics.113.151753.

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, et al. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS ONE (submitted).

Gore, M., P. Bradbury, R. Hogers, M. Kirst, E. Verstege, et al. 2007. Evaluation of target preparation methods for single-feature polymorphism detection in large complex plant genomes. Crop Sci. 47: S–135–S–148.

Gore, M.A., J.-M. Chia, R.J. Elshire, Q. Sun, E.S. Ersoz, et al. 2009. A first-generation haplotype map of maize. Science 326: 1115 –1117.

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

Habier, D., J. Tetens, F.-R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet. Sel. Evol. 42: 5.

Harrison, L.R., B.M. Colvin, J.T. Greene, L.E. Newman, and J.R. Cole. 1990. Pulmonary edema and hydrothorax in swine produced by fumonisin B1, a toxic metabolite of *Fusarium moniliforme*. J. Vet. Diagn. Invest. 2: 217–221.

Hayes, B.J., P.J. Bowman, A.C. Chamberlain, K. Verbyla, and M.E. Goddard. 2009. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet. Sel. Evol. 41: 51.

Heffner, E.L., M.E. Sorrells, and J.-L. Jannink. 2009. Genomic selection for crop improvement. Crop Sci. 49: 1–12.

Henderson, C.R. 1977. Best linear unbiased prediction of breeding values not in the model for records. J. Dairy Sci. 60: 783–787.

Henderson, C.R. 1984. Applications of Linear Models in Animal Breeding. University of Guelph, Guelph, ON, Canada.

Holland, J.B. 2007. Genetic architecture of complex traits in plants. Curr. Opin. Plant Biol. 10: 156–161.

Huang, X., X. Wei, T. Sang, Q. Zhao, Q. Feng, et al. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet. 42: 961–967.

Huang, X., Y. Zhao, X. Wei, C. Li, A. Wang, et al. 2012. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat. Genet. 44: 32–39.

Hung, H.-Y., and J.B. Holland. 2012. Diallel analysis of resistance to Fusarium ear rot and fumonisin contamination in maize. Crop Sci. 52: 2173–2181.

International HapMap Consortium. 2005. A haplotype map of the human genome. Nature 437: 1299–1320.

International Rice Genome Sequencing Project. 2005. The map-based sequence of the rice genome. Nature 436: 793–800.

Kellerman, T.S., W.F.O. Marasas, P.G. Thiel, W.C.A. Gelderblom, M. Cawood, et al. 1990. Leukoencephalomalacia in two horses induced by oral dosing of fumonisin B1. Onderstepoort J. Vet. Res. 57: 269–275.

Kerényi, Z., A. Moretti, C. Waalwijk, B. Olah, and L. Hornok. 2004. Mating type sequences in asexually reproducing *Fusarium* species. Appl. Environ. Microbiol. 70: 4419–4423.

Kim, S., V. Plagnol, T.T. Hu, C. Toomajian, R.M. Clark, et al. 2007. Recombination and linkage disequilibrium in *Arabidopsis thaliana*. Nat. Genet. 39: 1151–1155.

Kimanya, M.E., B. De Meulenaer, D. Roberfroid, C. Lachat, and P. Kolsteren. 2010. Fumonisin exposure through maize in complementary foods is inversely associated with linear growth of infants in Tanzania. Mol. Nutr. Food Res. 54: 1659–1667.

King, S.B., and G.E. Scott. 1981. Genotypic differences in maize to kernel infection by *Fusarium moniliforme*. Phytopathology 71: 1245–1247.

Koehler, B. 1942. Natural mode of entrance of fungi into corn ears and some symptoms that indicate infection. J. Agric. Res. 64: 421–442.

Korte, A., and A. Farlow. 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9: 29.

Kvas, M., W.F.O. (Walter F.O. Marasas, B.D. Wingfield, M.J. Wingfield, and E.T. Steenkamp. 2009. Diversity and evolution of *Fusarium* species in the *Gibberella fujikuroi* complex. Fungal Divers. 34: 1–21.

Kyprianou, M. 2007. Commission Regulation (EC) No 1126/2007 of 28 September 2007 amending Regulation (EC) No 1881/2006 setting maximum levels for certain contaminants in foodstuffs as regards Fusarium toxins in maize and maize products. Off. J. Eur. Union 255: 14–17.

Larsson, S.J., A.E. Lipka, and E.S. Buckler. 2013. Lessons from *Dwarf8* on the strengths and weaknesses of structured association mapping. PLoS Genet 9: e1003246.

Leslie, J.F., C.A.S. Pearson, P.E. Nelson, and T.A. Toussoun. 1990. *Fusarium* spp. from corn, sorghum, and soybean fields in the central and eastern United States. Phytopathology 80: 343–350.

Leslie, J.F., R.D. Plattner, A.E. Desjardins, and C.J.R. Klittich. 1992. Fumonisin B_1 production by strains from different mating populations of *Gibberella fujikuroi* (*Fusarium* section *Liseola*). Phytopathology 82: 341–345.

Li, Z.-M., J.-Q. Ding, R.-X. Wang, J.-F. Chen, X.-D. Sun, et al. 2011. A new QTL for resistance to Fusarium ear rot in maize. J. Appl. Genet. 52: 403–406.

Li, H., Z. Peng, X. Yang, W. Wang, J. Fu, et al. 2013. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat. Genet. 45: 43–50.

Löffler, M., B. Kessel, M. Ouzunova, and T. Miedaner. 2011. Covariation between line and testcross performance for reduced mycotoxin concentrations in European maize after silk channel inoculation of two Fusarium species. Theor. Appl. Genet. 122: 925–934.

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.

Ma, L.-J., H.C. van der Does, K.A. Borkovich, J.J. Coleman, M.-J. Daboussi, et al. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature 464: 367–373.

Melchinger, A.E., H.F. Utz, and C.C. Schön. 1998. Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics 149: 383–403.

Mesterházy, Á., M. Lemmens, and L.M. Reid. 2012. Breeding for resistance to ear rots caused by *Fusarium* spp. in maize – a review. Plant Breed. 131: 1–19.

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

Miller, J.D. 1994. Epidemiology of *Fusarium* ear diseases of cereals. p. 19–36. *In* Miller, J.D., Trenholm, H.L. (eds.), Mycotoxins in Grain: Compounds Other Than Aflatoxin. 1st ed. Eagan Press, St. Paul, MN.

Morris, G.P., P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah, et al. 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc. Natl. Acad. Sci. 110: 453–458.

Moser, G., B. Tier, R.E. Crump, M.S. Khatkar, and H.W. Raadsma. 2009. A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. Genet. Sel. Evol. 41: 56.

Munkvold, G.P., R.L. Hellmich, and L.G. Rice. 1999. Comparison of fumonisin concentrations in kernels of transgenic Bt maize hybrids and nontransgenic hybrids. Plant Dis. 83: 130–138.

Munkvold, G.P., R.L. Hellmich, and W.B. Showers. 1997. Reduced Fusarium ear rot and symptomless infection in kernels of maize genetically engineered for European corn borer resistance. Phytopathology 87: 1071–1077.

Nankam, C., and J.K. Pataky. 1996. Resistance to kernel infection by *Fusarium moniliforme* in the sweet corn inbred IL 125b. Plant Dis. 80: 593–598.

Olukolu, B.A., A. Negeri, R. Dhawan, B.P. Venkata, P. Sharma, et al. 2013. A connected set of genes associated with programmed cell death implicated in controlling the hypersensitive response in maize. Genetics 193: 609–620.

Parsons, M.W., and G.P. Munkvold. 2010. Associations of planting date, drought stress, and insects with Fusarium ear rot and fumonisin B1 contamination in California maize. Food. Addit. Contam. 27: 591–607.

Paterson, A.H., J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood, et al. 2009. The *Sorghum bicolor* genome and the diversification of grasses. Nature 457: 551–556.

Pérez-Brito, D., D. Jeffers, D. González-de-León, M. Khairallah, M. Cortés-Cruz, et al. 2001. QTL mapping of *Fusarium moniliforme* ear rot resistance in highland maize, Mexico. Agrociencia 35: 181–196.

Piepho, H.P. 2009. Ridge regression and extensions for genomewide selection in maize. Crop Sci. 49: 1165–1176.

Pray, C.E., J.P. Rheeder, M. Gouse, Y. Volkwyn, L. van der Westhuizen, et al. 2013. Bt maize and fumonisin reduction in South Afria: potential health impacts. p. 43–60. *In* Falck-Zepeda, J., Gruère, G., Sithole-Niang, I. (eds.), Genetically Modified Crops in Africa: Economic and Policy Lessons from Countries South of the Sahara. IFPRI, Washington, DC.

Proctor, R.H., R.D. Plattner, A.E. Desjardins, M. Busman, and R.A.E. Butchko. 2006. Fumonisin production in the maize pathogen *Fusarium verticillioides*: genetic basis of naturally occurring chemical variation. J. Agric. Food Chem. 54: 2424–2430.

Remington, D.L., and M.D. Purugganan. 2003. Candidate genes, quantitative trait loci, and functional trait evolution in plants. Int. J. Plant Sci. 164: S7–S20.

Remington, D.L., J.M. Thornsberry, Y. Matsuoka, L.M. Wilson, S.R. Whitt, et al. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci. U. S. A. 98: 11479 –11484.

Rheeder, J.P., W.F.O. Marasas, P.G. Thiel, E.W. Sydenham, G.S. Shephard, et al. 1992. *Fusarium moniliforme* and fumonisins in corn in relation to human esophageal cancer in Transkei. Phytopathology 82: 353.

Riedelsheimer, C., A. Czedik-Eysenberg, C. Grieder, J. Lisec, F. Technow, et al. 2012. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat. Genet. 44: 217–220.

Riley, R.T., D.M. Hinton, W.J. Chamberlain, C.W. Bacon, E. Wang, et al. 1994. Dietary fumonisin B1 induces disruption of sphingolipid metabolism in Sprague-Dawley rats: a new mechanism of nephrotoxicity. J. Nutr. 124: 594–603.

Robertson, L.A., C.E. Kleinschmidt, D.G. White, G.A. Payne, C.M. Maragos, et al. 2006. Heritabilities and correlations of Fusarium ear rot resistance and fumonisin contamination resistance in two maize populations. Crop Sci. 46: 353–361.

Robertson-Hoyt, L.A., M.P. Jines, P.J. Balint-Kurti, C.E. Kleinschmidt, D.G. White, et al. 2006. QTL mapping for Fusarium ear rot and fumonisin contamination resistance in two maize populations. Crop Sci. 46: 1734–1743.

Robertson-Hoyt, L.A., C.E. Kleinschmidt, D.G. White, G.A. Payne, C.M. Maragos, et al. 2007. Relationships of resistance to Fusarium ear rot and fumonisin contamination with agronomic performance of maize. Crop Sci. 47: 1770–1778.

Romay, M.C., M.J. Millard, J.C. Glaubitz, J.A. Peiffer, K.L. Swarts, et al. 2013. Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. 14: R55.

Santiago, R., A. Cao, R.A. Malvar, L.M. Reid, and A. Butrón. 2013. Assessment of corn resistance to fumonisin accumulation in a broad collection of inbred lines. Field Crops Res. 149: 193–202.

Schaefer, C.M., and R. Bernardo. 2013. Genomewide association mapping of flowering time, kernel composition, and disease resistance in historical Minnesota maize inbreds. Crop Sci. 53: 2518–2529.

Schnable, P.S., D. Ware, R.S. Fulton, J.C. Stein, F. Wei, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112–1115.

Shephard, G.S., M.E. Kimanya, K.A. Kpodo, G.J.B. Gnonlonfin, and W.C.A. Gelderblom. 2013. The risk management dilemma for fumonisin mycotoxins. Food Control 34: 596–600.

Shirima, C.P., M.E. Kimanya, J.L. Kinabo, M.N. Routledge, C. Srey, et al. 2013. Dietary exposure to aflatoxin and fumonisin among Tanzanian children as determined using biomarkers of exposure. Mol. Nutr. Food Res. 57: 1874–1881.

Sorensen, D., and D. Gianola. 2002. Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics. Springer, New York, NY.

Swigoňová, Z., J. Lai, J. Ma, W. Ramakrishna, V. Llaca, et al. 2004. Close split of sorghum and maize genome progenitors. Genome Res. 14: 1916–1923.

Sydenham, E.W., G.S. Shephard, P.G. Thiel, W.F.O. Marasas, and S. Stockenstrom. 1991. Fumonisin contamination of commercial corn-based human foodstuffs. J. Agric. Food Chem. 39: 2014–2018.

Technow, F., A. Bürger, 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 3: 197–203.

Technow, F., C. Riedelsheimer, T.A. Schrag, and A.E. Melchinger. 2012. Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. Theor. Appl. Genet. 125: 1181–1194.

Thornsberry, J.M., M.M. Goodman, J. Doebley, S. Kresovich, D. Nielsen, et al. 2001. *Dwarf8* polymorphisms associate with variation in flowering time. Nat Genet 28: 286–289.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91: 4414–4423.

Voss, K.A., J.B.G. Waes, and R.T. Riley. 2006. Fumonisins: current research trends in developmental toxicology. Mycotoxin Res. 22: 61–69.

Warren, H.L., and T. Kommedahl. 1973. Prevalence and pathogenicity to corn of Fusarium species from corn roots, rhizosphere, residues, and soil. Phytopathology 63: 1288–1290.

Watanabe, M., T. Yonezawa, K. Lee, S. Kumagai, Y. Sugita-Konishi, et al. 2011. Molecular phylogeny of the higher and lower taxonomy of the *Fusarium* genus and differences in the evolutionary histories of multiple genes. BMC Evol. Biol. 11: 322.

White, D.G. (Ed). 1999. Fusarium kernel or ear rot. p. 45–46. *In* Compendium of Corn Diseases. 3rd ed. American Phytopathological Society.

Xue, Y., M.L. Warburton, M. Sawkins, X. Zhang, T. Setter, et al. 2013. Genome-wide association analysis for nine agronomic traits in maize under well-watered and water-stressed conditions. Theor. Appl. Genet. 126: 2587–2596.

Yates, I.E., C.W. Bacon, and D.M. Hinton. 1997. Effects of endophytic infection by *Fusarium moniliforme* on corn growth and cellular morphology. Plant Dis. 81: 723–728.

Yu, J., and E.S. Buckler. 2006. Genetic association mapping and genome organization of maize. Curr. Opin. Biotechnol. 17: 155–160.

Yu, J., G. Pressoir, W.H. Briggs, I. Vroh Bi, M. Yamasaki, et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38: 203–208.

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

Zhao, Y., M. Gowda, W. Liu, T. Würschum, H.P. Maurer, et al. 2012. Accuracy of genomic selection in European maize elite breeding populations. Theor. Appl. Genet. 124: 769–776.

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

Zila, C. 2011. Utilizing exotic maize germplasm to improve Fusarium ear rot resistance and validate a photoperiod flowering time QTL. Available at http://repository.lib.ncsu.edu/ir/handle/1840.16/6957. M.S. thesis. North Carolina State University, Raleigh, NC.

CHAPTER 2: Proposal to Release Inbred Maize Lines NC301 and NC303

Prepared for submission to the Journal of Plant Registrations

J.B. Holland *****, J.C. Brewer, Magen S. Eller, Leilani A. Robertson-Hoyt, Gary A. Payne, Stella A. Salvo, Donald White, and Charles T. Zila

J.B. Holland, J.C. Brewer, Magen S. Eller, and Stella A. Salvo, USDA-ARS Plant Science Research Unit, Box 7620 North Carolina State University, Raleigh NC 27695-7620; J.B. Holland, Magen S. Eller, Leilani A. Robertson-Hoyt, and Charles T. Zila, Department of Crop Science, North Carolina State University, Raleigh NC 27695-7620; Gary A. Payne Department of Plant Pathology, North Carolina State University, Raleigh NC 27695-7597; Donald White, Department of Crop Sciences, University of Illinois, Urbana, IL 61801.

Abstract

Improvement of elite maize (*Zea mays* L.) breeding lines for resistance to Fusarium ear rot (caused by *Fusarium verticillioides* (Sacc.) Nirenberg) in the United States has been a challenge due to a lack of adapted resistance donor sources. To allow for better incorporation of Fusarium ear rot resistance alleles into elite germplasm pools, two maize inbred lines – NC301 and NC303, referred to as GEFR400-9-5-2-3-8 and GEFR400-9-5-2-3-1 during development, respectively – were developed by the USDA-ARS Plant Science Research Unit and North Carolina State University in Raleigh, NC in cooperation with the University of Illinois in Urbana, IL and Illinois Foundation Seeds. NC301 and NC303 are BC_4F_3-derived selections from a population created by backcrossing the unadapted donor line GE440 into the elite Stiff Stalk recurrent parent FR1064 (Illinois Foundation Seeds). Both NC301 and NC303 possess significantly ($P<0.001$) improved ear rot resistance and resistance to fumonsin contamination compared to FR1064 on an inbred *per se* basis but are not significantly different than FR1064 for many agronomic traits. Based on their improved disease resistance and favorable agronomic characteristics, NC301 and NC303 will be useful for improving Fusarium ear rot resistance in elite Stiff Stalk breeding populations.

Introduction

Fusarium ear rot disease of maize (*Zea mays* L.), caused by *Fusarium verticilliodes* (Sacc.) Nirenberg (synonym *F. moniliforme* Sheldon) (teleomorph: *Gibberella moniliformis*) and *F. proliferatum* (Matsushima) Nirenberg (teleomorph: *G. intermedia*), is prevalent in many maize producing regions. In addition to directly reducing yield potential, this disease

can also result in the accumulation of the mycotoxin fumonisin in affected grain (Bush, Carson, et al., 2004). Consumption of fumonisin-contaminated grain has been associated with human and animal diseases, including equine leukoencephalamalacia, porcine pulmonary edema, and human esophageal cancer (Munkvold and Desjardins, 1997, Thiel, Marasas, et al., 1992). Resistances to Fusarium ear rot and to the accumulation of fumonisin in maize grain are highly desirable, but complete immunity has not been reported (Mesterházy, Lemmens, et al., 2011, Munkvold and Desjardins, 1997, Smith and Madsen, 1949). Resistances to both of these aspects of the disease vary quantitatively, conditioned predominantly by numerous small effect genes (Robertson, Kleinschmidt, et al., 2006, Robertson-Hoyt, Jines, et al., 2006, Smith and Madsen, 1949).

Despite the complexity of inheritance and expression of resistances to Fusarium ear rot and fumonisin contamination, these traits have moderate to high heritability when evaluated in sufficiently replicated trials under artificial inoculation (Robertson, Kleinschmidt, et al., 2006, Eller, Payne, et al., 2008, Bolduan, Miedaner, et al., 2009). Unfortunately, many of the best sources of resistance are unadapted or obsolete, agronomically poor inbred lines (Clements, Maragos, et al., 2004). Thus, selection for improved Fusarium ear rot resistance needs to be performed in tandem with selection for improved agronomic performance to produce inbred lines that combine higher levels of resistance with acceptable agronomic potential. Eller et al. (2010) demonstrated that backcrossing the unadapted resistant inbred GE440 into the elite commercial inbred line FR1064 in combination with selection for improved ear rot resistance resulted in the

recovery of progeny lines with significantly better disease resistance than FR1064 without a significant loss of topcross yield potential.

Following additional generations of line purification and selection, we have developed two inbred lines, NC301 and NC303, resulting from the germplasm developed by Eller et al. (2010) that have demonstrated superior resistance to both Fusarium ear rot and contamination by fumonisin in combination with good agronomic performance in combination with an unrelated tester.

Materials and Methods

Breeding cross and line development

Dr. D.G. White (University of Illinois) crossed FR1064 to GE440, backcrossed the F_1 of this cross to FR1064, and selfed 215 BC_1F_1 plants to form 215 $BC_1F_{1:2}$ families. Dr. White sent these 215 families to USDA-ARS at North Carolina State University, where they were evaluated as described by Robertson et al. (2006). The ten families with lowest grain fumonisin concentrations identified by Robertson et al. (2006) were backcrossed to FR1064 an additional three times, as described by Eller et al. (2010), to create 455 BC_4F_1 families. These families were evaluated in two environments for Fusarium ear rot resistance, and 20 families with lowest mean ear rot values were selected for advancement (Eller, Payne, et al., 2010). Three to four $BC_4F_{1:3}$ lines per selected family were evaluated across two years, and ten were randomly selected from each of the 12 most resistant $BC_4F_{1:3}$. Sublines were either $BC_4F_{4:5}$ or $BC_4F_{3:4}$ generation, depending on availability of seed.

Field evaluation of inbred GEFR lines – experimental design

Inbred *per se* disease evaluations were conducted between 2009 and 2011 across multiple locations in North Carolina (Figure 2.1). In 2009, 80 $BC_4F_{4:5}$ and 40 $BC_4F_{3:4}$ experimental GEFR lines were evaluated for disease resistance at four North Carolina Department of Agriculture research stations (Clayton, Jackson Springs, Kinston, and Lewiston-Woodville). Six entries of FR1064 and two entries of GE440 were included as well bringing the total number of entries to 128. The experimental plots were single rows arranged in a randomized 8 × 16 α-lattice block design with two replications in each of the four locations for a total of 1024 experimental plots.

Based on the results from the 2009 field season, the 50 most resistant and 10 most susceptible lines were selected to be tested again in 2010. $BC_4F_{4:6}$ and $BC_4F_{3:5}$ seed stocks were used due to an insufficient amount of $BC_4F_{4:5}$ and $BC_4F_{3:4}$ remnant seed. Three of the four locations used in 2009 were chosen again to evaluate the selected subset of lines (Clayton, Kinston, and Lewiston-Woodville). In addition to the 60 selected GEFR lines, the entry list included six entries of FR1064, two entries of GE440, and two entries of NC478 to bring the total number of entries to 70. The experimental plots were arranged in a randomized 7 × 10 α-lattice block design with two replications in each of the three locations for a total of 420 plots.

The 20 lines with lowest mean Fusarium ear rot scores, based on combined 2009 and 2010 results, were selected for a final test in 2011. The same seed stocks and locations used in 2010 were used again in 2011. Two entries of FR1064, two entries of GE440, and one entry of NC478 were included bringing the total number of entries to 25. The experimental

plots were arranged in a randomized 5 × 5 α-lattice block design with three replications in each of the three locations for a total of 225 plots. Before harvest, the NC478 check plots were abandoned due to a planting error.

Field evaluation of topcross GEFR hybrids

Topcross hybrids of the GEFR lines were evaluated for yield potential and disease resistance between 2010 and 2012 at the same locations chosen for inbred *per se* evaluations (Figure 2.1). In 2009, each of the 120 GEFR lines as well as FR1064 and GE440 were topcrossed as males onto NC478 to produce F_1 hybrid seed. NC478 is a public inbred line of temperate non-Stiff Stalk and tropical pedigree origin ([PH X105A × H5)Agroceres 155] × NC262A; http://www.cropsci.ncsu.edu/maize/germplasm_450_498.html) with good combining ability with Stiff Stalk maize lines but a high level of susceptibility to Fusarium ear rot. Corresponding to the inbred line selections from the 2009 inbred disease study, 60 topcross hybrids were chosen to evaluate in 2010 representing the selected parental GEFR lines. The final entry list included two entries of the NC478×FR1064 topcross, one entry of the NC478×GE440 topcross, and one entry each of the commercial check hybrids Pioneer brand 31G66 and Dekalb brand DK697 to bring the total number of entries to 64. The topcross hybrids were evaluated for disease resistance in each of the four locations chosen to evaluate the GEFR inbred lines *per se* in 2009 (Clayton, Jackson Springs, Kinston, and Lewiston-Woodville). The experimental plots were single rows arranged in a randomized 8 × 8 α-lattice block design with two replications in each of the four locations for a total of 512 experimental plots. Each plot was planted with 25 kernels and was then thinned to a uniform

stand of 20 plants. Topcross hybrids were also evaluated separately for yield and other agronomic characteristics at the same locations, although the Jackson Springs yield trial was abandoned after planting due to a sting nematode infestation. The experimental plots were double rows with the same blocking structure and total plot number as the disease trial. Each row of the plot was planted with 25 kernels, and the whole plot was thinned to a uniform stand of 40 plants.

In 2011, 20 topcross hybrids were evaluated representing the lines remaining in inbred *per se* evaluations. The same checks were used as in 2010 bringing the total number of entries to 25, and trials were conducted at the same three locations as the inbred disease evaluations (Clayton, Kintson, and Lewiston-Woodville). Experimental plots in both the hybrid disease trial and hybrid yield trial were arranged in a randomized 5 × 5 α-lattice block design with three replications in each of three locations for a total of 225 plots.

In the summer nursery 2011, the 20 remaining GEFR lines were topcrossed onto a second tester for evaluations alongside the NC478 topcross hybrids in 2012. Each line was topcrossed as a female to the F_1 single-cross tester LH283×LH287. LH283 and LH287 are both proprietary inbred lines of non-Stiff Stalk pedigree origin from Holden's Foundation Seeds, LLC. Fouty-nine experimental entries consisting of the 20 NC478 topcrosses, 20 LH283×LH287 topcrosses, two entries each of NC478×FR1064 and FR1064×(LH283×LH287), one entry each of NC478×GE440 and GE440×(LH283×LH287), and one entry each of Pioneer brand 31G66, Pioneer brand 31G98, and Dekalb brand DK697 were evaluated in separate yield and disease trials in 2012. Both the yield and disease trials were arranged in a randomized 7 × 7 α-lattice block design with three replications per

location. The yield trial was conducted at Clayton, Kinston, and Lewiston-Woodville, while the disease trial was conducted at only Clayton and Kinston.

Inoculation technique

Inbred and hybrid disease trials across all years were inoculated using the method described in detail by Robertson et al. (2006). In 2009 and 2010, six isolates of *Fusarium verticillioides* described by Eller (2009) were used for inoculation. These isolates, referred to as NC-i6, NC-i7, NC-i9, NC-n16, NC-n17, and NC-n22, were collected from infected maize seeds produced in North Carolina and have been shown to cause ear rot and produce high levels of fumonisin. All six isolates can be obtained through the Fusarium Research Center collection (http://frc.cas.psu.edu/). In spring 2011, new isolates were isolated from susceptible plots of the 2010 GEFR inbred disease study. Four new isolates were identified with increased fumonisin production compared to four of the six original isolates when grown on sterile cracked maize kernels. The four new *F. verticillioides* isolates and NC-n16 and NC-n17 were used to inoculate inbred and hybrid disease trials in 2011 and 2012. Across all years, isolates were cultured independently on a medium of potato dextrose agar (Fisher Scientific, Pittsburgh, PA). Conidia were collected from the cultures by rinsing each Petri plate with distilled water and using a paint brush to help loosen conidia. Conidia suspension from the different isolates was then combined and strained through cheesecloth to remove residual agar debris. Spores were counted in the mixed suspension, and the suspension was then diluted to approximately 2×10^6 conidia mL^{-1}.

All experimental plots were inoculated twice. The first inoculation occurred one week after silking by injecting 5 mL of the conidia suspension into the silk channel of the primary ear of the first twelve plants in each plot. The second inoculation occurred two weeks after silking by injecting 5 mL of the conidia suspension near the base of the ear of the same plants inoculated in the first week. Injections were administered using an Allflex draw-off syringe (Allflex USA, DFW Airport, TX) fitted with a 16-gauge veterinary needle. Syringes were connected to a modified Solo backpack sprayer (Solo, Newport News, VA) to hold the conidia suspension, and one drop of Tween-20 was added to each liter of suspension to break water surface tension.

Phenotypic trait measurements

During the growing season, percent stand, anthesis date and silking date (Clayton only), plant height (cm), primary ear height (cm), root lodging, and stalk lodging data were collected for each plot in the hybrid yield trial experiments. Anthesis dates and silking dates were recorded as the date when 50% of plants within the plot were shedding pollen or silking, respectively. Plant and ear heights were measured on a random sample of two plants within each row of a double-row plot. Plant height was measured as the height (cm) from the soil to the topmost node directly below the tassel, and ear height was measured as the height from the soil to the node connected to the shank of the primary ear. Root lodging and stalk lodging were recorded as the number of plants lodged with intact stalks and the number of plants lodged with broken stalks, respectively. Root-lodged plants are defined as plants leaning greater than 30% from vertical but with intact stalks, and stalk-lodged plants are

defined as plants with stalks broken or cracked below the primary ear. Percentage of erect plants was calculated from the lodging scores by subtracting root lodging and stalk lodging scores from the stand count and dividing by stand count.

At maturity, ten of the twelve inoculated ears were collected by hand from each plot of the inbred and hybrid disease trials. Each of the ten ears within each experimental plot was visually scored for Fusarium ear rot symptoms. Scores were assigned to each ear in increments of 5% from 0% to 100% diseased based on the percentage of the ear presenting ear rot symptoms. Scored ears within a plot were then shelled in bulk and ground into a fine powder using a Romer II Series Mill (Romer Labs, Union, MO). A 20-g sample of ground grain from each plot was used to assay for fumonisin content. The amount of fumonisin B1 present in each sample was tested using a Diagnostix fumonisin ELISA assay kit (Diagnostix, Mississauga, ON, Canada).

At maturity, each in plot in the hybrid yield trials was mechanically harvested, and grain yield and moisture content (%) were recorded.

Statistical methods

ASReml 3.0 software (Gilmour, Gogel, et al., 2009) was used to analyze all data collected from the four field seasons of the various experiments. Particular location and year combinations were considered as distinct environments for computational purposes. In total, ten environments were analyzed in the inbred disease study, and nine environments were analyzed in the hybrid disease and yield studies. In both the inbred and hybrid disease studies, large residuals were associated with large predicted values for both ear rot score and

fumonisin content, so all analyses were performed on a natural logarithmic transformation on both responses.

Within each study, each environment was first analyzed separately using an α-lattice model fitting a fixed entry effect and random replication and block within replication effects. Models for ear rot scores and fumonisin content were also weighted by the number of ears scored within each plot. Environments with insignificant entry effects were reanalyzed by modifying the residual error matrix structure to an anisotropic correlated error structure (Brownie, Bowman, et al., 1993). Akaike's Information Criterion (Akaike, 1974) was used to determine if the correlated error structure improved model fit at those environments.

After checking each environment separately, a combined multi-environment analysis was performed for each trait within each experiment. The model for each trait consisted of a fixed entry effect, random environment, entry × environment, replication within environment, and block within replication effects, and a heterogeneous error variance structure across locations. An anisotropic correlated error structure was applied at any environment within the combined analysis that showed improved model fit in the individual location analyses.

Heritabilities of ear rot resistance and fumonisin content were calculated on both an inbred *per se* and a hybrid basis using an equation described by Cullis et al. (2006). This method produces a heritability estimate that is similar to heritability estimated on a per-plot basis (Hung, Browne, et al., 2012). The same models as above were used except that entry was considered a random effect sampled from the potential population of all possible

selected GEFR inbred lines in order to obtain a variance component for entry. Heritability was estimated as

$$\widehat{H}_c = 1 - \frac{V_{PPE}}{2\hat{\sigma}_G^2}$$

where V_{PPE} is the average prediction error variance for all pairwise comparison of entries (obtained directly from ASReml prediction output) and $\hat{\sigma}_G^2$ is the estimated genetic variance.

Characteristics

Significant ($P < 0.001$) genotypic variation was observed for all traits in the multi-environment analyses. Among the most resistant 20 GEFR lines tested across three years, mean ear rot ranged from 10.6% to 24.3% with an overall mean of 17.3% (Table 2.1). The mean fumonisin content observed in the best 20 GEFR lines ranged from 12.0 ppm to 24.8 ppm with an overall mean of 20.0 ppm. By comparison, FR1064 exhibited ear rot severity and a fumonisin content of 43.5% and 48.7 ppm, while GE440 had ear rot severity and a fumonisin content of 1.7% and 10.9 ppm, respectively. All GEFR inbred lines as well as GE440 differed significantly from FR1064 at $P = 0.05$ for both ear rot severity and fumonisin content. Heritability for the inbred lines *per se* was 0.84 for ear rot resistance and 0.80 for fumonisin content.

Mean ear rot of the NC478 topcrosses of the most resistant 20 GEFR lines tested across three years ranged from 7.6% to 12.4% with an average of 9.4%, and mean fumonisin content ranged from 13.4 ppm to 26.6 ppm with an average of 19.9 ppm (Table 2.1). Mean ear rot of the LH283×LH287 topcrosses tested in one year ranged from 8.0% to 16.6% with

an average of 12.6%, and mean fumonisin content ranged from 14.9 ppm to 36.7 ppm with an average of 24.4 ppm (Table 2.2). Six of the NC478 topcross hybrids were significantly better compared to the NC478×FR1064 check at $P = 0.05$ for both ear rot severity and fumonisin content, and three of the LH283×LH287 topcross hybrids were significantly better than the FR1064×(LH283×LH287) check for ear rot severity and fumonisin content. Heritability for the hybrids across all testers and years was 0.59 for ear rot resistance and 0.57 for fumonisin content.

Grain yield for the NC478 topcrosses across three years ranged from 8.253 Mg ha^{-1} to 9.381 Mg ha^{-1} with an overall average of 8.740 Mg ha^{-1} (Table 2.1). Sixteen of the topcross hybrids did not differ significantly from the NC478×FR1064 at at $P = 0.05$ for grain yield. Grain yield data from the LH283×LH287 topcross hybrids indicate that the FR1064×(LH283×LH287) topcross check had a mean of 6.384 Mg ha^{-1}, significantly less than the GE440×(LH283×LH287) topcross check (7.763 Mg ha^{-1}) at $P = 0.05$ (Table 2.2). Lower germination rates of FR1064×(LH283×LH287) (resulting from poor seed quality) compared to other entires may have contributed to this result. Further field testing will be required to accurately determine the yield potential of FR1064×(LH283×LH287) with respect to GE440×(LH283×LH287).

From among the advanced GEFR lines tested, we selected two inbreds that represent the best combinations of disease resistance and agronomic performance. The two release lines were originally coded as GEFR400-9-5-2-3-8 and GEFR400-9-5-2-3-1; both are $BC_4F_{4:5}$ lines derived from a common BC_4F_4 ancestor. GEFR400-9-5-2-3-8 was coded as NC301 for release and its sister line, GEFR400-9-5-2-3-1, was coded as NC303.

NC301 had significantly better resistance to Fusarium ear rot and fumonisin contamination than FR1064 but similar agronomic performance and yield potential across four years of inbred *per se* and hybrid evaluations. The only observed detrimental character of the NC301 topcross to NC478 was its significantly greater grain moisture (0.9 percentage points) than the FR1064 topcross (Table 2.1). The NC301 topcross had significantly lower plant and ear heights than the FR1064 topcross, despite GE440's tall stature. Otherwise, we observed no other significantly differences between the agronomic performance of NC301 and FR1064 topcrosses. NC301 was not significantly different from GE440 for fumonisin content on either an inbred *per se* or topcross basis.

NC303 was also significantly more resistant to both Fusarium ear rot and fumonisin contamination than the FR1064 recurrent parent. NC303 appears to have a lower level of resistance than NC301, but a higher yield potential in topcross hybrids (Table 2.1).

Discussion

Previous studies have shown that it is possible to select for high Fusarium resistance without negatively impacting other agronomic characteristics (Robertson-Hoyt, Kleinschmidt, et al., 2007). NC301 and NC303 are ideal candidates for release as germplasm lines for use by commercial breeding programs. Given the inherent difficulty in incorporating exotic sources of Fusarium resistance into elite germplasm, NC301 has excellent potential as an adapted source of resistance for commercial breeding efforts. NC303, although more susceptible to Fusarium ear rot than NC301 on an inbred *per se* basis, shows superior rot resistance compared to FR1064 as well as the highest yield potential among all experimental

entries across three years of testing. NC301 and NC303 could be used as resistance donor sources to improve elite Stiff Stalk breeding populations while minimizing the effect of linkage drag on agronomic traits compared to unadapted donor sources.

References

Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19: 716-723.

Bolduan, C., T. Miedaner, W. Schipprack, B.S. Dhillon and A.E. Melchinger. 2009. Genetic variation for resistance to ear rots and mycotoxin contamination in early European maize inbred lines. Crop Science 49: 2019-2028.

Brownie, C., D.T. Bowman and J.W. Burton. 1993. Estimating spatial variation in analysis of data from yield trials: a comparison of methods. Agronomy Journal 85: 1244-1253.

Bush, B.J., M.L. Carson, M.A. Cubeta, W.M. Hagler and G.A. Payne. 2004. Infection and Fumonisin production by *Fusarium verticilliodes* in developing maize kernels. Phytopathology 94: 88-93.

Clements, M.J., C.M. Maragos, J.K. Pataky and D.G. White. 2004. Sources of resistance to fumonisin accumulation in grain and fusarium ear rot and kernel rot of corn. Phytopathology 94: 251-260.

Cullis, B.R., A.B. Smith and N.E. Coombes. 2006. On the design of early generation variety trials with correlated data. Journal of Agricultural Biological and Environmental Statistics 11: 381-393.

Eller, M., G.A. Payne and J.B. Holland. 2008. Breeding for improved resistance to fumonisin contamination in maize. Toxin Reviews 27: 371-389.

Eller, M., G.A. Payne and J.B. Holland. 2010. Selection for reduced Fusarium ear rot and fumonisin content in advanced backcross maize lines and their topcross hybrids. Crop Science 50: 2249-2260.

Eller, M.R. 2009. Improving resistance to Fusarium ear rot and fumonisin contamination and increasing yield with exotic maize germplasm. North Carolina State University, Raleigh, NC.

Gilmour, A.R., B.J. Gogel, B.R. Cullis and R. Thompson. 2009. ASReml User Guide Release 3.0VSN International, Ltd., Hemel Hempstead, UK.

Hung, H.Y., C. Browne, K. Guill, N. Coles, M. Eller, A. Garcia, et al. 2012. The relationship between parental genetic or phenotypic divergence and progeny variation in the maize Nested Association Mapping population. Heredity 108: 490-499.

Mesterházy, Á., M. Lemmens and L.M. Reid. 2011. Breeding for resistance to ear rots caused by Fusarium spp. in maize – a review. Plant Breeding 131: 1-19.

Munkvold, G.P. and A.E. Desjardins. 1997. Fumonisins in maize - Can we reduce their occurrence? Plant Disease 81: 556-565.

Robertson-Hoyt, L.A., M.P. Jines, P.J. Balint-Kurti, C.E. Kleinschmidt, D.G. White, G.A. Payne, et al. 2006. QTL mapping for Fusarium ear rot and fumonisin contamination resistance in two maize populations. Crop Science 46: 1734-1743. doi:10.2135/cropsci2005.12-0450.

Robertson-Hoyt, L.A., C.E. Kleinschmidt, D.G. White, G.A. Payne, C.M. Maragos and J.B. Holland. 2007. Relationships of resistance to Fusarium ear rot and fumonisin contamination with agronomic performance of maize. Crop Science 47: 1770-1778.

Robertson, L.A., C.E. Kleinschmidt, D.G. White, G.A. Payne, C.M. Maragos and J.B. Holland. 2006. Heritabilities and correlations of Fusarium ear rot resistance and fumonisin contamination resistance in two maize populations. Crop Science 46: 1420-1420.

Smith, F.L. and C.B. Madsen. 1949. Susceptibility of inbred lines of corn to Fusarium ear rot. Agronomy Journal: 347-348.

Thiel, P.G., W.F.O. Marasas, E.W. Sydenham, G.S. Shepard and W.C.A. Gelderbloom. 1992. The implications of naturally occuring levels of fumonisins in corn for human and animal health. Mycopathologia 117: 3-9.

Table 2.1. Least square means for the twenty superior GEFR lines and checks for Fusarium ear rot resistance and fumonisin content as inbreds *per se* (averaged across ten environments) and for Fusarium ear rot resistance, fumonisin content, grain yield, and other agronomic traits as topcrosses to the tester NC478 (averaged across nine environments).

	Inbred *per se*		Topcross to NC478								
	Fusarium ear rot	Fumonisin content	Fusarium ear rot	Fumonisin content	Grain yield	Grain moisture	Days to anthesis	Days to silking	Plant height	Ear height	Erect plants
Experimental lines	%[†]	ppm[†]	%[†]	ppm[†]	Mg ha^{-1}	%	d	d	cm	cm	%
GEFR397-8-2-2-1-1-1	20.2[‡]	24.0[‡]	10.1	19.3	9.049	19.3	69.1	69.4	213.2	109.6	95.4
GEFR397-8-2-2-1-2-1	15.8[‡]	19.2[‡]	8.8[‡]	18.7	9.005	18.9	68.7	69.6	203.0	100.9	95.4
GEFR397-8-2-2-1-3-1	16.3[‡]	15.9[‡]	8.4[‡]	13.3[‡]	8.604	19.2	69.6	70.5	207.4	109.5	95.2
GEFR397-8-2-2-1-3-2	11.6[‡]	12.6[‡]	7.9[‡]	16.1[‡]	8.322	18.9	69.6	70.4	205.0	104.8	95.2
GEFR397-8-2-2-1-5-2	21.3[‡]	22.4[‡]	12.3	26.6	8.886	19.6	69.7	70.9	210.8	104.3	95.3
GEFR399-1-5-1-3-2-1	17.0[‡]	22.1[‡]	9.0[‡]	16.7[‡]	8.253	18.5	67.7	68.1	196.9	96.9	95.3
GEFR399-1-5-1-3-2-2	18.3[‡]	23.2[‡]	8.9[‡]	16.6[‡]	8.604	18.6	67.5	68.0	194.5	92.8	95.2
GEFR399-1-5-1-3-4-2	14.9[‡]	16.1[‡]	9.0[‡]	14.8[‡]	8.604	18.5	67.2	68.0	194.3	90.4	95.2
GEFR399-2-10-4-1-2-2	17.6[‡]	23.3[‡]	7.7[‡]	23.3	8.503	19.0	68.1	68.6	195.9	90.6	95.3
GEFR399-2-10-6-2-1-1	14.5[‡]	17.2[‡]	9.7	19.5	8.924	19.5	67.7	68.7	197.1	89.2	95.4
GEFR399-2-10-6-2-1-2	18.1[‡]	18.5[‡]	9.6	23.4	8.905	19.5	68.1	69.2	194.5	86.9	95.3
GEFR399-2-10-6-2-3-2	18.0[‡]	21.9[‡]	11.5	25.0	8.253	19.2	68.7	69.4	204.1	93.1	95.0
GEFR399-2-10-6-2-5-1	20.5[‡]	22.7[‡]	7.6[‡]	18.2	8.848	19.3	68.1	69.2	196.2	89.2	95.2
GEFR399-2-10-6-3-6	19.7[‡]	23.3[‡]	9.8	23.5	8.366	19.3	68.0	68.7	194.9	87.3	95.4
GEFR399-2-3-1-1-1	13.7[‡]	21.2[‡]	10.1	20.9	8.303	18.7	67.6	68.5	196.9	91.6	95.4
GEFR399-2-3-1-2-4-1	18.7[‡]	20.6[‡]	11.5	24.0	9.262	19.6	68.7	69.5	200.3	95.2	95.4
GEFR399-2-3-1-2-6-2	16.0[‡]	21.4[‡]	9.1[‡]	20.7	8.723	19.3	69.7	70.0	199.9	94.6	95.3
GEFR399-3-7-4-2-4-1	18.4[‡]	24.8[‡]	10.3	22.9	9.319	18.9	68.4	68.8	200.2	97.0	95.4
GEFR400-9-5-2-3-1 (NC303)	**24.3**[‡]	**18.2**[‡]	**9.2**	**18.4**	**9.381**	**19.6**	**69.4**	**70.0**	**195.6**	**94.0**	**95.3**
GEFR400-9-5-2-3-8 (NC301)	**10.6**[‡]	**12.0**[‡]	**8.0**[‡]	**16.6**[‡]	**8.692**	**19.8**	**68.3**	**69.4**	**191.2**	**87.3**	**95.4**
Checks											
FR1064	43.5	48.7	15.0	29.2	9.093	18.7	68.3	69.2	201.9	96.3	95.4
GE440	1.7[‡]	10.9[‡]	3.7[‡]	11.6[‡]	7.613	20.3	71.9	72.8	235.2	128.9	92.4
Pioneer 31G66	-	-	9.5	15.4	9.570	18.3	69.4	70.2	204.2	86.5	95.2
DeKalb DK697	-	-	11.9	19.3	9.751	18.9	71.4	72.8	204.3	104.4	95.2
Mean of experimental lines	17.3	20.0	9.4	19.9	8.740	19.2	68.5	68.0	199.6	95.3	95.3
Average LSD (0.05)	-	-	-	-	0.753	0.5	1.8	2.2	7.8	6.2	2.6

[†]Backtransformed from least square means from the analysis of natural log of the trait.
[‡]Significantly different from the FR1064 check at $P = 0.05$ on the transformed scale. Least significant difference is not appropriate on original scale.

Table 2.2. Least square means for the twenty superior GEFR lines and checks for Fusarium ear rot resistance, fumonisin content, grain yield, and other agronomic traits as topcrosses to the tester LH283×LH287 (averaged across two environments for disease traits and three environments for agronomic traits).

	Fusarium ear rot	Fumonisin content	Grain yield	Grain moisture	Days to anthesis	Days to silking	Plant height	Ear height	Erect plants
Experimental lines	%[†]	ppm[†]	Mg ha^{-1}	%	d	d	cm	cm	%
GEFR397-8-2-2-1-1-1	15.3	27.5	8.039	18.4	70.4	70.5	196.2	93.5	93.7
GEFR397-8-2-2-1-2-1	10.5	15.0[‡]	7.506	17.7	71.5	71.1	188.3	90.0	95.3
GEFR397-8-2-2-1-3-1	15.4	21.6	6.992	17.7	71.6	71.0	194.5	96.0	95.0
GEFR397-8-2-2-1-3-2	16.6	26.0	6.948	17.3	69.5	69.1	193.2	94.5	95.0
GEFR397-8-2-2-1-5-2	16.1	24.4	7.381	17.9	70.8	70.9	194.4	92.0	94.5
GEFR399-1-5-1-3-2-1	12.5	20.7[‡]	6.522	17.7	68.2	68.0	184.5	82.9	94.9
GEFR399-1-5-1-3-2-2	9.6[‡]	15.7[‡]	6.998	18.2	68.3	68.1	185.4	85.4	92.9
GEFR399-1-5-1-3-4-2	11.5	21.0	7.776	17.2	69.4	68.4	184.2	77.8	93.4
GEFR399-2-10-4-1-2-2	12.1	23.4	6.647	17.7	69.4	68.5	176.5	74.7	95.3
GEFR399-2-10-6-2-1-1	7.9[‡]	15.7[‡]	7.111	18.2	68.0	68.2	179.4	79.0	95.0
GEFR399-2-10-6-2-1-2	11.5	23.6	7.657	18.0	68.0	68.1	180.6	75.5	95.3
GEFR399-2-10-6-2-3-2	10.7	30.3	6.641	17.6	69.2	69.3	185.7	80.2	95.3
GEFR399-2-10-6-2-5-1	15.3	28.8	7.350	18.0	69.5	69.5	179.9	77.3	93.6
GEFR399-2-10-6-3-6	9.5[‡]	26.9	6.735	18.0	68.6	69.2	175.9	75.9	94.0
GEFR399-2-3-1-1-1	12.4	36.7	7.663	18.0	69.2	68.8	187.3	79.5	95.3
GEFR399-2-3-1-2-4-1	14.2	31.2	7.394	17.9	69.6	69.7	181.7	80.4	94.2
GEFR399-2-3-1-2-6-2	10.3	18.2[‡]	8.039	18.0	68.4	68.3	186.0	82.1	92.9
GEFR399-3-7-4-2-4-1	16.4	30.5	7.155	17.3	69.5	69.3	178.7	84.8	94.4
GEFR400-9-5-2-3-1 (NC303)	**14.0**	**34.8**	**7.751**	**18.0**	**69.7**	**69.4**	**179.8**	**78.6**	**95.3**
GEFR400-9-5-2-3-8 (NC301)	**9.0**[‡]	**14.9**[‡]	**7.952**	**17.8**	**69.8**	**69.2**	**175.9**	**77.8**	**94.7**
Checks									
FR1064	16.3	36.4	6.384	17.0	69.7	69.2	182.4	81.1	94.8
GE440	7.9	8.0	7.763	19.1	70.7	69.8	213.4	112.0	90.9
Pioneer 31G66	9.5	15.4	9.570	18.3	69.4	70.2	204.2	86.5	95.2
Pioneer 31G98	36.6	50.1	9.206	17.7	71.8	72.6	211.1	103.9	95.3
DeKalb DK697	11.9	19.3	9.751	18.9	71.4	72.8	204.3	104.4	95.2
Mean of experimental lines	12.6	24.3	7.313	17.8	69.4	69.2	184.4	82.9	94.5
Average LSD (0.05)	-	-	0.753	0.5	1.8	2.2	7.8	6.2	2.6

[†]Backtransformed from least square means from the analysis of natural log of the trait.
[‡]Significantly different from the FR1064 check at $P = 0.05$ on the transformed scale. Least significant difference is not appropriate on original scale.

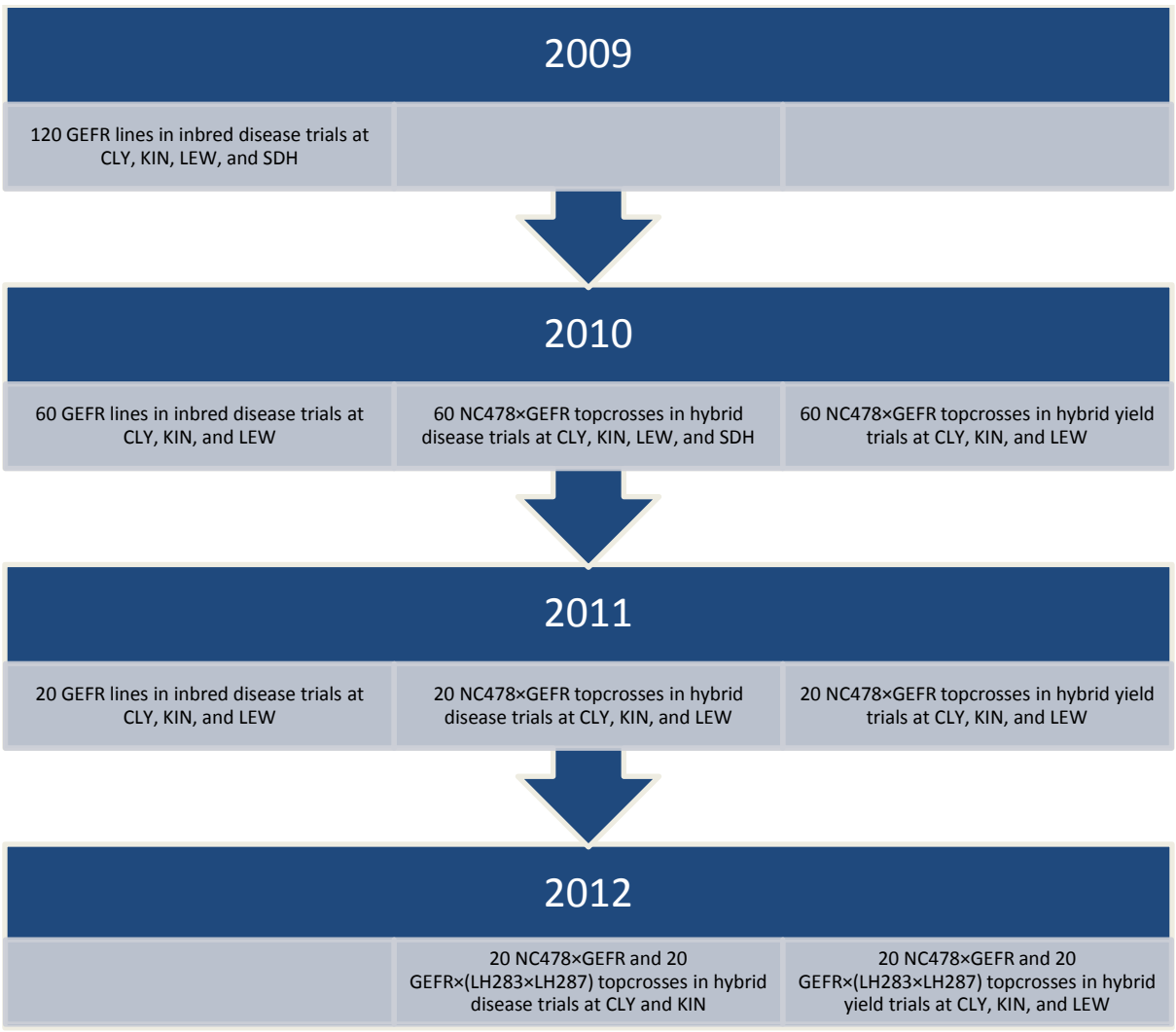


Figure 2.1. Flowchart showing the screening process for the GEFR backcross lines. The inbred disease evaluation pipeline is shown on the left, the hybrid disease evaluation pipeline in the middle, and the hybrid yield evaluation pipeline on the right. Location abbreviations are as follows: Clayton = CLY; Kinston = KIN; Lewiston-Woodville = LEW; and Jackson Springs = SDH.

CHAPTER 3: A Genome-Wide Association Study Reveals Genes Associated with Fusarium Ear Rot Resistance in a Maize Core Diversity Panel

Citation

Zila C.T., Samayoa L.F., Santiago R., Butrón A., Holland J.B., 2013 A genome-wide association study reveals genes associated with Fusarium ear rot resistance in a maize core diversity panel. G3: Genes, Genomes, Genetics **3**: 2095-2104.

A Genome-Wide Association Study Reveals Genes Associated with Fusarium Ear Rot Resistance in a Maize Core Diversity Panel

Charles T. Zila[*], L. Fernando Samayoa[§], Rogelio Santiago[§], Ana Butrón[§], and James B. Holland[*†1]

[*]Department of Crop Science, North Carolina State University, Raleigh, North Carolina 27695, [§]Misión Biológica de Galicia, CSIC**,** Pontevedra, Spain, 36080, and [†]U.S. Department of Agriculture—Agricultural Research Service, Plant Science Research Unit, Raleigh, North Carolina, 27695

GWAS for Fusarium Ear Rot Resistance in Maize

KEYWORDS: association analysis, disease resistance, genotype by environment interaction, maize, quantitative trait

[1] Corresponding author: USDA-ARS and Department of Crop Science, Campus Box 7620, Raleigh, NC, 27695-7616. Phone: (919) 513-4198. E-mail: james_holland@ncsu.edu.

Abstract

Fusarium ear rot is a common disease of maize that affects food and feed quality globally. Resistance to the disease is highly quantitative, and maize breeders have difficulty incorporating polygenic resistance alleles from unadapted donor sources into elite breeding populations without having a negative impact on agronomic performance. Identification of specific allele variants contributing to improved resistance may be useful to breeders by allowing selection of resistance alleles in coupling phase linkage with favorable agronomic characteristics. We report the results of a genome-wide association study (GWAS) to detect allele variants associated with increased resistance to Fusarium ear rot in a maize core diversity panel of 267 inbred lines evaluated in two sets of environments. We performed association tests with 47,445 SNPs while controlling for background genomic relationships with a mixed model and identified three marker loci significantly associated with disease resistance in at least one subset of environments. Each associated SNP locus had relatively small additive effects on disease resistance (±1.1% on a 0-100% scale), but nevertheless were associated with 3 to 12% of the genotypic variation within or across environment subsets. Two of three identified SNPs colocalized with genes that have been implicated with programmed cell death and were expressed at highest levels during the onset of disease symptoms. An analysis of associated allele frequencies within the major maize subpopulations revealed enrichment for resistance alleles in the tropical/subtropical and popcorn subpopulations compared to other temperate breeding pools.

Introduction

The hemibiotrophic fungus *Fusarium verticillioides* (Sacc) Nirenberg is endemic in most maize fields in the United States and is present in many arable regions of the world (VAN EGMOND *et al.* 2007). This fungus causes Fusarium ear rot disease of maize, especially in low rainfall high-humidity environments, such as the southern United States and some lowland tropics (MILLER and TRENHOLM 1994). Infection by *F. verticillioides* can result in decreased grain yields, poor grain quality, and contamination by the mycotoxin fumonisin, a suspected carcinogen associated with various diseases in livestock and humans (MILLER and TRENHOLM 1994; MARASAS 1996; PRESELLO *et al.* 2008).

The best strategy for controlling Fusarium ear rot and reducing the incidence of fumonisin contamination of grain is the development and deployment of maize hybrids with genetic resistance. Fusarium ear rot resistance is under polygenic control and strongly influenced by environmental factors; no fully immune genotypes have been discovered (KING and SCOTT 1981; NANKAM and PATAKY 1996; CLEMENTS *et al.* 2004). The complexity of this resistance trait has impeded breeding, such that most commercial maize hybrids have lower levels of resistance than are desirable (BUSH *et al.* 2004). Linkage-based mapping studies in biparental populations have shown that Fusarium ear rot resistance QTL have relatively small effects and are not consistent between populations (PÉREZ-BRITO *et al.* 2001; ROBERTSON-HOYT *et al.* 2006; DING *et al.* 2008; MESTERHÁZY *et al.* 2012).

Despite the genetic complexity of resistance to Fusarium ear rot and fumonisin accumulation, and despite the very low heritability of resistance measured on individual plants, resistance on the basis of family means from well-replicated studies is moderately to

highly heritable (ROBERTSON *et al.* 2006; ELLER *et al.* 2008; BOLDUAN *et al.* 2009). Robertson et al. (2006) and Bolduan et al. (2009) reported genotypic correlations between ear rot resistance and fumonisin accumulation of 0.87 in North Carolina and 0.92 in Germany, respectively, indicating that visual selection on Fusarium ear rot resistance should be effective in simultaneously reducing fumonisin contamination. The heritability estimates predict, and empirical selection studies demonstrate, that selection for improved ear rot resistance can be effective (ROBERTSON *et al.* 2006; BOLDUAN *et al.* 2009; ELLER *et al.* 2010). Unfortunately, most sources having high levels of ear rot resistance are older or exotic unadapted inbreds that lack the agronomic performance of modern elite maize lines (CLEMENTS *et al.* 2004; ELLER *et al.* 2008, 2010). Thus, breeders are faced with the difficulty of introducing polygenic resistance alleles of generally small effect linked to inferior polygenic alleles for agronomic performance if they attempt to incorporate improved genetic resistance from unadapted lines into elite breeding gene pools. Identification of specific allelic variants that confer improved resistance would permit maize breeders to select for rare recombinant chromosomes in backcross progeny that have desired target resistance allele sequences in coupling phase with the favorable elite polygenic background, facilitating the improvement of disease resistance without decreasing agronomic performance.

Resolving small effect QTL to causal genes for traits that are difficult to accurately measure phenotypically is exceedingly difficult in biparental mapping populations (HOLLAND 2007). Compared to traditional linkage-based analyses, association mapping offers higher mapping resolution while eliminating the time and cost associated with developing synthetic mapping populations (FLINT-GARCIA *et al.* 2005; YU and BUCKLER 2006). Historically, a

major limitation to association mapping in low linkage disequilibrium (LD) species such as maize has been the large number of genetic markers required to detect marker-trait associations. Limiting the search space to predetermined candidate genes allows for association mapping with a smaller number of markers but requires extensive knowledge of the biochemical pathway contributing to the trait of interest (REMINGTON and PURUGGANAN 2003). To date, nothing is known about the pathways contributing to Fusarium ear rot resistance in maize. However, the recent availability of the maize 50k SNP genotyping array (GANAL *et al.* 2011) has provided almost 50,000 single nucleotide polymorphism (SNP) markers scored on 279 of the 302 inbred lines of a commonly used maize core diversity panel (FLINT-GARCIA *et al.* 2005; COOK *et al.* 2012). The maize diversity panel captures much of the diversity present in public breeding programs worldwide. The large number of markers available on the diversity panel has enabled genome-wide association studies (GWAS) for several complex traits in maize including kernel composition traits (COOK *et al.* 2012) and the hypersensitive response (OLUKOLU *et al.* 2013). Olukolu et al. (2013) identified SNPs associated with the hypersensitive defense response in or adjacent to five genes not previously known *a priori* to affect disease resistance, but whose predicted gene functions all involved the programmed cell death pathway. In this study, we employed GWAS to identify SNPs associated with Fusarium ear rot resistance in the maize core diversity panel both within and across two contrasting environments – North Carolina, USA and Galicia, Spain.

Materials & Methods

Genotypes and experimental design

The maize core diversity panel (sometimes referred to as the "Goodman" association panel, because the seed stocks were originally assembled by Major Goodman at North Carolina State University (FLINT-GARCIA *et al.* 2005)) was evaluated in several years in both North Carolina, USA and Galicia, Spain. Only the 279 inbred lines with available genotypic data were considered in this study. In the Galicia experiment, a subset of 270 inbred lines from the maize diversity panel was evaluated for Fusarium ear rot resistance in a randomized 15 × 18 α-lattice block design with two replicates in 2010 and 2011. Nine lines with insufficient seed were dropped from the Galicia experiment before randomization. In the North Carolina experiment, the maize diversity panel was part of an evaluation of the entire USDA maize seed bank collection of inbred lines in 2010 (ROMAY *et al.* 2013) and subsets of that collection evaluated in 2011 and 2012. The genotypic data on the maize seed bank collection reported by Romay et al. (2013) were not available at the time of analysis. The 2010 seed bank collection evaluation included 2572 inbred line entries and was arranged in an augmented single replicate design. Experimental entries were divided into 18 sets of differing sizes based on maturity and field assignment. Each block within each set was augmented with a B73 check plot in a random assignment, and five other checks (IL14H, Ki11, P39, SA24, and Tx303) were included once per set in a random position. Romay et al. (2013) reported flowering time evaluations of the entire collection evaluated at three locations in 2010, including North Carolina. Here we include data only from North Carolina because it was the only environment used for Fusarium ear rot evaluation. In 2011 and 2012,

the maize core diversity panel was part of a larger sample of inbreds evaluated. The larger population consisted of 771 diverse entries divided into eight sets based on maturity and replicated across years. Although disease measurements were collected on all experimental entries in both years, genotypic data were not available on inbreds outside of the core diversity panel at the time of analysis. Sets were randomized within the field, and each set was blocked using an α-lattice design. As with the seed bank collection evaluation, each block was augmented by a randomly assigned B73 check plot, and five other checks (GE440, NC358, NK794, PHB47, and Tx303) were included once per set.

The three North Carolina environments were artificially inoculated with local toxigenic *Fusarium verticillioides* isolates using the toothpick method (CLEMENTS *et al.* 2003). Approximately one week after flowering, a toothpick containing *F. verticillioides* spores was inserted directly into the primary ear of five plants in each plot. At maturity, inoculated ears were harvested and visually scored for Fusarium ear rot symptoms. Scores were assigned to each ear in increments of 5% from 0% to 100% diseased based on the percentage of the ear presenting disease symptoms (ROBERTSON *et al.* 2006; Figure A.1). In Galicia, between seven and 14 days after flowering, five primary ears per plot were inoculated with 2 mL of a spore suspension of the local toxigenic isolate of *F. verticillioides.* The spore suspension contained 10^6 spores mL^{-1} and was prepared following the protocol established by Reid et al. (1996) with some modifications. Inoculum was injected into the center of the ear using a four-needle vaccinator which perforated the husks and injured three to four kernels. Ears from each plot were collected two months after inoculation and were individually rated for Fusarium ear rot symptoms using a seven-point scale (1=no visible

disease symptoms, 2=1-3%, 3=4-10%, 4=11-25%, 5=26-50%, 6=51-75%, and 7=76-100% of kernels exhibiting visual symptoms of infection, respectively) devised by Reid and Zhu (2005). Phenotypic data on the seven-point scale from the Galicia environments were transformed to the 0-100% scale used in North Carolina in the analyses. Reliable data could not be obtained for some line-environment combinations because seed set for some plots was limited due to poor adaptation. Raw data are provided in supplemental dataset File A.1. Climate data from on-farm weather stations were obtained from http://www.climate.ncsu.edu and http://www.mbg.csic.es/eng/index.php.

Genotypic data

The genotypic data were 47,445 SNPs from the Illumina maize 50k genotyping array filtered by Olukolu et al. (2013). The original array consists of 49,585 SNPs designed by Ganal et al. (2011). Olukolu et al. (2013) filtered the data set to include only those SNP markers that mapped to defined single locations in the maize genome and had <20% missing data (http://www.genetics.org/content/suppl/2012/12/05/genetics.112.147595.DC1/genetics.112.147595-3.txt).

Statistical analyses

Estimation of least square means and heritabilities

The Galicia and North Carolina experiments were analyzed separately and then combined in a single multi-environment analysis. Each year of data within each experiment

was first analyzed separately by fitting a mixed linear model including line as a fixed effect, silking date as a fixed linear covariate, and replication (Galicia only), block within replication (Galicia only), set (North Carolina only), and block within set (North Carolina only) as random effects. The mixed linear model for the Galicia experiment across years included line as a fixed effect, silking date as a fixed linear covariate, and year, line×year interaction, replication within year, and block within replication as random effects. The North Carolina experiment was analyzed across years with a model including line as a fixed effect, silking date as a fixed linear covariate, and year, line×year interaction, set within year, and block within set as random effects. In the combined experiment analysis, each combination of location and year was considered an environment. The combined analysis model included a fixed line effect, silking date as a fixed covariate nested within environment, a random line×environment interaction effect, and nested random experimental design effects (replication within environment and block within replication at Galicia and set within environment and block within set at North Carolina). All analyses were weighted by the number of ears scored within each plot and utilized a heterogeneous error variance structure. In both experiments, larger predicted ear rot values were associated with larger residuals, so a natural logarithmic transformation of raw ear rot scores (which largely eliminated the relationship between residual variance and predicted values) was used for all analyses. Least square means were estimated for 267 inbred lines within each experiment and across experiments (File A.2) using ASReml version 3 software (GILMOUR *et al.* 2009). Means for twelve lines were not estimable due to missing phenotypic observations in all environments (generally due to poor seed production).

We conducted a second analysis treating inbred lines as random effects for the purposes of estimating heritability for Fusarium ear rot resistance in the diversity panel. The same models as above were used except lines were treated as random effects to obtain estimates of genetic variance. Line mean-basis heritability was estimated as

$$\widehat{H}_c = 1 - \frac{\sigma^2_{PPE}}{2\hat{\sigma}^2_G}$$

where σ^2_{PPE} is the average prediction error variance for all pairwise comparisons of lines and $\hat{\sigma}^2_G$ is the estimated genetic variance (CULLIS *et al.* 2006). We estimated line mean-basis heritabilities for each environment individually, across the North Carolina environments, across the Galicia environments, and we also estimated line mean-basis heritability for the combined data set across all environments. The model used to estimate line mean-basis heritability in the combined data set was further modified by nesting the random line effect within environment and modeling the genotype-environment effect (**G**) matrix as unstructured, thereby allowing estimation of unique genetic variance within each environment and a unique genetic correlation between each pair of environments. For the purpose of estimating heritability, the average of the ten pair-wise covariance estimates between environments (which are expected to equal the genotypic variance) was used in the denominator of the above equation.

Silking date heritabilities were also calculated for each environment and across environments. The same models used to compute ear rot heritabilities were used to estimate silking date heritabilites, but silking date was treated as the dependent variable instead of as a fixed linear covariate.

Association analyses

A genetic kinship matrix (**K**; File A.3) based on observed allele frequencies (VANRADEN 2008; method 1) was created using R software version 3.0.0 (R CORE TEAM 2013). A subset of 4000 SNP markers were used to estimate **K**. Markers were uniformly distributed across the genome (at least 60 kbp between adjacent markers) and had no missing data after excluding heterozygous genotypes. Olukolu et al. (2013) used a kinship matrix produced by Tassel software (BRADBURY *et al.* 2007), which is appropriate for population structure correction for GWAS. In addition to population structure correction, we also wanted to estimate the polygenic background genetic variance component, so we estimated a new **K** matrix that is scaled appropriately to represent realized genomic average identity by descent relationships among the lines (VANRADEN 2008).

Tassel version 4.1.24 was used for the genome-wide association analyses based on a mixed linear model (BRADBURY *et al.* 2007). The least square means for inbred lines were used as the input phenotypes, and each set of means (North Carolina, Galicia and combined) was analyzed separately (File A.2). The mixed linear model implemented by Tassel was

$$y = X\beta + Zu + e$$

where **y** is the vector of ear rot least square means (on the natural-log scale), **β** is a vector of fixed effects including SNP marker effects, **u** is a vector of random additive genetic effects from background QTL for lines, **X** and **Z** are design matrices, and **e** is a vector of random residuals. The variance of the **u** vector was modeled as

$$\mathrm{Var}(u) = K\sigma_a^2$$

where **K** is the $n \times n$ matrix of pairwise kinship coefficients ranging 0—2 and σ_a^2 is the estimated additive genetic variance (YU *et al.* 2006).

Restricted maximum likelihood estimates of variance components were obtained using the optimum compression level and population parameters previously determined (P3D) options in Tassel (ZHANG *et al.* 2010). The optimum compression level option reduces the dimensionality of **K** by clustering *n* lines into *s* groups, thereby reducing computational time and potentially improving model fit. The *P*-values for each of the 47,445 tests of associations between one SNP and ear rot resistance within each analysis were used to estimate the false positive discovery rate (FDR) using the QVALUE version 1.0 package in R (STOREY and TIBSHIRANI 2003). SNPs significant at $q < 0.10$ in the initial GWAS scan for a particular environment set were then included together in a joint SNP association model together using the GLM procedure in SAS software version 9.2 (SAS INSTITUTE INC 2010) to estimate the total amount of variation explained by the SNPs together and to re-estimate their effects jointly. Candidate genes either containing or located adjacent to associated SNPs were identified using the MaizeGDB genome browser (ANDORF *et al.* 2010).

Allele frequency analysis

Lines were grouped into one of five major maize subpopulations (stiff stalk, non-stiff stalk, tropical/subtropical, popcorn, and sweet corn) based on the population structure analysis of the maize core diversity panel reported by Flint-Garcia et al. (2005; http://panzea.org/db/gateway?file_id=pop_structure_xls). Lines of mixed ancestry (the result of admixture among the subpopulations) were dropped from the analysis. Based on the

results of the association analyses, the frequencies of alleles that reduced disease severity at significant SNPs were estimated within each subpopulation using the FREQ procedure using SAS software version 9.2 (SAS INSTITUTE INC 2010). At each SNP locus, a Fisher's exact test in R software version 3.0.0 (R CORE TEAM 2013) was used to test the null hypothesis that frequency of the allele conferring increased disease resistance was the same across all five subpopulations.

Results

Line means and heritability

Significant ($P < 0.001$) genotypic variation for ear rot resistance was observed in both the North Carolina and Galicia experiments. The mean ear rot observed among 267 inbred lines of the association panel ranged from 4.4% to 100% with an overall mean of 41.1% in North Carolina and from 0% to 89.3% with an overall mean of 7.4% in Galicia (Table A.1; File A.2). In the combined analysis, mean ear rot ranged from 1.6% to 79.6% with an overall mean of 22.1%. The silking date covariate was highly heritable ($\widehat{H}_c = 0.98$ in the combined analysis) and was significantly associated with ear rot resistance in the North Carolina and combined analyses ($P < 0.001$), but not in the Galicia analysis ($P = 0.099$; Table A.1).

A significant ($P < 0.001$) line×environment interaction was detected in the combined analysis. Results of the mixed model analysis that estimate unique genotypic covariances for each pair of environments indicated that the two Galicia environments had a much stronger genotypic correlation ($r = 0.93$; Table 3.1 and Figure A.2) than did any other pair of environments (range, $r = 0.28$ to 0.51; Table 3.1 and Figure A.2). Thus, there was little

genotype×environment interaction between the two Galicia environments, and the heritability of line means across the two years in Galicia was 0.71. In contrast, pair-wise genotypic correlations were much lower among the North Carolina environments and between North Carolina and Galicia environments (Table 3.1 and Figure A.2), generating much of the observed genotype×environment interaction in the combined analysis. Despite the strong genotype×environment interaction among North Carolina environments, heritability of genotype means across the three years in North Carolina (0.73) was higher than within any single North Carolina environment (Table A.1). In addition, heritability of line means across all five environments was 0.75, higher than within any single environment or group of environments (Table A.1). Therefore, we conducted separate association analyses on three different sets of genotypic mean values for ear rot: (1) means from three North Carolina environments, (2) means from two Galicia environments, and (3) means from the combined analysis of all five environments.

Association mapping of Fusarium ear rot resistance

The optimum compression option in Tassel clustered the 267 lines into 229 groups in the Galicia analysis and 197 groups in the North Carolina and combined analyses (Table 3.2). Background genetic effects modeled by **K** accounted for 31% of the total variation among line means in the North Carolina analysis, 57% of the total phenotypic variation in the Galicia analysis, and 48% of the total phenotypic variation in the combined analysis (Table 3.2). In the analysis of means from North Carolina environments, two SNPs were identified as significantly associated with ear rot resistance at $q \leq 0.05$ (raw P-value = 2.4×10^{-7}), and

one additional SNP was identified at $q \leq 0.10$ (Table 3.3 and Figure 3.1). In the combined analysis, one SNP was identified as significantly associated with ear rot resistance at $q \leq 0.05$ and coincided with one of the SNPs identified in the North Carolina analysis. No SNPs significant at $q \leq 0.10$ were identified in the Galicia analysis, where the minimum raw *P*-value among SNP association tests was 2.1×10^{-4}.

Candidate genes colocalized with associated SNPs

Genes containing or nearby SNPs significantly associated with ear rot resistance were characterized using the filtered predicted gene set from the annotated B73 reference maize genome (SCHNABLE *et al.* 2009). Two of the three genes identified in the North Carolina analysis have predicted functions that have been implicated in disease response pathways in other plant species (TSUNEZUKA *et al.* 2005; HÉMATY *et al.* 2009). The SNP at physical position 151,295,233 bp on chromosome 9, which was identified in both the North Carolina and combined analyses, is located in an intronic region of a cellulose synthase-like family A/mannan synthase gene (Table 3.3). Mean LD r^2 between the chromosome 9 SNP and other SNPs dropped below 0.1 within approximately 100 kbp (Figure 3.2). The other two SNPs identified in the North Carolina analysis on chromosomes 1 and 5 were located inside of a gene of unknown function and nearby a heat-shock 60-kDa protein (HSP60), respectively. Mean LD r^2 between the chromosome 1 and chromosome 5 SNPs and other SNPs dropped below 0.1 within approximately 10 kbp and 100 kbp, respectively (Figure 3.2). Although the chromosome 1 and 9 SNPs were not significantly associated with ear rot resistance in Galicia, the allele effects at these loci were consistent between North Carolina and Galicia

(Table 3.3). However, the allele effect at the chromosome 5 SNP locus showed a change in direction between North Carolina (+1.149%, Table 3.3) and Galicia (-0.017%).

Allele frequencies at candidate genes

We estimated the allele frequency at the three SNPs significantly associated with ear rot resistance in five of the major maize subpopulations – stiff stalk temperate (SS), non-stiff stalk temperate (NSS), tropical/sub-tropical (TS), popcorn (PC), and sweet corn (SC) (FLINT-GARCIA *et al.* 2005). European flint types are poorly represented in this maize core diversity panel and thus were not considered. Popcorn and sweet corn types were considered in the analysis, but comparisons to either of these two subpopulations may be less reliable than comparisons to other subpopulations due to smaller sample size (Table 3.4). The allele that reduced disease severity at the chromosome 1 SNP locus is only present in the NSS and TS subpopulations but not at high enough frequencies to be considered significantly different from the other three subpopulations (P=0.15, Table 3.4). The allele with reduced disease severity at the chromosome 5 SNP locus is significantly (P=6.2×10^{-6}) over-represented in TS and PC lines relative to other temperate (SS, NSS, and SC) lines. At the chromosome 9 SNP locus, the allele associated with reduced disease severity is significantly (P=3.846×10^{-4}) overrepresented in PC lines compared to the other four subpopulations (Table 3.4). Averaging least square means from the combined analysis across members of each subpopulation, the SS, NSS, TS, PC, and SC subpopulations had average ear rot scores of 24.0%, 24.3%, 14.6%, 17.9%, and 46.5% respectively (Table 3.4).

Discussion

Heritability and false discovery rate estimation

The mean ear rot severity observed across experimental entries was 41.1% in North Carolina and 7.4% in Galicia (Table A.1). Mean ear rot in North Carolina 2012 was particularly high (55%; Table A.1). The very strong genotypic correlation between Galician environments (Table A.1 and Figure A.2) justified their grouping as one environmental set in the analysis. Genotypic effects were significantly correlated between each pair of North Carolina environments, but at much lower magnitude (Table 3.1 and Figure A.2). Genotypic values in North Carolina 2010 had slightly higher correlations with the genotypic values in Galicia than in other years of North Carolina (Table 3.1), so grouping the three North Carolina environments has little justification based on genotype-by-environment patterns. Nevertheless, this environment grouping has a natural interpretation in terms of geography and adaptation, and the heritability of line means across these environments was higher than any individual environment, such that analysis of the three years as a group simplified interpretation of results.

The relationship between the *F. verticillioides* isolates used in each location is unknown; as such, it is possible that differences in pathogen aggressiveness could have contributed to the disparity in mean ear rot values across environments. In addition, differences in inoculation methods, as well as variation in temperature and precipitation levels, may have allowed for more favorable disease development in North Carolina as compared to Galicia. Although precipitation levels varied across all five environments,

average daily temperatures (both pre- and post-flowering) were higher in all three North Carolina environments compared to the two Galicia environments (Table A.2).

Heritabilities observed across environments in this study ($\widehat{H}_c \geq 0.71$) are consistent with estimates from biparental populations (ROBERTSON *et al.* 2006) and a small sample of North American and European public inbred lines (BOLDUAN *et al.* 2009). These heritability estimates were obtained with a model that assumed each line is a random sample from the reference population of global maize inbreds, modeled by a genotypic variance-covariance structure equal to the genotypic variance component multiplied by an identity matrix. For the purpose of controlling population structure in association analysis, adjusted line means from the original model were then used as observations in a mixed model analysis that modeled the genotypic variance-covariance structure as proportional to the realized genomic relationship matrix, thus incorporating the different pairwise relationships among the lines. This mixed model was simplified by the compression method of ZHANG *et al.* (2010), which clusters lines according to genetic similarity and replaces the full pair-wise realized genomic relationship matrix with a reduced matrix of average relationships among the groups. The optimal level of clustering or compression is determined empirically based on model fit to the observed phenotypic data. A compressed relationship matrix can have better model fit than the original matrix when the empirically observed covariance relationships among lines follow the group relationship averages better than the individual pairwise relationships. Typically, this can happen when closely related lines are grouped and estimate of the group phenotypes and their relationships with other group phenotypes are improved. The optimal

compression level can vary among phenotypes for the same set of lines, as observed in this study.

Among environment groups, the proportion of phenotypic variance explained by background genetic effects (**K**) was much smaller in North Carolina (31%, Table 3.2) compared to Galicia (57%). Besides the small polygenic additive effects captured by the kinship matrix, rare allele variants (minor allele frequency < 0.05) with larger effects, as well as epistatic interactions, may explain some of the genotypic variation not captured by either **K** or the significantly associated SNPs (MANOLIO *et al.* 2009).

Analyzing the Galicia environments separately from the North Carolina environments revealed no significant SNPs, whereas the North Carolina analysis identified three SNPs significantly associated with Fusarium ear rot resistance (Table 3.3). Examination of the empirical distribution of *P*-values for the Galicia analysis revealed a slight skew toward higher *P*-values, whereas the North Carolina and combined analyses exhibited excesses of small *P*-values (Figure A.3, Figure A.4, and Figure A.5). The Storey and Tibshirani (2003) method used to compute the false discovery rate assumes that the distribution of *P*-value for truly null tests follows a flat distribution, such that if the observed proportion of very low *P*-values is lower than expected based on the flat distribution, the false discovery rate will be high even for the lowest *P*-values, as we observed in the Galicia analysis. Whereas a few significant SNPs were identified in the North Carolina and combined analyses at $q < 0.10$, no SNP had a *q*-value of less than 0.9 in the Galicia analysis ((Figure A.3, Figure A.4, and Figure A.5). The disparity between the two individual experiment analyses highlights the importance of conducting individual environment association analyses in the presence of

significant genotype by environment (G×E) interaction. It should be noted, however, that the appropriate threshold proportion of variation due to G×E interaction to warrant individual location analyses instead of an overall combined analysis is not clear.

One possible mode of G×E interaction is the relative increase or decrease of additive allelic effects among different loci between environments (FALCONER and MACKAY 1996). Comparison of the absolute value of the allele effect at each of the identified SNP loci between North Carolina and Galicia revealed that allele effects were larger in North Carolina across all three loci (Table 3.3), congruent with the higher mean ear rot values in North Carolina (Table A.1). The largest proportion of phenotypic variance explained by **K** was in Galicia (Table 3.2), and when combined with comparatively smaller allele effects, suggested that more loci may have contributed to ear rot resistance in Galicia than North Carolina, and on average each locus had a smaller additive effect on disease phenotype in Galicia. Collectively, these two points may explain the deficiency of SNPs significantly associated with ear rot resistance in the Galicia analysis.

Association analyses

Three SNPs significantly associated with ear rot resistance were identified in the North Carolina analysis (Table 3.3), and all localized to separate chromosomes. One of these three SNPs, located on chromosome 9, was also identified in the combined analysis. None of the three SNPs localized to any of the linkage map bins containing resistance QTL reported by Robertson et al. (2006) and Ding et al. (2008). However, the proportion of phenotypic variance explained by each SNP is consistent with individual QTL r^2 values reported by each

of the two aforementioned mapping studies. The chromosome 9 SNP explained the largest proportion of the variation in line mean values for ear rot resistance (R^2=11.5% in NC and R^2=9.6% in the combined analysis, Table 3.3), while the chromosome 1 and chromosome 5 SNPs explained 8.8% and 9.6% of the variation in line mean values for ear rot resistance in North Carolina, respectively. Modeling all three SNPs together collectively explained 26% of the line mean variation in ear rot resistance in North Carolina.

Although all three SNPs explained a relatively large portion of the total variation in line means, each SNP had a relatively small additive effect on ear rot resistance (±1.1 percentage points ear rot severity on the back-transformed scale, Table 3.3). Additive genetic variance estimates for each SNP was computed based on allele effects and frequencies (Table 3.3), and when scaled to the total line mean variance coincided with the SNP R^2 values computed by Tassel (Table 3.3). In every case, an increase in disease resistance (decrease in ear rot severity) was associated with the rare allele at each locus. Resistance alleles at the chromosome 1 and 5 SNP loci were overrepresented in the tropical subpopulation relative to the other temperate subpopulations (Table 3.3), consistent with enriched disease resistance observed in tropical maize for some foliar diseases of maize (WISSER *et al.* 2011; OLUKOLU *et al.* 2013) and the lower level of ear rot disease observed in tropical lines in this study.

Using the same association panel and marker set as this study, Olukolu et al. (2013) reported that LD in the maize core diversity panel is variable across chromosomes and subpopulations. The authors also reported that marker pairs separated by more than 10 kbp had r^2<0.1 on average, which is consistent with estimates of r^2<0.1 between marker pairs separated by 5-10 kbp on average in tropical subpopulations and 10-100 kbp on average in

temperature subpopulations (LU *et al.* 2011). Increased marker coverage, such as the genotype-by-sequencing (GBS) data (ELSHIRE *et al.* 2011) used in Romay et al. (2013), in conjunction with a larger association panel, may be able to uncover more SNPs in higher LD with ear rot resistance loci. Assuming an association panel of between 350 and 400 inbred lines, Van Inghelandt et al. (2011) indicated that as few as 4,000 markers would be necessary in a GWAS to detect individual QTL explaining greater than 10% of the total phenotypic variation for a complex trait within the stiff stalk subpopulation, whereas 65,000 markers would be required to detect QTL at the same threshold within European flint types. In a sample of 2,815 inbred lines from the National Plant Germplasm System (USA) representing the same heterotic groups described in this study, Romay et al. (2013) reported that the utilization of over 680,000 GBS markers was sufficient to detect most known candidate genes associated with flowering time in maize. Even so, polymorphisms that strongly associated with the lower LD tropical/subtropical subpopulation (such as *ZmCTT*) were more difficult to detect compared to polymorphisms that more frequently associated with higher LD temperate subpopulations (such as *Vgt1*). The results of Romay et al. (2013) indicate that although increased marker coverage and association panel size can improve the power of a GWAS, special care needs to be given to ensure that lower LD subpopulations, such as the tropical/subtropical subpopulation, are adequately represented in an association panel in order to capture rare allele variants associated with those subpopulations.

Candidate genes

We used the B73 maize genome reference sequence to identify genes that either included or were nearby SNPs significantly associated with ear rot resistance. The chromosome 9 gene (GRMZM2G178880) that was identified in both the North Carolina and combined analyses belongs to the cellulose synthase-like family A (*CslA*) protein family. Given that the associated SNP localized to an intron segment within this gene, it is likely that this SNP is in LD with the causal variant and not the causal variant itself. The expression of this gene is highest in the endosperm of the developing seed kernel between 20 and 24 days after flowering during the growing season (SEKHON *et al.* 2011; http://www.plexdb.org). Peak expression of this gene coincides with the initial onset of Fusarium ear rot symptoms, which occurs approximately 28 days after flowering (BUSH *et al.* 2004). Genes in the *CslA* family encode for non-cellulose polysaccharides (such as mannan polymers) that form part of the wall matrix in plant cells (DHUGGA 2005; LIEPMAN *et al.* 2005). In the model grass species *Brachypodium distachyon*, mannan polymers make up a significant portion of the seed endosperm (GUILLON *et al.* 2011). Dismantling of mannan-rich cell walls may play an important role in programmed cell death (PCD) in host-pathogen interactions (GADJEV *et al.* 2008; RODRÍGUEZ-GACIO *et al.* 2012). Although the interaction between *Fusarium verticillioides* and maize is complex, cell wall structure and PCD may play a role in quantitative resistant to the disease (CHIVASA *et al.* 2005).

The SNP on chromosome 5 is located downstream of an HSP60 gene (GRMZM2G111477). Expression levels of this gene are highest in the developing endosperm 12 days after flowering (SEKHON *et al.* 2011; http://www.plexdb.org). HSP60s

are chaperonins that are involved with protein folding under plant stress primarily in the mitochondria and cholorplasts (WANG *et al.* 2004). The role of HSP60s in programmed cell death has been demonstrated in mutants of both rice and Arabidopsis (ISHIKAWA *et al.* 2003; TSUNEZUKA *et al.* 2005). The SNP on chromosome 1 is contained within the coding region of GRMZM2G703598. Unfortunately, this gene has no predicted function and has no sequence orthology with related grass species.

In conclusion, we have utilized a GWAS approach to identify three novel loci associated with improved resistance to Fusarium ear rot in maize. The identified loci each explain a relatively small proportion of the overall phenotypic variance for ear rot, and each locus has a very small additive genetic effect on resistance, consistent with the highly quantitative nature of the *F. verticillioides*-maize pathosystem. The large amount of variation captured by the kinship matrix, in combination with high false discovery rates, suggests that additive polygenic variation across many loci underlies resistance to Fusarium ear rot. Given the rapid decay of LD along the chromosomes in the maize core diversity panel (OLUKOLU *et al.* 2013), future studies employing increased marker density and larger association panels may be able to identify other novel loci associated with ear rot resistance. Maize breeders can employ targeted allele selection for these three resistance alleles, but may need to also select for recombinations near them as they are introgressed into elite maize from unadapted or undesirable genotypes (such as the tropical maize or popcorn germplasm pools that appear to be enriched for resistance alleles). In addition, given the substantial additive polygenic variation for ear rot resistance, phenotypic and genomic selection approaches should be

effective as long as high quality phenotypic evaluations of resistance can be performed to permit direct selection or provide training data for genomic selection models.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation (projects DBI-0321467, IOS-0820619 and IOS-0604923), USDA-ARS, the National Plan for Research and Development of Spain (AGL2009-12770), and the Excma. Diputación Provincial de Pontevedra. USDA-ARS supported the graduate fellowship of C.T. Zila. R. Santiago acknowledges postdoctoral contract "Isidro Parga Pondal" supported by the Autonomous Government of Galicia and the European Social Fund. The authors wish to thank Jason Brewer, Steve Pigozzo, David Horne, Brittany Scott, Sarah Davidson-Dyer, and Colt Jackson for technical assistance and Bode Olukolu, Funda Ogut, and Shang Xue for assistance with the analyses.

References

Andorf C. M., Lawrence C. J., Harper L. C., Schaeffer M. L., Campbell D. A., *et al.*, 2010 The Locus Lookup tool at MaizeGDB: identification of genomic regions in maize by integrating sequence information with physical and genetic maps. Bioinformatics **26**: 434–436.

Bernardo R., 2002 Chapter 5: Phenotypic and Genetic Variances. In: *Breeding for Quantitative Traits in Plants*, Stemma Press, Woodbury, MN, pp. 89–116.

Bolduan C., Miedaner T., Schipprack W., Dhillon B. S., Melchinger A. E., 2009 Genetic variation for resistance to ear rots and mycotoxins contamination in early European maize inbred lines. Crop Sci. **49**: 2019–2028.

Bradbury P. J., Zhang Z., Kroon D. E., Casstevens T. M., Ramdoss Y., *et al.*, 2007 TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics **23**: 2633–2635.

Bush B. J., Carson M. L., Cubeta M. A., Hagler W. M., Payne G. A., 2004 Infection and fumonisin production by *Fusarium verticillioides* in developing maize kernels. Phytopathology **94**: 88–93.

Chivasa S., Simon W. J., Yu X.-L., Yalpani N., Slabas A. R., 2005 Pathogen elicitor-induced changes in the maize extracellular matrix proteome. Proteomics **5**: 4894–4904.

Clements M. J., Kleinschmidt C. E., Maragos C. M., Pataky J. K., White D. G., 2003 Evaluation of inoculation techniques for Fusarium ear rot and fumonisin contamination of corn. Plant Dis. **87**: 147–153.

Clements M. J., Maragos C. M., Pataky J. K., White D. G., 2004 Sources of resistance to fumonisin accumulation in grain and Fusarium ear and kernel rot of corn. Phytopathology **94**: 251–260.

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

Cullis B., Smith A., Coombes N., 2006 On the design of early generation variety trials with correlated data. J. Agric. Biol. Environ. Stat. **11**: 381–393.

Dhugga K. S., 2005 Plant Golgi cell wall synthesis: From genes to enzyme activities. Proc. Natl. Acad. Sci. U. S. A. **102**: 1815–1816.

Ding J.-Q., Wang X.-M., Chander S., Yan J.-B., Li J.-S., 2008 QTL mapping of resistance to Fusarium ear rot using a RIL population in maize. Mol. Breed. **22**: 395–403.

Eller M. S., Holland J. B., Payne G. A., 2008 Breeding for improved resistance to fumonisin contamination in maize. Toxin Rev. **27**: 371–389.

Eller M. S., Payne G. A., Holland J. B., 2010 Selection for reduced Fusarium ear rot and fumonisin content in advanced backcross maize lines and their topcross hybrids. Crop Sci. **50**: 2249–2260.

Elshire R. J., Glaubitz J. C., Sun Q., Poland J. A., Kawamoto K., *et al.*, 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE **6**: e19379.

Falconer D. S., Mackay T. F. C., 1996 Chapter 19: Correlated Characters. In: *Introduction to Quantitative Genetics*, Longman, Essex, UK, pp. 312–334.

Flint-Garcia S. A., Thuillet A., Yu J., Pressoir G., Romero S. M., *et al.*, 2005 Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J. **44**: 1054–1064.

Gadjev I., Stone J. M., Gechev T. S., 2008 Programmed cell death in plants: New insights into redox regulation and the role of hydrogen peroxide. Int. Rev. Cell Mol. Biol. **270**: 87–144.

Ganal M. W., Durstewitz G., Polley A., Bérard A., Buckler E. S., *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.

Gilmour A. R., Gogel B. J., Cullis B. R., Thompson R., 2009 *ASReml User Guide Release 3.0*. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Guillon F., Bouchet B., Jamme F., Robert P., Quéméner B., *et al.*, 2011 *Brachypodium distachyon* grain: characterization of endosperm cell walls. J. Exp. Bot. **62**: 1001–1015.

Hématy K., Cherk C., Somerville S., 2009 Host–pathogen warfare at the plant cell wall. Curr. Opin. Plant Biol. **12**: 406–413.

Holland J. B., 2007 Genetic architecture of complex traits in plants. Curr. Opin. Plant Biol. **10**: 156–161.

Ishikawa A., Tanaka H., Nakai M., Asahi T., 2003 Deletion of a chaperonin 60β gene leads to cell death in the Arabidopsis *lesion initiation 1* mutant. Plant Cell Physiol. **44**: 255–261.

King S. B., Scott G. E., 1981 Genotypic differences in maize to kernel infection by *Fusarium moniliforme*. Phytopathology **71**: 1245–1247.

Liepman A. H., Wilkerson C. G., Keegstra K., 2005 Expression of cellulose synthase-like (*Csl*) genes in insect cells reveals that *CslA* family members encode mannan synthases. Proc. Natl. Acad. Sci. U. S. A. **102**: 2221–2226.

Lu Y., Shah T., Hao Z., Taba S., Zhang S., *et al.*, 2011 Comparative SNP and haplotype analysis reveals a higher genetic diversity and rapider LD decay in tropical than temperate germplasm in maize. PLoS ONE **6**: e24861.

Manolio T. A., Collins F. S., Cox N. J., Goldstein D. B., Hindorff L. A., *et al.*, 2009 Finding the missing heritability of complex diseases. Nature **461**: 747–753.

Marasas W. F., 1996 Fumonisins: history, world-wide occurrence and impact. Adv. Exp. Med. Biol. **392**: 1–17.

Mesterházy Á., Lemmens M., Reid L. M., 2012 Breeding for resistance to ear rots caused by *Fusarium* spp. in maize – a review. Plant Breed. **131**: 1–19.

Miller J. D., Trenholm H. L. (Eds.), 1994 The biosynthetic diversity of secondary metabolites; Epidemiology of Fusarium ear diseases of cereals. In: *Mycotoxins in Grain: Compounds Other Than Aflatoxin*, Eagan Press, St. Paul, pp. 3–36.

Nankam C., Pataky J. K., 1996 Resistance to kernel infection by *Fusarium moniliforme* in the sweet corn inbred IL 125b. Plant Dis. **80**: 593–598.

Olukolu B. A., Negeri A., Dhawan R., Venkata B. P., Sharma P., *et al.*, 2013 A connected set of genes associated with programmed cell death implicated in controlling the hypersensitive response in maize. Genetics **193**: 609–620.

Pérez-Brito D., Jeffers D., González-de-León D., Khairallah M., Cortés-Cruz M., *et al.*, 2001 QTL mapping of *Fusarium moniliforme* ear rot resistance in highland maize, Mexico. Agrociencia **35**: 181–196.

Presello D. A., Botta G., Iglesias J., Eyhérabide G. H., 2008 Effect of disease severity on yield and grain fumonisin concentration of maize hybrids inoculated with *Fusarium verticillioides*. Crop Prot. **27**: 572–576.

R Core Team, 2013 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reid L. M., Hamilton R. I., Mather D. E., 1996 *Screening Maize for Resistance to Gibberella Ear Rot*. Agriculture and Agri-Food Canada, Ottawa, ON, Canada.

Reid L. M., Zhu X., 2005 *Screening Corn for Resistance to Common Diseases in Canada*. Agriculture and Agri-Food Canada, Ottawa, ON, Canada.

Remington D. L., Purugganan M. D., 2003 Candidate genes, quantitative trait loci, and functional trait evolution in plants. Int. J. Plant Sci. **164**: S7–S20.

Robertson L. A., Kleinschmidt C. E., White D. G., Payne G. A., Maragos C. M., *et al.*, 2006 Heritabilities and correlations of Fusarium ear rot resistance and fumonisin contamination resistance in two maize populations. Crop Sci. **46**: 353–361.

Robertson-Hoyt L. A., Jines M. P., Balint-Kurti P. J., Kleinschmidt C. E., White D. G., *et al.*, 2006 QTL mapping for Fusarium ear rot and fumonisin contamination resistance in two maize populations. Crop Sci. **46**: 1734–1743.

Rodríguez-Gacio M. del C., Iglesias-Fernández R., Carbonero P., Matilla Á. J., 2012 Softening-up mannan-rich cell walls. J. Exp. Bot. **63**: 3976–3988.

Romay M. C., Millard M. J., Glaubitz J. C., Peiffer J. A., Swarts K. L., *et al.*, 2013 Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. **14**: R55.

SAS Institute Inc, 2010 *SAS® 9.2 Intelligence Platform: System Administration Guide*. SAS Institute Inc, Cary, NC.

Schnable P. S., Ware D., Fulton R. S., Stein J. C., Wei F., *et al.*, 2009 The B73 maize genome: complexity, diversity, and dynamics. Science **326**: 1112–1115.

Sekhon R. S., Lin H., Childs K. L., Hansey C. N., Buell C. R., *et al.*, 2011 Genome-wide atlas of transcription during maize development. Plant J. **66**: 553–563.

Storey J. D., Tibshirani R., 2003 Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. **100**: 9440–9445.

Tsunezuka H., Fujiwara M., Kawasaki T., Shimamoto K., 2005 Proteome analysis of programmed cell death and defense signaling using the rice lesion mimic mutant *cdr2*. Mol. Plant. Microbe Interact. **18**: 52–59.

van Egmond H. P., Schothorst R. C., Jonker M. A., 2007 Regulations relating to mycotoxins in food; Perspectives in a global and European context. Anal. Bioanal. Chem. **389**: 147–157.

Van Inghelandt D., Reif J. C., Dhillon B. S., Flament P., Melchinger A. E., 2011 Extent and genome-wide distribution of linkage disequilibrium in commercial maize germplasm. Theor. Appl. Genet. **123**: 11–20.

VanRaden P. M., 2008 Efficient methods to compute genomic predictions. J. Dairy Sci. **91**: 4414–4423.

Wang W., Vinocur B., Shoseyov O., Altman A., 2004 Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends Plant Sci. **9**: 244–252.

Wisser R. J., Kolkman J. M., Patzoldt M. E., Holland J. B., Yu J., *et al.*, 2011 Multivariate analysis of maize disease resistances suggests a pleiotropic genetic basis and implicates a *GST* gene. Proc. Natl. Acad. Sci. **108**: 7339–7344.

Yu J., Buckler E. S., 2006 Genetic association mapping and genome organization of maize. Curr. Opin. Biotechnol. **17**: 155–160.

Yu J., Pressoir G., Briggs W. H., Vroh Bi I., Yamasaki M., *et al.*, 2006 A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. **38**: 203–208.

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

Table 3.1. Genotypic covariance/variance/correlation matrix for Fusarium ear rot from the combined analysis of a maize diversity panel evaluated in five environments. The diagonal (bold) is an estimate of genetic variance ($\hat{\sigma}_G^2$) plus the genotype by environment interaction ($\hat{\sigma}_{GE}^2$) within each environment. Estimates of genetic variance (covariance between pairs of environments) are shown below the diagonal, and genetic correlations between inbred lines in each pair of environments are shown above the diagonal.

Environment	NC 2010	NC 2011	NC 2012	Galicia 2010	Galicia 2011
NC 2010	**0.27**	0.42	0.44	0.51	0.44
NC 2011	0.15	**0.45**	0.38	0.33	0.28
NC 2012	0.19	0.21	**0.68**	0.36	0.35
Galicia 2010	0.15	0.12	0.17	**0.32**	0.93
Galicia 2011	0.11	0.09	0.14	0.25	**0.23**

Table 3.2. Number of lines, number of groups, compression level, polygenic additive background genetic variance component, residual genotypic variance component, and proportion of total line mean variance explained by additive relationship matrix from the three mixed-linear model (MLM) analyses.

	N[a]	Groups[b]	Compression[c]	$(\hat{\sigma}_G^2)^d$	$(\hat{\sigma}^2)^d$	$\left(\frac{\hat{\sigma}_G^2}{\hat{\sigma}_G^2+\hat{\sigma}^2}\right)^e$
North Carolina	247	197	1.25	0.09	0.20	0.31
Galicia	254	229	1.11	0.18	0.14	0.57
Combined	267	197	1.36	0.10	0.11	0.48

[a] Total number of lines included in the analysis.
[b] Number of groups determined by optimum compression.
[c] Compression level is the average number of individuals per group.
[d] Polygenic additive background genetic variance and residual genotypic variance components are estimated in Tassel by fitting the kinship matrix (**K**) in the mixed linear model without any SNP marker effects.
[e] Background genetic variance divided by total phenotypic variance.

Table 3.3. Chromosome locations (AGP v2 coordinates), allele effect estimates, genes containing or adjacent to SNP, and other summary statistics for the three SNPs significantly associated with Fusarium ear rot resistance in the North Carolina analysis and the single SNP associated with resistance in the combined analysis. Statistics from environments in which the SNPs were not significantly associated with ear rot are also shown for comparison.

Chromosome	SNP physical position (bp)	*P*-value	*q*-value	Allele	N[a]	Allele effect (%)[b]	Additive variance estimate[c]	$(R^2)^d$	Gene containing or adjacent to SNP
North Carolina analysis									
1	63,540,590	5.5×10^{-6}	0.084	A	224	+0.945	0.036	8.8	GRMZM2G703598
				G	22	0.0			
5	30,997,717	2.2×10^{-6}	0.050	G	225	+1.149	0.042	9.6	GRMZM2G111477
				A	19	0.0			
9	151,295,233	2.4×10^{-7}	0.011	A	67	-0.365	0.041	11.5	GRMZM2G178880
				G	176	0.0			
Galicia analysis									
1	63,540,590	0.826^{NS}	1.000	A	231	+0.035	9.55×10^{-5}	1.9×10^{-2}	GRMZM2G703598
				G	22	0.0			
5	30,997,717	0.918^{NS}	1.000	G	228	-0.017	2.49×10^{-5}	4.2×10^{-3}	GRMZM2G111477
				A	23	0.0			
9	151,295,233	0.198^{NS}	1.000	A	71	-0.115	0.003	0.7	GRMZM2G178880
				G	179	0.0			
Combined analysis									
1	63,540,590	4.5×10^{-3}	0.689	A	244	+0.425	0.010	3.1	GRMZM2G703598
				G	22	0.0			
5	30,997,717	2.6×10^{-3}	0.689	G	240	+0.428	0.011	3.5	GRMZM2G111477
				A	24	0.0			
9	151,295,233	9.1×10^{-7}	0.042	A	74	-0.292	0.024	9.6	GRMZM2G178880
				G	189	0.0			

[a] N, total number of lines with the specific SNP genotype.
[b] Allele effects are reported back-transformed to the original 0-100% disease severity scale.
[c] Additive variance for an inbred population was computed as two times the product of the separate allele frequencies times the genotypic value from Tassel squared using the formula $2pqa^2$ from Bernardo (2002).
[d] R^2, proportion of total line mean variance explained by SNP as computed by Tassel.

Table 3.4. Allele frequencies of significantly associated SNPs in the five major maize subpopulations.

Chromo-some	SNP physical position (bp)	Allele increasing resistance	Allele frequency (%)					*P*-value	N[b]					Ear rot mean (%)[c]				
			SS[a]	NSS	TS	PC	SC		SS	NSS	TS	PC	SC	SS	NSS	TS	PC	SC
1	63,540,590	G	0.0	8.4	15.4	0.0	0.0	0.1488	28	107	65	8	6	24.0	24.3	14.6	17.9	46.5
5	30,997,717	A	0.0	3.8	26.6	37.5	0.0	6.193×10^{-6}	28	106	64	8	6					
9	151,295,233	A	14.3	34.9	26.6	100.0	33.3	3.846×10^{-4}	28	106	64	7	6					

[a] SS, Stiff Stalk; NSS, non-Stiff Stalk; TS, tropical/sub-tropical; PC, popcorn; SC, sweet corn.

[b] N, total number of lines within each subpopulation.

[c] Overall phenotypic ear rot means are the average of least square means from the combined analysis across members of each subpopulation.

Figure 3.1. Results of the three GWAS showing significant associations (points above red FDR = 0.10 threshold lines) in the North Carolina (A), Galicia (B), and combined (C) analyses. The vertical axis indicates $-\log_{10}$ of *P*-value scores, and the horizontal axis indicates chromosomes and physical positions of SNPs.

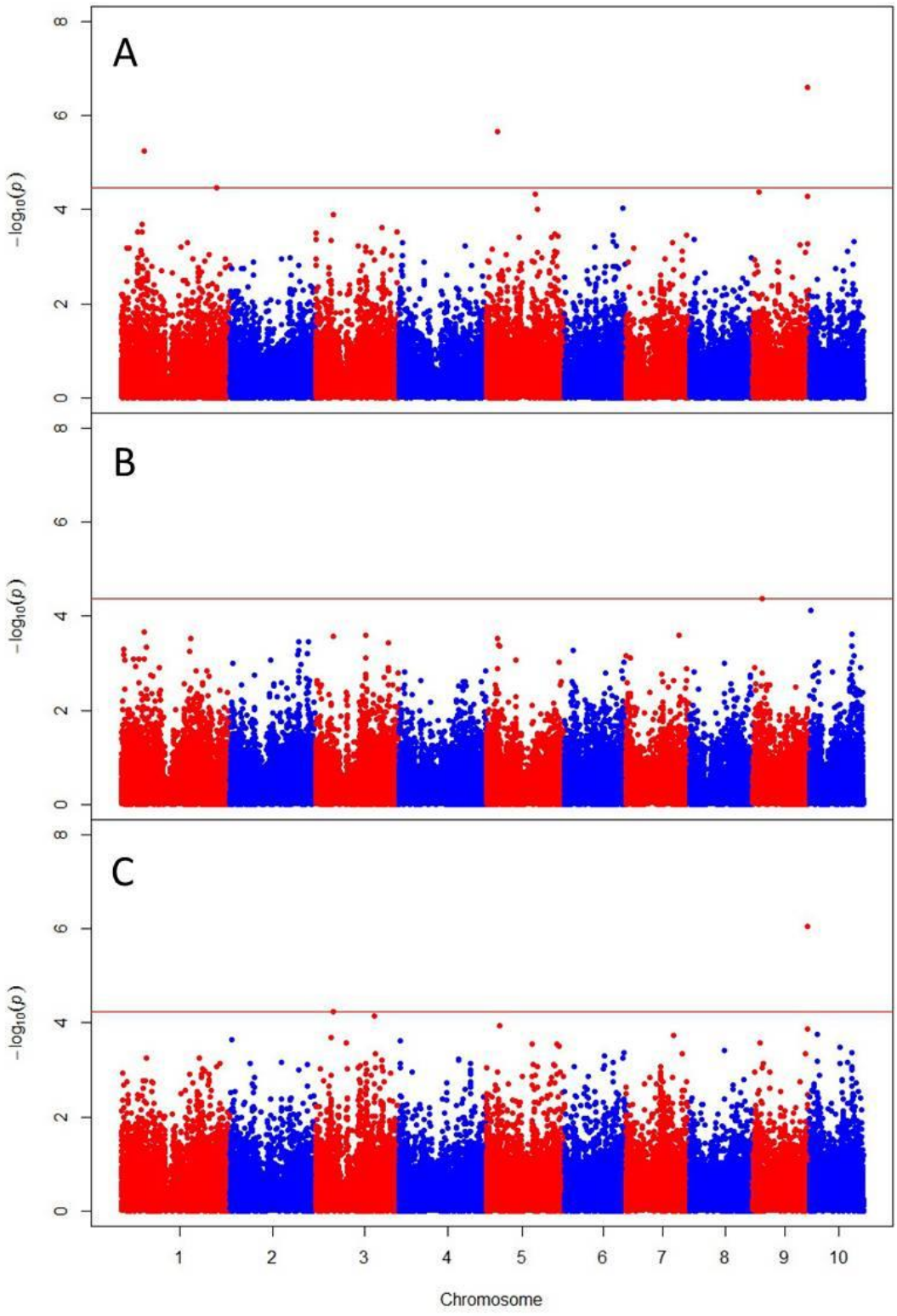
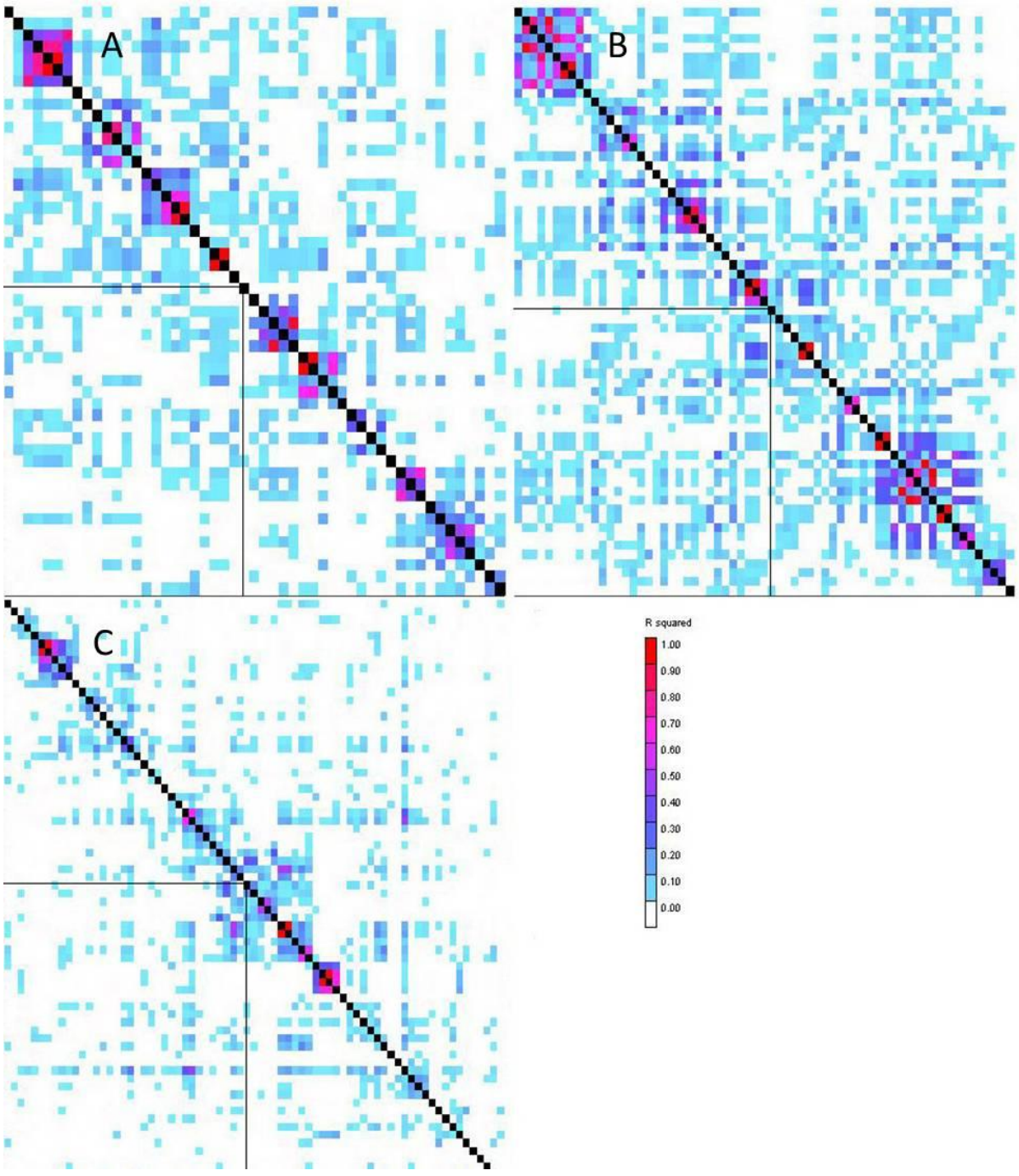

A
B
C
$-\log_{10}(p)$
0
2
4
6
8
1
2
3
4
5
6
7
8
9
10
Chromosome

Figure 3.2. LD heatmaps showing LD measure (r^2) calculated for each pairwise combination of SNPs in an approximately ±1 Mbp region surrounding each SNP significantly associated with ear rot resistance in the North Carolina analysis. (A) LD around chromosome 1 SNP. (B) LD around chromosome 5 SNP. (C) LD around chromosome 9 SNP. The significant SNP on each chromosome is highlighted by the perpendicular black lines within each heatmap. Colors indicate the magnitude of each pairwise r^2 measure (r^2=1 is red to r^2=0 is white).


A
B
C
R squared
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00

CHAPTER 4: Integration of Genome-Wide Association and Genomic Selection Methods for Genetic Prediction of a Complex Disease Resistance Trait in Maize

Prepared for submission to Genetics

Charles T. Zila[*], Funda Ogut[*], Maria C. Romay[†], Candice. A. Gardner[‡], Edward S. Buckler[†,§,**], and James B. Holland[*,††1]

[*]Department of Crop Science, North Carolina State University, Raleigh, North Carolina 27695

[†]Institute for Genomic Diversity, Biotechnology bldg., Cornell University, Ithaca, NY, 14853, USA

[‡]U.S. Department of Agriculture—Agricultural Research Service, North Central Regional Plant Introduction Station, Ames, IA, 50014, USA

[§]U.S. Department of Agriculture—Agricultural Research Service, Plant, Soil, and Nutrition Research Unit, Ithaca, NY, 14853, USA

[**]Department of Plant Breeding and Genetics, Bradfield Hall, Cornell University, Ithaca, NY, 14853, USA

GWAS and GS for Fusarium Ear Rot Resistance in Maize

KEYWORDS: association analysis, disease resistance, genomic selection, maize, quantitative trait

[1]Corresponding author: USDA-ARS and Department of Crop Science, Campus Box 7620, Raleigh, NC, 27695-7616. Phone: (919) 513-4198. E-mail: james_holland@ncsu.edu.

Abstract

Resistance to Fusarium ear rot of maize is a highly quantitative and complex trait. Marker-trait associations to date have had small additive effects and were inconsistent between previous studies, likely due to the combined effects of genetic heterogeneity and low power of detection of many small effect variants. The complexity of inheritance of resistance hinders the use marker-assisted selection for ear rot resistance. Genomic selection (GS) offers an alternative approach that is optimal for predicting values of traits with highly polygenic additive architecture. However, GS methods (even those that model variable marker effects) are not geared to identify and estimate effects of specific variants, or to select specific favorable variants with larger effects that are in repulsion phase disequilibrium with desired genomic backgrounds. We conducted a genome-wide association study (GWAS) for Fusarium ear rot resistance in a panel of 1689 diverse inbred lines from the USDA maize gene bank with 200,978 SNPs while controlling for background genetic relationships with a mixed model and identified seven SNPs associated with disease resistance. The three most significantly associated SNPs were then included as linear fixed effects in a genomic best linear unbiased prediction (G-BLUP) model and compared to a traditional G-BLUP model using cross-validation to see if the inclusion of GWAS associations along with a GS polygenic background model can improve predictive ability across a diverse panel with low (21%) heritability. Although the three SNPs had small effects, their inclusion increased G-BLUP prediction accuracies from 11% to 15%.

Introduction

Fusarium ear rot disease of maize, caused by the fungus *Fusarium verticillioides* (Sacc) Nirenberg, is endemic to maize production systems in the United States and worldwide (VAN EGMOND *et al.* 2007). The fungus is present as a symptomless endophyte in most maize seed lots (WARREN and KOMMEDAHL 1973; LESLIE *et al.* 1990; YATES *et al.* 1997); pathogenic colonization of developing maize kernels is common in the low rainfall high-humidity climates of the southern United States and lowland tropics (MILLER 1994). Infection by *F. verticillioides* can result in decreased grain yield, reduced grain quality, and grain contamination by the mycotoxin fumonisin. Fumonisin is a suspected carcinogen and is associated with various diseases in livestock and humans (MILLER 1994; MARASAS 1996; PRESELLO *et al.* 2008). In areas of the world where maize is a dietary staple and occurrence of Fusarium ear rot infection is high (such as sub-Saharan Africa), consumption of infected grain has been linked to esophageal cancer in adults and growth retardation in children (RHEEDER *et al.* 1992; KIMANYA *et al.* 2010; SHIRIMA *et al.* 2013).

The most effective method for controlling Fusarium ear rot infection and reducing fumonisin contamination is through the deployment of maize hybrids possessing genetic resistance. Resistance to the disease is under polygenic control, and no fully immune genotypes have been discovered (KING and SCOTT 1981; NANKAM and PATAKY 1996; CLEMENTS *et al.* 2004). Previous linkage-based and association mapping studies have shown that resistance QTL have relatively small effects and are not consistent between populations (PÉREZ-BRITO *et al.* 2001; ROBERTSON-HOYT *et al.* 2006; DING *et al.* 2008; MESTERHÁZY *et al.* 2012; ZILA *et al.* 2013). The complex nature of resistance has made it difficult for maize

breeders to effectively incorporate novel resistance alleles into adapted breeding pools; as a result, most commercial maize hybrids have lower levels of resistance than desired (BUSH *et al.* 2004). Although the heritability of resistance to Fusarium ear rot and fumonisin contamination is low on a per plot-basis, resistance on an entry mean-basis from replicated bi-parental and diversity panel studies is moderately to highly heritable (ROBERTSON *et al.* 2006; ELLER *et al.* 2008; BOLDUAN *et al.* 2009; ZILA *et al.* 2013). Empirical studies demonstrate that phenotypic selection for improved ear rot resistance can be effective (ROBERTSON *et al.* 2006; BOLDUAN *et al.* 2009; ELLER *et al.* 2010). However, most novel sources of disease resistance are unadapted inbreds with poor agronomic performance that often come from tropical or other exotic germplasm pools (CLEMENTS *et al.* 2004; ZILA *et al.* 2013).

Genome-wide association studies (GWAS) can be a powerful tool in the identification of specific allele variants that confer improved resistance to various diseases in maize. Utilizing a maize core diversity panel of 279 public inbred lines (FLINT-GARCIA *et al.* 2005) and over 47,000 SNPs from the Illumina maize 50k array (GANAL *et al.* 2011), Zila et al. (2013) identified three genes associated with improved resistance to Fusarium ear rot. However, the three loci associated with improved ear rot resistance all had small allelic effects (±1.1% on a percentage ear rot scale), and each individual locus was associated between 3 to 12% of the observed variation in line means after accounting for the additive polygenic background genetic variance captured by the genomic kinship matrix. In contrast, the genomic kinship matrix itself explained nearly half of the observed variation in line means across all analyses (ZILA *et al.* 2013). The substantial additive polygenic variance

observed suggests that genomic selection (GS) models could be an effective tool in improving resistance to Fusarium ear rot in maize breeding pools as opposed to selecting upon individual allele variants identified through GWAS.

In genomic selection, prediction models aim to estimate breeding values for untested selection candidates based on their genomic information (BERNARDO and YU 2007). A variety of statistical methodologies have been proposed for GS, including genomic best linear unbiased prediction (G-BLUP; VANRADEN 2008), ridge regression BLUP (RR-BLUP; PIEPHO 2009), and various Bayesian regression models (GIANOLA 2013). Although both RR-BLUP and G-BLUP are similarly based on the BLUP method (HENDERSON 1984), RR-BLUP fits all genome-wide SNPs simultaneously as random effects, whereas G-BLUP combines all genomic information into a realized genomic relationship matrix that is used to model the variance-covariance structure of all individuals. Habier et al. (2007) demonstrated that G-BLUP and RR-BLUP provide equivalent predictions, but G-BLUP is more computationally efficient than RR-BLUP (CLARK and VAN DER WERF 2013). A limitation to both BLUP methods is that all marker loci are assumed to contribute equally to overall genetic variance, which can be an erroneous assumption in cases where a few major QTL exist that explain a much larger proportion of the overall variation compared to background QTL (BERNARDO 2014). In contrast, several Bayesian regression methods exist that allow genetic variance estimates at different SNP loci to vary across the genome (GIANOLA 2013). However, Bayesian approaches can be computationally intensive compared to G-BLUP and RR-BLUP, and studies with real plant breeding data suggest no consistent difference in prediction accuracies between Bayesian and BLUP methods (HAYES *et al.* 2009; LORENZANA

and BERNARDO 2009; MOSER *et al.* 2009; LORENZ *et al.* 2011). Bernardo (2014) has suggested a simple solution to account for known major genes in GS models by including them as fixed effects in order to improve prediction accuracies. In a simulation study, Bernardo (2014) observed significant gains in prediction accuracies by including between one and three large effect QTL (each with $R^2 > 10\%$) in the GS models, but prediction accuracies were reduced by including fixed QTL effects in cases where heritability, population size, and the proportion of genetic variance explained by the major genes were all low.

Empirical evaluation of the predictive potential of GS models in maize was first reported by Albrecht et al. (2011), where relatively high prediction accuracies ($r^2 \leq 0.74$) for grain yield were reported in a large doubled haploid (DH) population. Since Albrecht et al. (2011), other studies have demonstrated the potential of GS for improving complex polygenic traits in maize (TECHNOW *et al.* 2012, 2013; WINDHAUSEN *et al.* 2012; RIEDELSHEIMER *et al.* 2013; CROSSA *et al.* 2013). None of these studies have found a conclusive difference in predictive ability between the various GS methods. However, Windhausen et al. (2012) reported low prediction accuracy when the genomic relationship between training and prediction sets was low, highlighting the importance that prediction models need to be trained in appropriate germplasm, no matter what statistical model is used.

The USDA-ARS North Central Regional Plant Introduction Station (NCRPIS) located in Ames, IA maintains a large and diverse collection of maize inbred lines that represents a century of public and private maize breeding efforts in the United States and from across the globe (ROMAY *et al.* 2013). Within the last year, almost 680,000 genotype-

by-sequencing (GBS; ELSHIRE *et al.* 2011) markers on 2,815 accessions from the NCRPIS collection have become available through the efforts of Romay et al. (2013). The availability of this large set of markers on the NCRPIS collection provides the opportunity for significantly expanding the sample of maize diversity and the marker density for GWAS and GS studies in maize. The objectives of this study were to evaluate 1689 diverse inbred lines from the NCRPIS collection and a subset of their topcross hybrids for resistance to Fusarium ear rot across several years and to conduct genome-wide association studies of resistance to this important disease using a filtered set of 200,978 GBS SNPs from Romay et al. (2013). We also assessed the accuracy of phenotype prediction across the entire set of lines using G-BLUP and compared the accuracy of standard G-BLUP to prediction from a related model that also included the fixed effects of several SNPs identified as associated with resistance from GWAS.

Materials & Methods

Germplasm and experimental design

In 2010, the NCRPIS collection of inbred lines (ROMAY *et al.* 2013) was evaluated for disease resistance at the Central Crops Research Station in Clayton, NC. The 2010 field experiment consisted of 2572 inbred line entries and was arranged in an augmented single replicate design. Experimental entries were divided into 18 sets of differing sizes based on maturity and field assignment, and sets were then randomly subdivided into incomplete blocks (where the maximum block size across sets was 23 plots). Each block within each set was augmented with a B73 check plot in a random position, and five other checks of varying

maturities (IL14H, Ki11, P39, SA24, and Tx303) were included once per set in a random position.

In 2011 and 2012, a novel association mapping panel consisting of 771 diverse inbred line entries was evaluated for disease resistance in Clayton, NC. Based on phenotypic information from the 2010 field experiment, a subset of 681 inbred lines from the NCRPIS collection representing a range of both pedigrees and disease severity scores was chosen for the panel. An additional 90 lines, mostly modern public lines available from North Carolina State University as well as a few lines developed by private industry with recently expired Plant Variety Protection Act (exPVPA) coverage that had become available through the NCRPIS in the spring of 2011 were included. The complete panel of 771 entries was divided into eight sets based on maturity and replicated across the two years using an augmented design. Within years, sets were randomized within the field, and each set was blocked using an α-lattice design (PATTERSON and WILLIAMS 1976). Similar to the NCRPIS evaluation, each block was augmented by a randomly assigned B73 check plot, and five other checks representing a range of maturities and disease reactions (GE440, NC358, 794, B47, and Tx303) were included once per set.

Topcross F_1 hybrids representing a subset of inbred lines from the 2011-2012 association panel were also evaluated in Clayton, NC in 2011 and 2012. Due to seed availability, topcross seed was limited to a sample of 405 inbred lines from the total 771 entries of the association panel. F_1 hybrid seed was generated by crossing inbred lines to either the stiff stalk exPVPA inbred tester B47 or the non-stiff stalk exPVPA inbred tester PHZ51 (or both). Overall, 92 lines were crossed only to B47, 162 lines were crossed only to

PHZ51, and 151 lines were crossed to both testers, resulting in a total of 556 F_1 hybrid entries in the topcross panel. In the 2011 and 2012 field experiments, topcross entries were classified by tester and maturity (early or late, for a total of four tester×maturity combinations), and each tester×maturity combination was randomly subdivided into three groups. One random group of each tester×maturity combination was assigned to a set, for total of three sets (with four groups per set). Similar to the inbred association panel, sets were randomized within the field in each year, groups were randomized within set, and each group was then subdivided into incomplete blocks, but the topcross hybrids were grown in different field blocks than the inbreds. Each block was augmented with a B73×PHZ51 topcross check plot in a random position, and two other hybrids that exhibited relatively good resistance to Fusarium ear rot in previous experiments (Pioneer 31G66 and NC478×GE440) were included once per group. Lastly, one additional check plot of P39×PHZ51 or CML52×PHZ51 was included once per group depending on maturity (early or late, respectively).

Inoculation and phenotyping methods

The 2010 NCPRIS collection experiment and the 2011/2012 inbred association panel experiments were inoculated with local toxigenic *Fusarium verticillioides* isolates using the toothpick method (CLEMENTS *et al.* 2003; ZILA *et al.* 2013). Approximately one week after flowering, a toothpick containing dried *F. verticillioides* conidia was inserted near the base of the primary ear of five plants in each plot. At maturity, inoculated ears were harvested and visually scored for Fusarium ear rot symptoms. Scores were assigned to each ear in

increments of 5% from 0% to 100% diseased based on the percentage of the ear displaying disease symptoms (ROBERTSON *et al.* 2006).

Topcross hybrid experiments in 2011 and 2012 were inoculated with a suspension of *F. verticillioides* conidia using the method described by Robertson et al. (2006). Approximately one week after flowering, 5 mL of a liquid suspension containing 2×10^6 conidia mL^{-1} was injected into the silk channel of the primary ear of five plants in each plot. One week following the first inoculation, 5 mL of the conidia suspension was injected near the base of the primary ear of the same plants inoculated in the first week. At maturity, inoculated ears were harvested and visually scored using the same protocol as the inbred disease experiments. Raw data from both the inbred and topcross experiments are provided in supplemental datasets File B.1 and File B.2, respectively.

Genotypic data

The genotypic data used in this study consisted of 200,978 SNPs filtered from the GBS markers developed by Romay et al. (2013). The original set of markers consisted of 681,257 SNPs generated by the approach described by Elshire et al. (2011) and Glaubitz et al. (in press) with missing data imputed using the haplotype-based imputation method described by Romay et al. (2013). SNP data are available at http://panzea.org/db/gateway?file_id=Romay_etal_2013_imputed_geno_data. In addition, the Romay et al. (2013) marker set was augmented with GBS data for the ninety inbred lines in the 2011/2012 association panel that were not present in the NCPRIS collection in 2010. GBS data for the aforementioned lines were obtained through the Institute for Genomic

Diversity at Cornell Unversity, Ithaca, NY (http://www.igd.cornell.edu). Even after haplotype-based imputation, some missing genotypes exist because the imputation method of Romay et al. (2013) does not impute missing data when the observed scores within a test haplotype window do not sufficiently match the reference haplotype set. Therefore, the augmented SNP marker set was then filtered to include only those markers that had less than 20% missing data (after haplotype-based imputation) and a minor allele frequency (MAF) greater than 0.05. Duplicate samples present in the Romay et al. (2013) data set were also removed from the augmented data set; after this filtering step, genotypic data were available for a total of 2480 inbred lines from across all years combined. The final genotypic data set used in the GWAS analyses is provided in supplemental dataset File B.3.

Statistical analyses

Estimation of least square means

A number of distinct subsets of the complete phenotype data set were used for different analyses; Figure B.1 describes the relationships between the various data subsets analyzed. Fusarium ear rot data from the 2010 NCPRIS collection experiment and the 2011/2012 inbred association panel experiments were first analyzed separately to determine the best fitting spatial model within each year, and then the best models within each year were combined together to form a single multi-environment trial (MET) analysis. Within each year, a model was first fit with a fixed entry effect, fixed first, second, third, and fourth order polynomial trend effects in both the row and column directions (BROWNIE *et al.* 1993), and flowering time as a fixed linear covariate. Only those fixed trend effects significant at *P*

< 0.01 were chosen to remain in the model, and flowering time was also dropped from the model if it was not significant at $P < 0.05$. Once significant fixed effects were selected, random effects were chosen using Akaike's Information Criterion (AKAIKE 1974) to compare four different models within each year: a model fitting only the significant fixed effects; a model fitting significant fixed effects and random set and block within set effects; a model fitting fixed effects and an anisotropic correlated error structure (BROWNIE *et al.* 1993); and a model fitting fixed effects, random set and block within set effects, and an anisotropic correlated error structure. All models were weighted by the number of ears scored within each plot, and a natural logarithmic transformation of raw ear rot scores was used in all analyses due to an association between the magnitude of predicted ear rot values and residuals. All analyses were performed using ASReml version 3 software (GILMOUR *et al.* 2009).

Once the best model within each year was selected, a single MET analysis was conducted by nesting the various best spatial models within year. Fixed effects from the individual year analyses were checked again for significance in the combined model, and those which became insignificant in the combined model were dropped. The combined model consisted of a fixed entry effect, random year and entry×year effects, a heterogeneous error variance structure within each year, and the various spatial effects nested within their respective years: a random set effect in 2010, a random block within set effect in 2010, a fixed first order trend in the row direction in 2011, and a fixed second order trend in the column direction in 2011. Of the 2480 inbred lines with available genotypic data, least squares means were estimated for 1689 lines from the combined model (Figure B.1; File

B.4). Means were not estimable for the other 791 lines due to missing phenotypic observations in all years (typically due to extreme time to maturity or poor seed production). Given the imbalance in the number of experimental entries in 2010 versus 2011/2012, a second filtered least squares mean data set was created that included only the 734 inbred lines for which we had data from all three years of testing (Figure B.1; File B.4).

Ear rot data from the 2011/2012 topcross experiments were analyzed using the same model selection protocol as the inbred experiments. The only difference in model selection in the topcross experiments was the testing of random set, group within set, and block within group effects in addition to other fixed and random effects tested in the inbred models. The combined model for the topcross experiments consisted of a fixed entry effect, random year and entry×year effects, a heterogeneous error variance structure within each year, and the significant spatial and experimental design factors nested within years: a fixed flowering time covariate in both years, an anisotropic correlated error structure in the row direction in both years, and a fixed first order trend in the row direction in 2011. From the combined model, least squares means were estimated for all 556 topcross hybrid entries. Means were then divided into two separate data sets based on tester. The B47 topcross set contained 243 means, and the PHZ51 topcross set contained 313 means (File B.4).

Heritability of Fusarium ear rot resistance was estimated within the inbred association panel and topcross hybrid experiments. The same models used to estimate least square means were used to estimate heritability except entries were treated as random effects to obtain estimates of genetic variance. Entry mean-basis heritability was estimated as

$$\widehat{H}_C = 1 - \frac{\sigma^2_{PPE}}{2\hat{\sigma}^2_G}$$

where σ^2_{PPE} is the average prediction error variance for all pairwise comparisons of entries and $\hat{\sigma}^2_G$ is the estimated genetic variance (CULLIS *et al.* 2006). Five entry mean-basis heritabilities were estimated: across the full inbred association panel, within the filtered inbred subset of 734 lines, across all topcross hybrids, within the B47 topcrosses, and within the PHZ51 topcrosses.

Genotypic correlations between inbred rot resistance and hybrid rot resistance were estimated using individual location least square means for inbred entries and their corresponding topcross hybrids in a multivariate mixed model in ASReml. The least squares means used to calculate genetic correlations were only from years in which both inbred entries and hybrids were evaluated simultaneously (2011 and 2012). The model statement in ASReml was specified as

$$Y_{INB}, Y_{B47}, Y_{PHZ51} = Trait + Trait.Year + Trait.Entry$$

where $Y_{INB}$is the inbred *per se* rot score variate, Y_{B47} is the B47 topcross hybrid rot score variate, $Y_{PHZ51}$is PHZ51 topcross hybrid rot score variate, $Trait$ fits the mean for all three disease variates, $Trait.Year$ fits a fixed year effect for each disease variate, and $Trait.Entry$ fits the random genotype effect for each disease variate. Each term in the model was associated with one variance component for each trait and three covariance components between the three traits.

Association analyses

A genetic kinship matrix (**K**; File B.5) for all 2480 inbred lines based on observed allele frequencies (VANRADEN 2008; method 1) was created using R software version 3.0.1 (R CORE TEAM 2013). A subset of 10,241 SNP markers from the entire genotypic data set of 2480 inbred lines was used to produce **K**. The subset of markers was created by selecting markers from the complete marker set with less than 1% missing data. Missing genotypes remaining in the marker subset were imputed using a stochastic approach described by Zapata-Valenzuela et al. (2013). This method imputes a categorical genotype based on the frequency of all genotypes observed at the same locus across all individuals. This method imputes genotypic values that are expected to maintain the genotypic frequencies observed across the non-missing data. A principal components analysis (PCA) in R was used to obtain the first two principal components of **K** in order to study the association of population structure with mean Fusarium ear rot scores.

The R package GAPIT version 3.35 (Lipka *et al.*, 2012) was used for the genome-wide association analyses based on a compressed mixed linear model (ZHANG *et al.* 2010). Analyses were conducted on four sets of means: the entire set of inbred lines (1689 entries); the filtered set of inbred lines tested in all years (734 entries); the B47 topcross set (243 entries); and the PHZ51 topcross set (313 entries). In each set of means, missing values were included to allow for the same kinship matrix to be used across all analyses. The mixed linear model implemented by GAPIT was

$$y = X\beta + Zu + e$$

where **y** is the vector of ear rot least squares means on the natural-log scale, **β** is a vector of

fixed effects including SNP marker effects, **u** is a vector of random additive genetic effects from background QTL for lines, **X** and **Z** are design matrices, and **e** is a vector of random residuals. The variance of the **u** vector was modeled as

$$\mathrm{Var}(\mathrm{u}) = \mathrm{K}\sigma_{\mathrm{a}}^{2}$$

where **K** is the 2480×2480 matrix of pairwise kinship coefficients and σ_{a}^{2} is the estimated additive genetic variance (YU *et al.* 2006). The full **K** matrix was used for all analyses.

Restricted maximum likelihood estimates of variance components were obtained using the optimum compression level and population parameters previously determined (P3D) options in GAPIT (ZHANG *et al.* 2010). The positive false discovery rate (FDR) across all 200,978 tests of association between one SNP and ear rot resistance was estimated by GAPIT using the method described by Benjamini and Hochberg (1995). The MaizeGDB genome browser (ANDORF *et al.* 2010) was used to identify genes either containing or located adjacent to significant SNP hits from the GWAS.

The 1689 phenotyped lines of the full inbred panel were grouped into one of five major maize subpopulations (stiff stalk, non-stiff stalk, tropical, popcorn, and sweet corn) based on pedigree information compiled by Romay et al. (2013; http://genomebiology.com/content/supplementary/gb-2013-14-6-r55-s1.xlsx). Pedigree descriptors of the additional North Carolina State University lines added to the experiment in 2011 were obtained from http://www.cropsci.ncsu.edu/maize/germplasm.html and appended to the Romay et al. (2013) data set. Lines of mixed ancestry ("unclassified") were dropped from the analysis. Landraces were also dropped due to very small sample size. The frequencies of alleles that reduced disease severity at significantly associated SNPs from the

GWAS were estimated within each subpopulation in R software, and a Fisher's exact test was used to test the null hypothesis that the frequency of the allele conferring increased disease resistance was the same across all five subpopulations.

Cross validation of genome-wide prediction models

Fivefold cross validation was used to evaluate the accuracy and bias of two different prediction models. One model was a standard G-BLUP model that allows prediction of un-phenotyped lines based on their realized genomic relationships with lines that were phenotyped (HABIER *et al.* 2007), and the second was a G-BLUP model augmented by the addition of fixed effects of a small number of SNPs detected with GWAS. Prediction models were built using the plot-basis phenotype data and nearly the same multi-environment models originally used to estimate line least squares means (Figure B.1) to incorporate the variability in precision that exists among the inbred line genotypic means. For the genomic selection analyses, the raw data were filtered to retain only inbred lines that had at least one observation across all three years of testing as well as no missing data at any of the seven SNP loci identified as significant (FDR < 0.20) by GWAS. This was required to ensure that the G-BLUP and G-BLUP + GWAS prediction models were fit using identical observations. After filtering, observations on 1436 inbred lines remained in the raw data set (Figure B.1).

As with the original analyses used to estimate line means, within-plot variability in ear rot scores was used as a weighting factor to account for differences in precision among plot values. Unlike the original analyses, however, the prediction models treated lines as random effects and modeled the variance-covariance relationships among line values using

the realized genomic relationship matrix (**K**). The G-BLUP model predicted phenotypic values for lines based solely on this mixed model. The G-BLUP + GWAS model incorporated the fixed effect of several SNPs identified by GWAS in addition to modeling the random line variance-covariance structure with the realized genomic relationship matrix. Fixed SNPs to include in the G-BLUP + GWAS prediction model were backward selected by fitting all seven significant GWAS SNPs as fixed linear effects in the G-BLUP model. Only those SNPs that were significant at $P < 0.05$ in this combined together were retained for use in the G-BLUP + GWAS prediction models. Best linear unbiased predictions (BLUPs) were computed for each of the 1436 lines from the combined spatial model (without **K**) in order to provide a baseline comparison for the two prediction models. All prediction models were fit using ASReml version 3.0 software (GILMOUR *et al.* 2009).

The predictive accuracies of the G-BLUP model (based on only the kinship matrix) and the G-BLUP + GWAS model (where in addition to the **K** matrix, significant GWAS SNPs were fit together as fixed covariates) were evaluated via five-fold cross validation. In each cross-validation fold, 80% of the 1436 lines with complete data were selected for inclusion in the training data set, and the remaining 20% allocated to the validation data set. Each of the five cross-validation folds was disjoint. Prediction accuracies for each model (G-BLUP or G-BLUP + GWAS) were evaluated by predicting the phenotype of each of the 20% of lines in the validation data set and comparing predictions to their phenotypic BLUPs estimated from the full data set using an identity matrix to model the variance-covariance structure among lines. Line BLUPs were regressed on the predicted values from either G-BLUP or G-BLUP + GWAS, and the regression analysis R^2 value (coefficient of

determination) was recorded as the prediction accuracy. The regression coefficient ($\hat{\beta}$) was also recorded as the bias of the prediction model, where 1 indicates no inflation in line BLUPs. The cross-validation process was replicated ten times for both methods (with random resampling of cross-validation folds across replicates), yielding 100 model runs in total. The accuracy and bias of each model were evaluated as the mean validation R^2 and $\hat{\beta}$ values across the 50 cross-validation analyses for each model.

Results

Line means and heritability

Significant ($P < 0.001$) genotypic variation for ear rot resistance was observed in both the inbred association panel and topcross experiments. Ear rot least squares means among 1689 entries of the inbred association panel ranged from 0.2% to 100% with a mean score of 38.5% (Table 4.1 and File B.4). Least square means for topcross hybrids ranged from 2.5% to 84.8% with a mean score of 21.0%. Entry mean-basis heritability of ear rot resistance in the full inbred association panel was 0.21, while in the balanced subset of 734 entries all tested across three years it was 0.61. Heritability of topcross rot resistance averaged across testers (for the set of lines evaluated in combination with both testers) 0.63, while heritabilites of resistance within the B47 and PHZ51 topcross sets individually were 0.46 and 0.18, respectively. The genotypic correlations between inbred ear rot resistance and resistance in topcrosses to B47 and PHZ51 were 0.39 and 0.42, respectively. The genotypic correlation between performance of B47 topcrosses and PHZ51 topcrosses was 0.48. On an inbred *per se* basis, B47 had a mean ear rot score of 28.1%, while PHZ51 had a mean score

of 58.7% (File B.4).

Genome-wide association mapping of Fusarium ear rot resistance

Background polygenic effects modeled by **K** accounted for 31% of the variation among entry means in the full inbred association panel analysis and 42% of the entry mean variation in the balanced subset inbred association panel (Table 4.2). Principal component decomposition of **K** revealed little association between mean rot scores in the inbred association panel and overall population structure (Figure 4.1). In the topcross analyses, **K** accounted for 31% of the variation among B47 topcross entry means and 39% of the variation among PHZ51 topcross entry means (Table 4.2). From the analysis of the full inbred association panel, two SNPs were identified as significantly associated with ear rot resistance at FDR adjusted *P*-values < 0.20 (Table 4.3; Figure 4.2). At the same significance threshold, five other SNPs were identified as significantly associated with ear rot resistance from the analysis on the filtered inbred data set (Table 4.3; Figure 4.2). No significant SNPs at FDR < 0.20 were identified from either the B47 topcross analysis or the PHZ51 topcross analysis (Figure 4.3), where the minimum raw *P*-values among SNP association tests were 1.3×10^{-5} and 2.3×10^{-5}, respectively. SNPs identified from either of the two inbred analyses explained relatively small proportions of the observed variance in entry means after accounting for the background polygenic effects (individual SNP R^2 values ranged from 1.3% to 3.0%, Table 4.3), and each SNP also had a small allelic effect (-0.13% to -0.27% back-transformed to the original percentage ear rot scale). All significant associations had

negative allelic effect, indicating that the minor allele was associated with lower ear rot (increased diseased resistance) at all loci.

The frequency of disease resistance alleles were estimated at the seven significantly associated SNPs in the same five major maize subpopulations analyzed by Zila et al. (2013) – stiff stalk temperate (SS), non-stiff stalk temperate (NSS), tropical/subtropical (TS), popcorn (PC), and sweet corn (SC) (ROMAY *et al.* 2013). Alleles associated with increased disease resistance at all seven SNP loci were significantly ($P \leq 1.7\times10^{-5}$) overrepresented in the tropical and/or popcorn groups compared to the three other temperate groups (Table 4.4). Disease resistance alleles at all seven SNP loci were absent or nearly absent in the SS, NSS, and SC subpopulations. However, examination of the average of least squares means across lines sampled within a subpopulation showed no major difference in disease severity between the groups, largely agreeing with the principal component analysis of the **K** matrix (Table 4.4; Figure 4.1).

Genes colocalized with associated SNPs

Genes containing SNPs significantly associated with ear rot resistance were characterized using the filtered predicted gene set from the annotated B73 reference genome (SCHNABLE *et al.* 2009). For the purpose of describing local LD around associated SNPs, an LD block was defined as the region surrounding the SNP in which mean LD (r^2) was at least 0.2. All seven SNPs identified across both inbred association panel analyses were within predicted genes on the maize physical map, and six of the seven localized to exonic regions of those genes. Of the two SNPs identified in the full association panel analysis, the first SNP

was in a sucrose synthase gene (GRMZM2G060659) located in an LD block extending approximately 0.2 Mbp on chromosome 5 (Figure 4.4). Examination of the lines carrying the minor allele at this locus revealed no relationship between population structure due to kernel type (namely the sweet corn and popcorn groups) and presence of the minor allele. The other SNP was in a DNA replication factor *CDT1*-like gene (GRMZM2G035665) located at the end of a 0.1 Mbp LD block on chromosome 9 (Figure 4.4). All five SNPs identified in the balanced subset of the inbred association panel analysis were located on chromosome 4. Four of those SNPs were located in a 1.8 Mbp region between physical positions 7,566,354 bp and 9,353,851 bp, representing a region of high linkage disequilibrium (Figure 4.4) covering a genetic distance of less than 1 cM (LIU *et al.* 2009). Two of the SNPs in this region localized to an exonic region of an F-box domain gene, one localized to a thioredoxin gene, and the last localized to a gene of no known function (GRMZM2G012821, GRMZM2G419836, and GRMZM2G372364, respectively). The fifth SNP identified on chromosome 4 located at position 124,930,006 bp localized to an exon of a loricrin-related gene (GRMZM2G106752).

Comparison of genome-wide prediction models

After backward selection from an initial model containing all seven significant GWAS SNPs, only three SNPs remained significant at $P < 0.05$. Both SNPs from the full inbred association panel analysis remained significant (chromosome 5 and chromosome 9; Table 4.3), while only one SNP from chromosome 4 was selected (located at physical position 9,353,851 bp). Average allelic effects of the three SNPs across the ten reps of five-fold cross-validation (50 data samples in total) were consistent with effects estimated in

GWAS based on line means. Effects estimated from cross-validation were -0.142±0.023%, -0.166±0.024%, and -0.119±0.013% at the chromosome 4, 5, and 9 SNPs respectively, while the corresponding GWAS estimates were -0.254%, -0.170%, and -0.134%, respectively (Table 4.3). Including the three GWAS SNPs as fixed effects along with random line effects with variance-covariance modeled using **K** in a genomic prediction model gave a prediction accuracy of 15% averaged across the 50 cross-validation replicate folds (Table 4.5). Comparatively, fitting a traditional G-BLUP prediction model resulted in a significantly ($P < 0.001$) lower average prediction accuracy of 11% across ten replicates (Table 4.5). In addition, G-BLUP + GWAS resulted in reduced bias in the predictions of line BLUPs ($\hat{\beta}$ = 0.52±0.07) compared to G-BLUP ($\hat{\beta}$ = 0.44±0.06; Table 4.5).

Discussion

Heritability and genotypic correlation between experiments

The removal of 953 unreplicated inbred lines that were present only in the 2010 NCPRIS collection experiment from the across year data set had a substantial impact on entry mean-basis heritability ($\widehat{H}_C = 0.21$ on full data set versus $\widehat{H}_C = 0.61$ on filtered data set). This large difference in heritability provided justification for conducting separate GWAS on the complete and filtered inbred association panel data sets (Figure B.1). Improved heritability of the mean values from the filtered panel will contribute to increased power of GWAS, but this is balanced by the loss of diversity and reduced allele replication in the subset compared to the complete set of inbreds. Analyses on the full versus filtered inbred data sets identified different genomic regions significantly associated with Fusarium ear rot

resistance (Table 4.3). These differing results presumably reflect the tradeoffs between higher heritability and larger sample size that affect GWAS power.

Although the heritability estimate for ear rot resistance averaged across testers in the topcross experiment ($\widehat{H}_C = 0.63$) was comparable to that of the filtered inbred data set, no SNPs were identified as being significantly associated with ear rot resistance in either the B47 or PHZ51 topcross data sets. In addition, genetic correlations between inbred *per se* resistance and hybrid performance in the two sets of topcrosses were moderately low ($r_g \leq 0.42$). Several factors may have contributed to the lack of significant SNP loci in the topcross GWAS analyses. A suspension of *F. verticillioides* was used to inoculate the topcross experiments, while the toothpick method was used in the inbred association panels. The suspension method was preferred in the hybrid trials due to its potential to produce more consistent disease pressure (CLEMENTS *et al.* 2003), but it was logistically difficult to apply to the inbred association panel experiments due to both the scale of the inbred experiments as well as the broad range in flowering times of inbred entries compared to hybrid entries. Estimates of genetic variance in the heritability calculations also revealed reduced genetic variance in the topcross experiment compared to the inbred experiments (Table 4.1). Smaller genotypic sample size of the topcross experiment also contributes to reduced power of detection of SNP associations.

Association mapping

Two SNPs significantly associated with ear rot resistance, located on chromosomes 5 and 9, respectively, were identified in the full inbred association panel analysis, and five

additional SNPs (representing two different LD blocks) were identified on chromosome 4 in the filtered inbred panel analysis (Table 4.3). Although all SNPs localized to genic regions, no obvious relationship exists between the predicted functions of these genes and Fusarium ear rot resistance; however the currently limited understanding of pathways contributing to resistance restricts our ability to predict what genes might be involved in resistance to this complex disease. These SNP associations are different than those previously reported by Zila et al. (2013) based on analysis a subset of 267 lines with a smaller and largely distinct set of SNPs. The closest pair of associations between the two studies were the SNPs on chromosome 5, which localized to the same genomic bin; however, they are 34 Mbp distant from each other physically, and 14.4 cM apart genetically (LIU *et al.* 2009). The differences between the results presented here and those reported by Zila et al. (2013) may be due to sample size and sampling of alleles and also due to differences in the SNPs tested for association. None of the three SNPs reported as associated with ear rot resistance by Zila et al. (2013), located on chromosomes 1 (63,540,590 bp), 5 (30,997,717 bp), and 9 (151,295,233 bp), were present in the filtered GBS Romay et al. (2013) marker set, and thus we had no potential to detect them in this study. The nearest neighboring filtered GBS SNP to each of the three SNPs reported by Zila et al. (2013) was located 82 bp (raw $P = 0.44$), 2902 bp (raw $P = 0.74$), and 299 bp away (raw $P = 0.11$), respectively. However, the chromosome 9 SNP from the Zila et al. (2013) study located was present in the original unfiltered Romay et al. (2013) marker set, but a follow-up analysis of this single marker in GAPIT using the full inbred panel found it insignificant (raw $P = 0.78$). Finally, no SNPs

from this analysis colocalized with any QTL intervals identified by Robertson-Hoyt et al. (2006).

Although this study used four times the number of SNP markers (200k versus 47k) and an association panel almost six times as large as those used by Zila et al. (2013), the number of genic regions identified as significantly associated with ear rot was about the same for the two studies (four and three, respectively). Furthermore, the proportion of phenotypic variance among entry means explained on average by the **K** matrix across the two inbred analyses and two topcross analyses was similar to results reported by Zila et al. (2013). These results suggest that the genetic architecture of resistance to Fusarium ear rot is highly polygenic, with substantial genetic variability generated by a large number of effective variants, each with individually small effects. Even with increased marker coverage and a larger association panel, the results of this study highlight the limitations of GWAS to precisely identify allele variants with small effects on complex traits.

Marker coverage in this study is still insufficient to provide SNPs in high LD with all segregating sequence variants. Thus, it is possible that a further increase in marker density might reveal more SNP associations and possibly some genetic variants with larger effects. However, if the genetic architecture really is highly polygenic, then the benefit of increasing marker density on increasing the likelihood of tagging additional causal variations by LD association is likely to offset by the increasingly stringent significance thresholds imposed by the larger number of association tests conducted. The additional benefit of adding markers is also somewhat limited if most of the markers have low minor allele frequency, as is the case for the GBS markers used here (ROMAY *et al.* 2013). The SNP associations detected in this

study had minor allele frequencies ranging from 0.04 to 0.15 (missing phenotypic observations caused some markers to have MAF < 0.05 in the GWAS), compared to minor allele frequencies below 0.05 for more than half of the complete GBS marker set. Besides having low power of detection just due to reduced allele replication, rare alleles tend to be highly associated with population structure since they are usually limited to a single subpopulation, thereby further reducing their potential for trait association following correction for population structure. In this study, we removed SNPs with MAF < 0.05 to ensure reliable associations based on sufficient replication across lines. If rare alleles are a major component of the genetic architecture, however, we may have missed many important associations by dropping SNPs with low allele frequencies that would represent the best possible associations with rare functional alleles. Further studies would be required to better understand the compromises between improving reliability of results by removing rare SNPs versus potentially missing important but rare functional variants.

No significant SNPs were identified in either topcross analysis, and examination of the empirical distribution of *P*-values from the four analyses revealed a tendency towards higher *P*-values in the two topcross analyses compared to the two inbred panel analyses (Figure B.2, Figure B.3, Figure B.4, and Figure B.5). Heterosis plays a significant part in Fusarium ear rot resistance, reducing both genetic variance and the mean level of disease in F_1 hybrids compared to inbred parents (HUNG and HOLLAND 2012), which can reduce the ability to discriminate levels of disease resistance in topcross hybrids. Further, within a set of hybrids created from crosses to a common tester, each topcross hybrid has an equal contribution of half of its alleles at all loci from the common tester, which also reduces

genetic variation among the hybrids. The reduction of genetic variance, along with the smaller sample sizes, reduced the power of detection in hybrids relative to inbreds.

Cross validation of genome-wide prediction models

GWAS associations in situations of low LD that potentially identify causal genes can provide insights into the genetic and biological pathways that may be involved in complex traits that currently are poorly understood. Better understanding of these pathways requires validation of GWAS associations, followed by investigation of the functional effects of the associated genes. The possible role of the genes containing SNPs identified in this study as associated with Fusarium ear rot is obscure, but the identification of these genes at least provides candidates for follow-up biological investigation.

The potential utility of GWAS associations for breeding and phenotype prediction of highly complex and polygenically controlled traits may be questioned, however. The SNP associations in this study each corresponded to small genetic effects. The large proportion of genotypic variance that is associated with the polygenic background component suggests that there are many genes whose individual effects are insufficient to be detected in GWAS, but whose collective effects on Fusarium ear rot resistance are substantial. In this case, the value of selection on a handful of SNPs with small individual effects seems limited compared to the greater potential for genomic prediction based on the polygenic background (LORENZ *et al.* 2011). Rather than compare the potential utility of a few GWAS SNPs to the G-BLUP model, however, we considered that potential utility of a few GWAS SNPs *in addition to* the G-BLUP model.

Although GS prediction accuracies are highest when applied within highly-related breeding pools and may be poor across diverse germplasm (WINDHAUSEN *et al.* 2012), we chose to use the complete diversity panel for comparing G-BLUP and G-BLUP + GWAS prediction models. Although we do not expect to achieve high prediction accuracies from any method in this diverse germplasm set for a trait with low heritability, this is a reasonable 'difficult' test case for comparing different phenotype prediction methods. The finding that adding three SNPs, chosen based on their association to the trait from GWAS, to the G-BLUP model increased prediction accuracy from an average of 11% to 15% is somewhat remarkable, in light of the highly polygenic nature of the trait. Although Bernardo (2014) reported that individual loci should explain at least 10% of the total genetic variance in a biparental family to warrant their inclusion as fixed effects in a prediction model for that same family, our findings suggest that predictive ability can still be improved by including significant SNP loci that individually explain much less than 10% of total genetic variance in more diverse breeding pools. This result suggests that these GWAS associations are robust, even if they are of small effect magnitude. The substantial boost in prediction accuracy achieved from inclusion of these individual SNP effects in the model also suggests that the LD relationships between these SNPs and nearby causal variants are consistent across this diverse germplasm set (which is how they were able to be detected initially), and thus their predictive power across diverse germplasm is relatively good.

Zila et al. (2013) suggested that GWAS could be a useful tool for identifying specific disease resistance allele variants in unadapted maize germplasm, thereby allowing maize breeders to more effectively introgress specific allele variants into adapted germplasm.

However, the small effects of resistance loci identified in this study and Zila et al. (2013) suggest that introgressing a few specific resistance loci may not have a large overall impact on resistance levels within temperate breeding populations. Directly targeting low frequency SNP alleles, particularly when they are harbored in unadapted subpopulations like the tropical and popcorn populations identified both here and by Zila et al. (2013), combined with GS for the polygenic background for both the target trait and general adaptation traits (which will favor selection of individuals with higher proportions of adapted alleles), however, may be a useful compromise to leverage the benefits of both approaches to prediction and selection.

ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation (projects DBI-0321467, IOS-0820619, and IOS-0604923) and by USDA-ARS. USDA-ARS supported the graduate fellowship of C.T. Zila. The authors wish to thank Jason Brewer, Steve Pigozzo, David Horne, Brittany Scott, Sarah Davidson-Dyer, and Colt Jackson for technical assistance, Bode Olukolu and Shang Xue for help with the analyses, Josie Bloom for help with laboratory procedure, Major Goodman and Matt Krakowsky for seed availability, and Sharon Mitchell, Charlotte Acharya, and Dallas Kroon from the IGD at Cornell University for GBS data on new lines.

References

Akaike H., 1974 A new look at the statistical model identification. IEEE Trans. Autom. Control **19**: 716–723.

Albrecht T., Wimmer V., Auinger H.-J., Erbe M., Knaak C., *et al.*, 2011 Genome-based prediction of testcross values in maize. Theor. Appl. Genet. **123**: 339–350.

Andorf C. M., Lawrence C. J., Harper L. C., Schaeffer M. L., Campbell D. A., *et al.*, 2010 The Locus Lookup tool at MaizeGDB: identification of genomic regions in maize by integrating sequence information with physical and genetic maps. Bioinformatics **26**: 434–436.

Benjamini Y., Hochberg Y., 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B **57**: 289–300.

Bernardo R., 2014 Genomewide selection when major genes are known. Crop Sci. **54**: 76–88.

Bernardo R., Yu J., 2007 Prospects for genomewide selection for quantitative traits in maize. Crop Sci. **47**: 1082–1090.

Bolduan C., Miedaner T., Schipprack W., Dhillon B. S., Melchinger A. E., 2009 Genetic variation for resistance to ear rots and mycotoxins contamination in early European maize inbred lines. Crop Sci. **49**: 2019–2028.

Brownie C., Bowman D. T., Burton J. W., 1993 Estimating spatial variation in analysis of data from yield trials: a comparison of methods. Agron. J. **85**: 1244–1253.

Bush B. J., Carson M. L., Cubeta M. A., Hagler W. M., Payne G. A., 2004 Infection and fumonisin production by *Fusarium verticillioides* in developing maize kernels. Phytopathology **94**: 88–93.

Clark S. A., van der Werf J., 2013 Genomic best linear unbiased prediction (gBLUP) for the estimation of genomic breeding values, pp. 321–330 in *Genome-Wide Association Studies and Genomic Prediction*, Methods in Molecular Biology, edited by Gondro C, van der Werf J, and Hayes B. Humana Press, New York, NY.

Clements M. J., Kleinschmidt C. E., Maragos C. M., Pataky J. K., White D. G., 2003 Evaluation of inoculation techniques for Fusarium ear rot and fumonisin contamination of corn. Plant Dis. **87**: 147–153.

Clements M. J., Maragos C. M., Pataky J. K., White D. G., 2004 Sources of resistance to fumonisin accumulation in grain and Fusarium ear and kernel rot of corn. Phytopathology **94**: 251–260.

Crossa J., Beyene Y., Kassa S., Pérez P., Hickey J. M., *et al.*, 2013 Genomic prediction in maize breeding populations with genotyping-by-sequencing. Genes, Genomes, Genetics **3**: 1903–1926.

Cullis B., Smith A., Coombes N., 2006 On the design of early generation variety trials with correlated data. J. Agric. Biol. Environ. Stat. **11**: 381–393.

Ding J.-Q., Wang X.-M., Chander S., Yan J.-B., Li J.-S., 2008 QTL mapping of resistance to Fusarium ear rot using a RIL population in maize. Mol. Breed. **22**: 395–403.

Eller M. S., Holland J. B., Payne G. A., 2008 Breeding for improved resistance to fumonisin contamination in maize. Toxin Rev. **27**: 371–389.

Eller M. S., Payne G. A., Holland J. B., 2010 Selection for reduced Fusarium ear rot and fumonisin content in advanced backcross maize lines and their topcross hybrids. Crop Sci. **50**: 2249–2260.

Elshire R. J., Glaubitz J. C., Sun Q., Poland J. A., Kawamoto K., *et al.*, 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE **6**: e19379.

Flint-Garcia S. A., Thuillet A., Yu J., Pressoir G., Romero S. M., *et al.*, 2005 Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J. **44**: 1054–1064.

Ganal M. W., Durstewitz G., Polley A., Bérard A., Buckler E. S., *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.

Gianola D., 2013 Priors in whole-genome regression: the Bayesian alphabet returns. Genetics: genetics.113.151753.

Gilmour A. R., Gogel B. J., Cullis B. R., Thompson R., 2009 *ASReml User Guide Release 3.0*. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Glaubitz J. C., Casstevens T. M., Lu F., Harriman J., Elshire R. J., *et al.* TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS ONE (accepted).

Habier D., Fernando R. L., Dekkers J. C. M., 2007 The impact of genetic relationship information on genome-assisted breeding values. Genetics **177**: 2389–2397.

Hayes B. J., Bowman P. J., Chamberlain A. C., Verbyla K., Goddard M. E., 2009 Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet. Sel. Evol. **41**: 51.

Henderson C. R., 1984 *Applications of Linear Models in Animal Breeding*. University of Guelph, Guelph, ON, Canada.

Hung H.-Y., Holland J. B., 2012 Diallel analysis of resistance to Fusarium ear rot and fumonisin contamination in maize. Crop Sci. **52**: 2173–2181.

Kimanya M. E., Meulenaer B. De, Roberfroid D., Lachat C., Kolsteren P., 2010 Fumonisin exposure through maize in complementary foods is inversely associated with linear growth of infants in Tanzania. Mol. Nutr. Food Res. **54**: 1659–1667.

King S. B., Scott G. E., 1981 Genotypic differences in maize to kernel infection by *Fusarium moniliforme*. Phytopathology **71**: 1245–1247.

Leslie J. F., Pearson C. A. S., Nelson P. E., Toussoun T. A., 1990 *Fusarium* spp. from corn, sorghum, and soybean fields in the central and eastern United States. Phytopathology **80**: 343–350.

Liu S., Yeh C.-T., Ji T., Ying K., Wu H., *et al.*, 2009 *Mu* transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. PLoS Genet **5**: e1000733.

Lorenz A. J., Chao S., Asoro F. G., Heffner E. L., Hayashi T., *et al.*, 2011 Genomic selection in plant breeding: knowledge and prospects, pp. 77–123 in *Advances in Agronomy*, edited by Sparks D. L. Academic Press, San Diego, CA.

Lorenzana R. E., Bernardo R., 2009 Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor. Appl. Genet. **120**: 151–161.

Marasas W. F., 1996 Fumonisins: history, world-wide occurrence and impact. Adv. Exp. Med. Biol. **392**: 1–17.

Mesterházy Á., Lemmens M., Reid L. M., 2012 Breeding for resistance to ear rots caused by *Fusarium* spp. in maize – a review. Plant Breed. **131**: 1–19.

Miller J. D., 1994 The biosynthetic diversity of secondary metabolites; Epidemiology of Fusarium ear diseases of cereals, pp. 3–36 in *Mycotoxins in Grain: Compounds Other Than Aflatoxin*, edited by Miller J. D. and Trenholm H. L. Eagan Press, St. Paul,.

Moser G., Tier B., Crump R. E., Khatkar M. S., Raadsma H. W., 2009 A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. Genet. Sel. Evol. **41**: 56.

Nankam C., Pataky J. K., 1996 Resistance to kernel infection by *Fusarium moniliforme* in the sweet corn inbred IL 125b. Plant Dis. **80**: 593–598.

Patterson H. D., Williams E. R., 1976 A new class of resolvable incomplete block designs. Biometrika **63**: 83–92.

Pérez-Brito D., Jeffers D., González-de-León D., Khairallah M., Cortés-Cruz M., *et al.*, 2001 QTL mapping of *Fusarium moniliforme* ear rot resistance in highland maize, Mexico. Agrociencia **35**: 181–196.

Piepho H. P., 2009 Ridge regression and extensions for genomewide selection in maize. Crop Sci. **49**: 1165–1176.

Presello D. A., Botta G., Iglesias J., Eyhérabide G. H., 2008 Effect of disease severity on yield and grain fumonisin concentration of maize hybrids inoculated with *Fusarium verticillioides*. Crop Prot. **27**: 572–576.

R Core Team, 2013 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rheeder J. P., Marasas W. F. O., Thiel P. G., Sydenham E. W., Shephard G. S., *et al.*, 1992 *Fusarium moniliforme* and fumonisins in corn in relation to human esophageal cancer in Transkei. Phytopathology **82**: 353.

Riedelsheimer C., Endelman J. B., Stange M., Sorrells M. E., Jannink J.-L., *et al.*, 2013 Genomic predictability of interconnected biparental maize populations. Genetics **194**: 493–503.

Robertson L. A., Kleinschmidt C. E., White D. G., Payne G. A., Maragos C. M., *et al.*, 2006 Heritabilities and correlations of Fusarium ear rot resistance and fumonisin contamination resistance in two maize populations. Crop Sci. **46**: 353–361.

Robertson-Hoyt L. A., Jines M. P., Balint-Kurti P. J., Kleinschmidt C. E., White D. G., *et al.*, 2006 QTL mapping for Fusarium ear rot and fumonisin contamination resistance in two maize populations. Crop Sci. **46**: 1734–1743.

Romay M. C., Millard M. J., Glaubitz J. C., Peiffer J. A., Swarts K. L., *et al.*, 2013 Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. **14**: R55.

Schnable P. S., Ware D., Fulton R. S., Stein J. C., Wei F., *et al.*, 2009 The B73 maize genome: complexity, diversity, and dynamics. Science **326**: 1112–1115.

Shirima C. P., Kimanya M. E., Kinabo J. L., Routledge M. N., Srey C., *et al.*, 2013 Dietary exposure to aflatoxin and fumonisin among Tanzanian children as determined using biomarkers of exposure. Mol. Nutr. Food Res. **57**: 1874–1881.

Technow F., Bürger A., Melchinger A. E., 2013 Genomic prediction of Northern corn leaf blight resistance in maize with combined or separated training sets for heterotic groups. Genes, Genomes, Genetics **3**: 197–203.

Technow F., Riedelsheimer C., Schrag T. A., Melchinger A. E., 2012 Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. Theor. Appl. Genet. **125**: 1181–1194.

van Egmond H. P., Schothorst R. C., Jonker M. A., 2007 Regulations relating to mycotoxins in food; Perspectives in a global and European context. Anal. Bioanal. Chem. **389**: 147–157.

VanRaden P. M., 2008 Efficient methods to compute genomic predictions. J. Dairy Sci. **91**: 4414–4423.

Warren H. L., Kommedahl T., 1973 Prevalence and pathogenicity to corn of Fusarium species from corn roots, rhizosphere, residues, and soil. Phytopathology **63**: 1288–1290.

Windhausen V. S., Atlin G. N., Hickey J. M., Crossa J., Jannink J.-L., *et al.*, 2012 Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. Genes, Genomes, Genetics **2**: 1427–1436.

Yates I. E., Bacon C. W., Hinton D. M., 1997 Effects of endophytic infection by *Fusarium moniliforme* on corn growth and cellular morphology. Plant Dis. **81**: 723–728.

Yu J., Pressoir G., Briggs W. H., Vroh Bi I., Yamasaki M., *et al.*, 2006 A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. **38**: 203–208.

Zapata-Valenzuela J., Whetten R. W., Neale D., McKeand S., Isik F., 2013 Genomic estimated breeding values using genomic relationship matrices in a cloned population of loblolly pine. Genes, Genomes, Genetics **3**: 909–916.

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

Zila C. T., Samayoa L. F., Santiago R., Butrón A., Holland J. B., 2013 A genome-wide association study reveals genes associated with Fusarium ear rot resistance in a maize core diversity panel. Genes, Genomes, Genetics **3**: 2095–2104.

Table 4.1. Sample size (N), mean ear rot severity, genotypic variance component estimates ($\hat{\sigma}_G^2$), average prediction error variance (σ_{PPE}^2) and heritability ($\widehat{H}_C$) estimates for Fusarium ear rot resistance in the full inbred association panel, filtered association panel, across the topcross experiment, and within the B47 and PHZ51 topcrosses, respectively.

	N	Mean (%)[a]	$(\hat{\sigma}_G^2)^b$	$(\sigma_{PPE}^2)^c$	$\widehat{H}_C$
Full inbred panel	1689	38.5	0.15	0.24	0.21
Filtered inbred panel	734	33.0	0.18	0.14	0.61
Topcrosses	556	21.0	0.13	0.10	0.63
B47	243	23.1	0.15	0.16	0.46
PHZ51	313	19.4	0.06	0.10	0.18

[a] Mean ear rot severity is reported as the average of the entry least square means (back-transformed to the original 0-100% disease severity scale).
[b] Estimated genetic variance component from ASReml.
[c] Average prediction error variance among all pair-wise comparisons of entries from ASReml.

Table 4.2. Number of lines, number of groups and compression level of the full 2480×2480 kinship matrix, and proportion of total line mean variance explained by additive relationship matrix from the four mixed-linear model (MLM) analyses.

	N[a]	Groups[b]	Compression[c]	$\left(\frac{\hat{\sigma}^2_G}{\hat{\sigma}^2_G+\hat{\sigma}^2}\right)^d$
Full inbred panel	1689	2100	1.18	0.31
Filtered inbred panel	734	2000	1.24	0.42
B47 topcrosses	243	1760	1.41	0.31
PHZ51 topcrosses	313	1770	1.40	0.39

[a] Total number of entries included in the analysis.
[b] Number of groups determined by optimum compression (note that the complete kinship matrix for 2480 lines was used for all analyses).
[c] Compression level is the average number of individuals per group.
[d] Polygenic additive background genetic variance divided by total phenotypic variance. This ratio was estimated in GAPIT by fitting the kinship matrix (**K**) in the mixed linear model without any SNP marker effects.

Table 4.3. Chromosome locations (AGP v2 coordinates), allele effect estimates, genes containing SNP, and other summary statistics for the seven SNPs significantly associated with Fusarium ear rot resistance from the two inbred association panel analyses.

Chromosome	SNP physical position (bp)	Raw *P*-value	FDR adjusted *P*-value	Minor allele frequency	Allele effect (%)[a]	(R^2)[b]	Gene containing SNP
Full inbred panel analysis							
5	64,771,372	8.83×10^{-7}	0.089	0.07	-0.170	1.3	GRMZM2G060659
9	19,532,465	8.44×10^{-8}	0.017	0.15	-0.134	1.5	GRMZM2G035665
Filtered inbred panel analysis							
4	7,566,354	7.34×10^{-7}	0.074	0.10	-0.230	2.9	GRMZM2G372364
4	7,618,125	2.67×10^{-6}	0.175	0.10	-0.225	2.6	GRMZM2G012821
4	7,618,284	3.96×10^{-6}	0.175	0.11	-0.205	2.5	GRMZM2G012821
4	9,353,851	6.14×10^{-7}	0.074	0.07	-0.254	3.0	GRMZM2G419836
4	124,930,006	4.36×10^{-6}	0.175	0.04	-0.271	2.5	GRMZM2G106752

[a] Allele effects are reported back-transformed to the original 0-100% disease severity scale. Effects are in reference to the minor allele.

[b] R^2, proportion of total entry mean variance remaining after accounting for background polygenic variance associated with a SNP as computed by GAPIT.

Table 4.4. Allele frequencies of significantly associated SNPs in the five major maize subpopulations and *P*-value of Fisher's exact test of the null hypothesis of equal allele frequencies across subpopulations.

		Resistance allele frequency (%)[a]						N[b]				
Chromo-some	SNP physical position (bp)	SS[c]	NSS	TS	PC	SC	*P*-value	SS	NSS	TS	PC	SC
4	7,566,354	1.2	0.0	32.0	60.4	0.0	$<2.2\times10^{-16}$	164	171	222	48	51
4	7,618,125	0.6	0.0	30.7	47.9	0.0	$<2.2\times10^{-16}$	164	171	215	48	52
4	7,618,284	0.6	0.0	36.5	66.0	0.0	$<2.2\times10^{-16}$	159	168	211	47	51
4	9,353,851	0.6	0.6	31.9	0.0	0.0	$<2.2\times10^{-16}$	161	168	213	61	50
4	124,930,006	0.0	1.8	8.8	3.3	0.0	1.7×10^{-5}	162	166	238	60	51
5	64,771,372	0.0	4.7	8.1	14.8	0.0	3.2×10^{-6}	164	170	246	61	51
9	19,532,465	2.5	7.2	26.6	26.7	2.0	4.9×10^{-15}	158	167	241	60	51
	Ear rot mean (%)[d]	39.6	39.6	43.0	41.3	61.1						

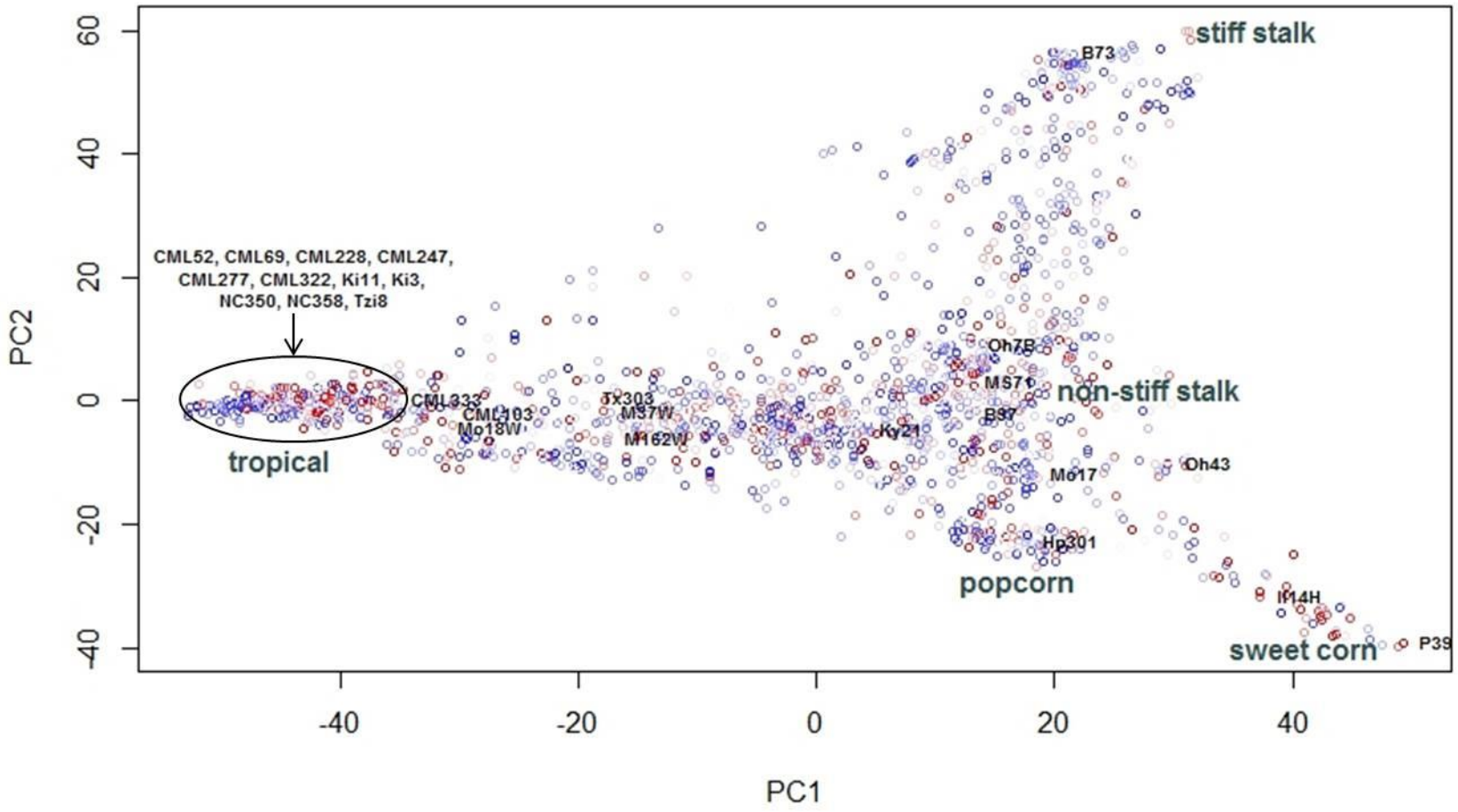
[a] At all SNP loci the minor allele is associated with increased disease resistance.
[b] N, total number of lines within each subpopulation with marker calls at a particular SNP locus.
[c] SS, stiff stalk; NSS, non-stiff stalk; TS, tropical/subtropical; PC, popcorn; SC, sweet corn.
[d] Overall phenotypic ear rot means are the average of least squares means across members of each subpopulation.

Table 4.5. Prediction accuracies (R^2) and regression coefficient estimates for the G-BLUP and G-BLUP + GWAS five-fold cross validation models. Values are reported as averages across the five-folds within each replicate.

	G-BLUP		G-BLUP + GWAS	
Rep	$\hat{\beta}^a$	$(R^2)^b$	$\hat{\beta}^a$	$(R^2)^b$
1	0.43	11.0	0.50	14.5
2	0.42	9.0	0.51	13.1
3	0.44	10.6	0.50	14.6
4	0.44	10.3	0.53	14.8
5	0.45	11.1	0.53	15.4
6	0.44	10.2	0.51	13.9
7	0.46	11.5	0.51	15.1
8	0.45	10.5	0.53	15.1
9	0.47	11.7	0.54	15.5
10	0.43	11.0	0.50	14.5
Mean	0.44	10.7	0.52	14.6
SE[c]	0.06	1.5	0.07	2.1

[a] Coefficients from the regression of line BLUPs on predicted breeding values from GBLUP or GBLUP + GWAS.
[b] Coefficient of determination from regression model.
[c] Standard error of mean.

Figure 4.1. Genetic relationships between the 1689 lines of the full inbred association panel visualized using a principal component analysis of the **K** matrix. The horizontal and vertical axes are the first and second principal components, respectively. The color gradient from blue to red of the points represents the relative mean Fusarium ear rot score of each line (blue is most resistant and red is most susceptible). Five major recognized heterotic group clusters are labeled in large gray font, and the 26 nested association mapping (NAM) population founders and Mo17 are labeled in small black font for reference.


PC2
PC1
60
40
20
0
-20
-40
-40
-20
0
20
40
CML52, CML69, CML228, CML247,
CML277, CML322, Ki11, Ki3,
NC350, NC358, Tzi8
tropical
CML333
CML103
Mo18W
Tx303
M37W
M162W
Ky21
Oh7B
MS71
B97
Mo17
B73
Oh43
Hp301
popcorn
non-stiff stalk
stiff stalk
I14H
P39
sweet corn

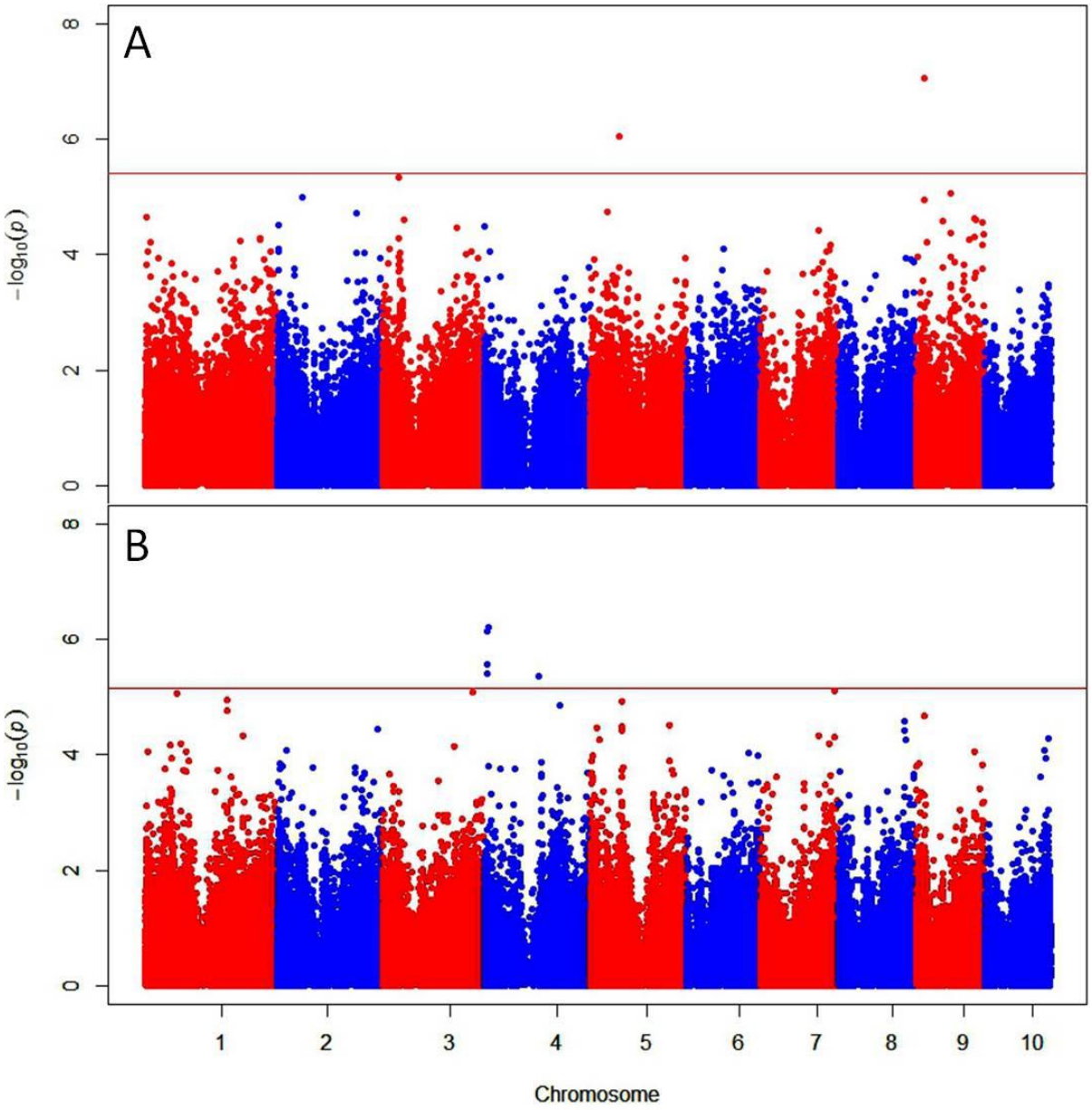


Figure 4.2. Manhattan plots showing significant associations (points above the red FDR = 0.20 threshold lines) from the full inbred association panel (A) and filtered inbred association panel (B) GWAS analyses. The vertical axis indicates $-\log_{10}$ of *P*-value scores, and the horizontal axis indicates chromosomes and physical position of SNPs.

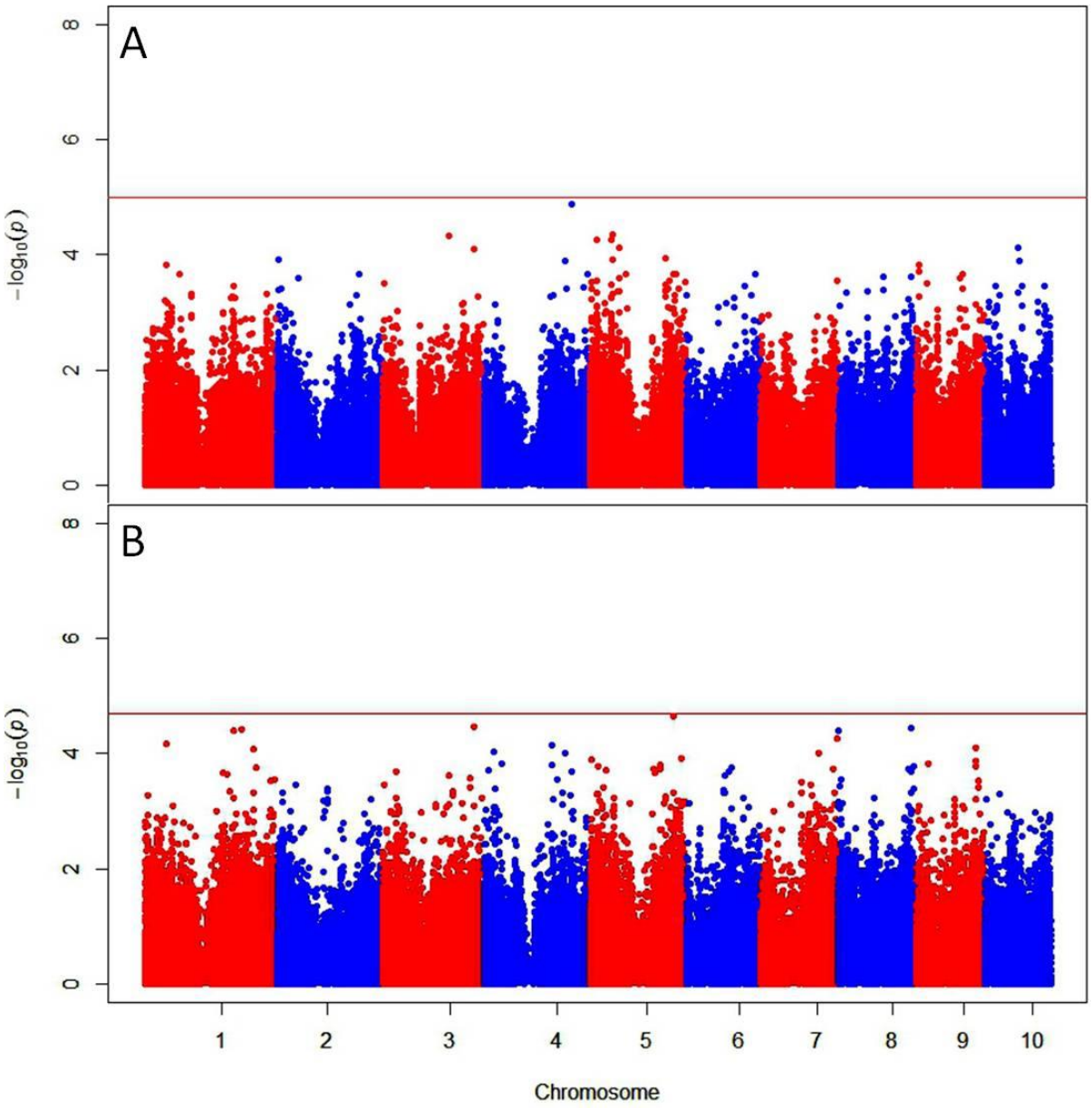
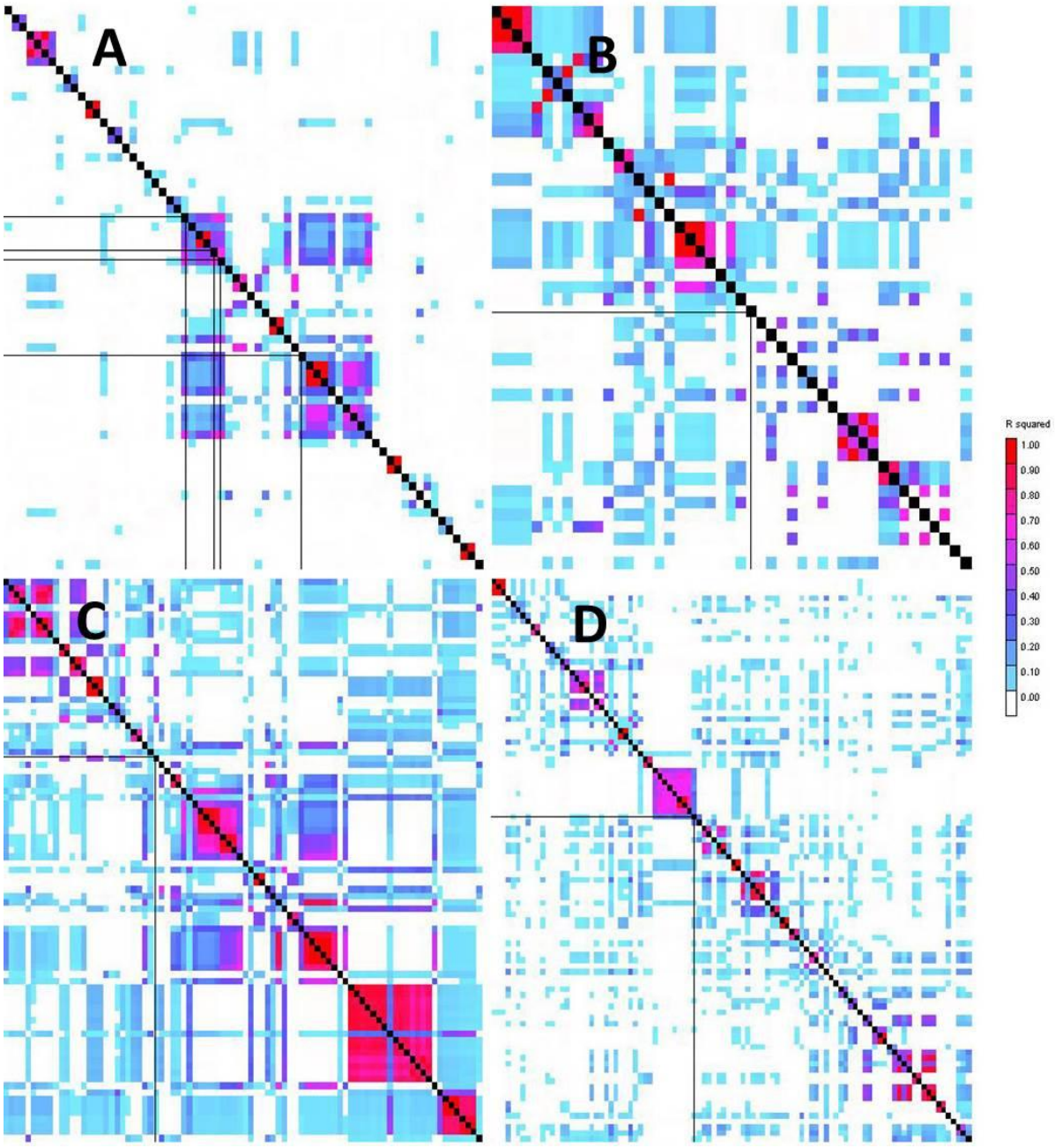


Figure 4.3. Manhattan plots showing significant associations (points above the red FDR = 0.20 threshold lines) from the B47 topcross (A) and PHZ51 topcross (B) GWAS analyses. The vertical axis indicates $-\log_{10}$ of *P*-value scores, and the horizontal axis indicates chromosomes and physical position of SNPs.

Figure 4.4. LD heatmaps showing LD measure (r^2) calculated for each pair-wise combination of SNPs in an approximately ±0.5 Mbp region surrounding each SNP significantly associated with ear rot resistance in the two inbred association panel analyses. (A) LD around the four SNPs chromosome 4 SNPs located in the 7.6 Mbp to 9.4 Mbp interval. (B) LD around chromosome 4 SNP at physical position 124.9 Mbp. (C) LD around chromosome 5 SNP. (D) LD around chromosome 9 SNP. The significant SNP(s) on each chromosome is highlighted by the perpendicular black lines within each heatmap. Colors indicate the magnitude of each pair-wise r^2 measure (r^2=1 is red to r^2=0 is white).


A
B
C
D
R squared
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00

APPENDICES

APPENDIX A: Supplemental Material for Chapter 3

Table A.1. Heritability estimates for Fusarium ear rot resistance, mean ear rot severity, heritability estimates for silking date, regression coefficients for silking date covariates, and significance level of regression coefficients. Estimates are reported for each environment individually, across years within the North Carolina and Galicia environments, and combined across all environments.

	Fusarium ear rot		Silking date		
Environment	$\widehat{H}_c$	Mean (%)[a]	$\widehat{H}_c$	$\hat{\beta}$ (%/day)[b]	*P*-value
NC 2010	0.44	46.7	0.92	0.03	0.023
NC 2011	0.47	26.8	0.38	0.04	0.024
NC 2012	0.71	55.1	0.78	0.01	0.340
Galicia 2010	0.53	7.6	0.93	0.05	0.022
Galicia 2011	0.51	3.4	0.90	0.02	0.248
NC, all years	0.73	41.1	0.95	0.02	<0.001
Galicia, all years	0.71	7.4	0.92	0.02	0.099
Combined	0.75	22.1	0.98	0.02	<0.001

[a] Mean ear rot severity is reported as the average of the line least square means calculated within and across environments. Means are reported back-transformed to the original 0-100% disease severity scale.

[b] Regression coefficients for the silking date covariate in the Fusarium ear rot models. Coefficients are reported back-transformed to the original 0-100% disease severity scale. As an example, a one day increase in silking date in NC 2010 increased the ear rot score for an observation by 0.03%.

Table A.2. Climate data for the three North Carolina and two Galicia environments. Average daily minimum temperature, average daily maximum temperature, average daily overall temperature, and cumulative precipitation level are reported for two time intervals in each environment: planting date to the average silking date (date at which at least 50% of the plots within an environment had silked) and average silking date to 45 days post-silking.

	Planting date to average silking date				Average silking date to 45 days post-silking			
Environment	Average daily min. temp. (°C)	Average daily max. temp. (°C)	Average daily temp. (°C)	Cumulative precipitation (mm)	Average daily min. temp. (°C)	Average daily max. temp. (°C)	Average daily temp. (°C)	Cumulative precipitation (mm)
NC 2010	17.5	28.4	23.0	252.7	21.9	32.7	27.3	120.7
NC 2011	17.4	28.9	23.1	70.6	22.1	33.5	27.8	90.9
NC 2012	16.1	27.3	21.7	9.1	22.3	32.1	27.2	258.8
Galicia 2010	7.0	19.5	13.3	356.2	13.3	25.6	19.5	81.3
Galicia 2011	10.8	25.8	19.7	63.1	11.4	26.5	18.7	137.2

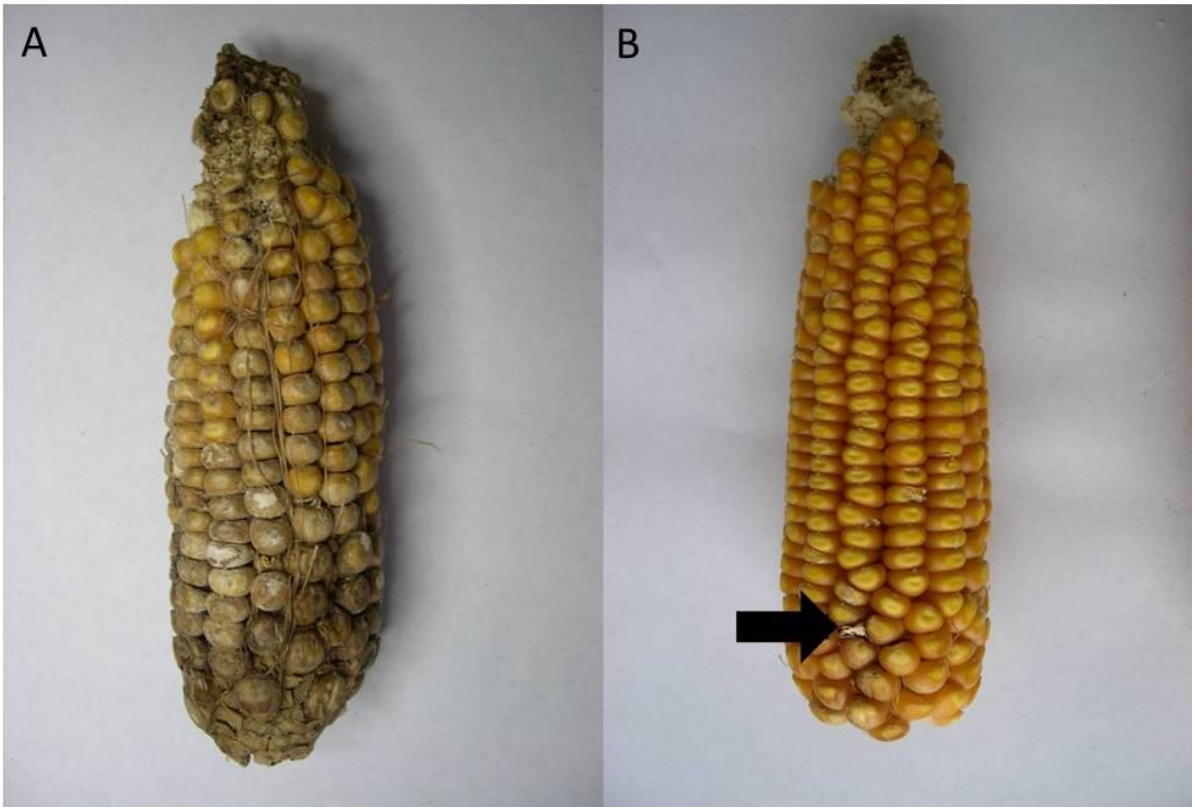


Figure A.1. (A) Example of a susceptible (100% severity) phenotype. (B) Example of a resistant (0% severity) phenotype. The arrow indicates the point of inoculation in the resistant ear.

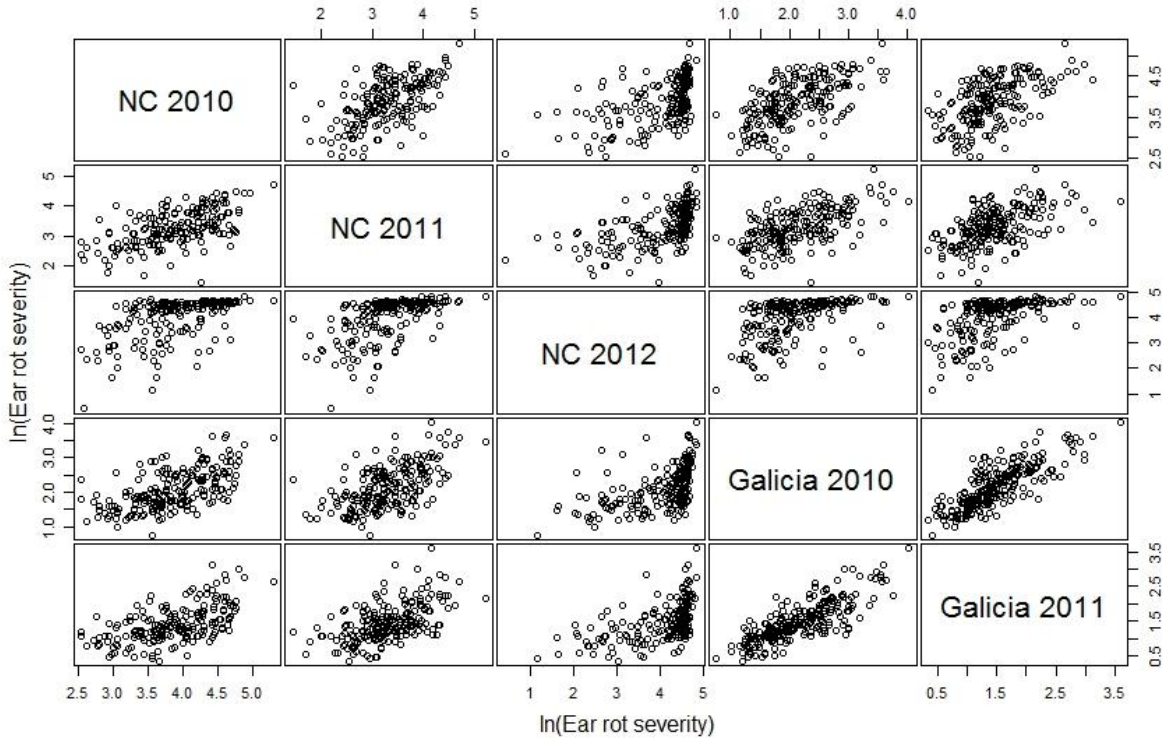


Figure A.2. Scatter plot matrix illustrating the genotypic relationship of Fusarium ear rot resistance between environments. The model used to estimate variance components and genetic correlations in the combined analysis was used to predict least square means for each inbred line within each environment (treating line as a fixed effect instead of random). Means for each line on the natural log transformed scale are plotted against one another in each pair-wise combination of environments.

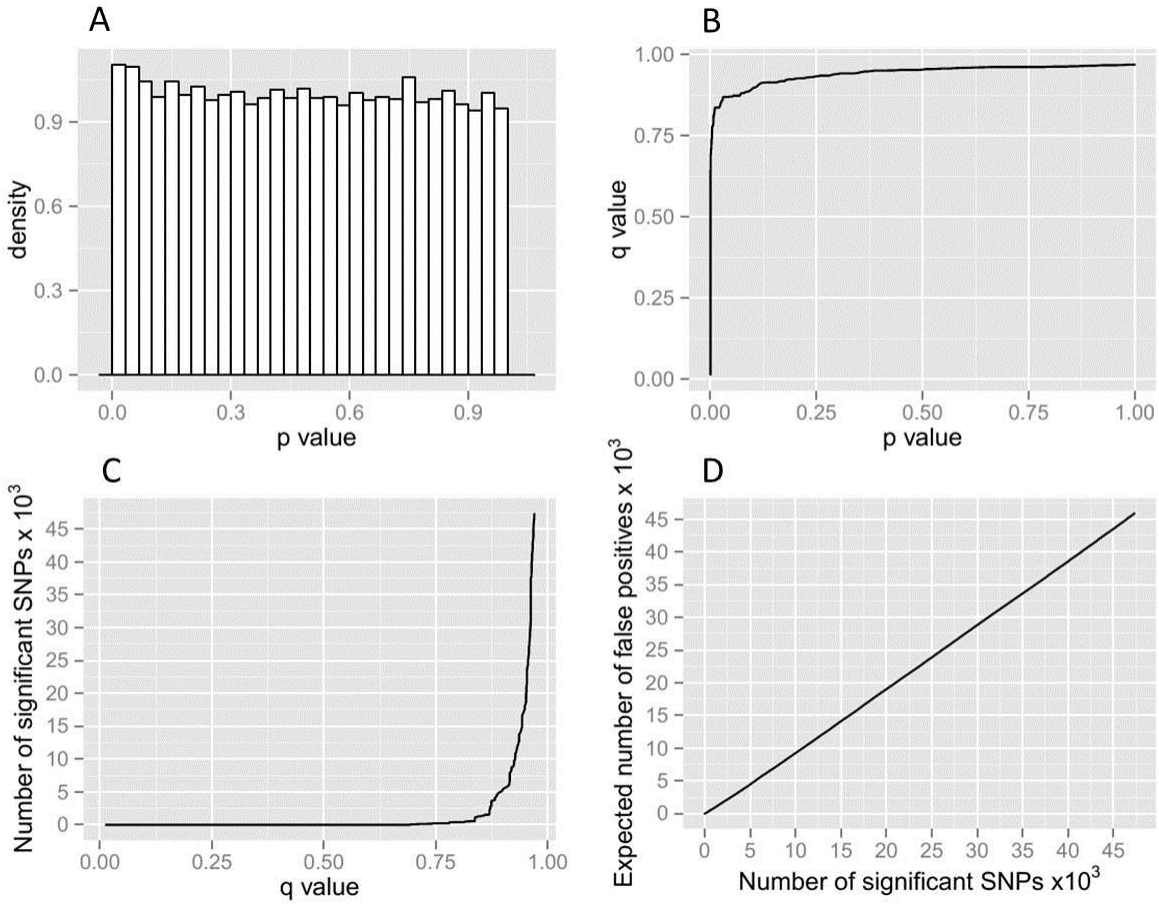


Figure A.3. Estimating the false discovery rate (FDR) for SNP marker association with Fusarium ear rot resistance in the North Carolina analysis. (A) A density histogram showing *p*-value distribution of 47,445 SNPs following GWAS. (B) The *q*-values plotted against their respective *p*-values. (C) The number of SNPs plotted against each of the respective *q*-value estimates. (D) The expected number of false positive SNPs versus the total number of significant SNPs given the *q*-values.

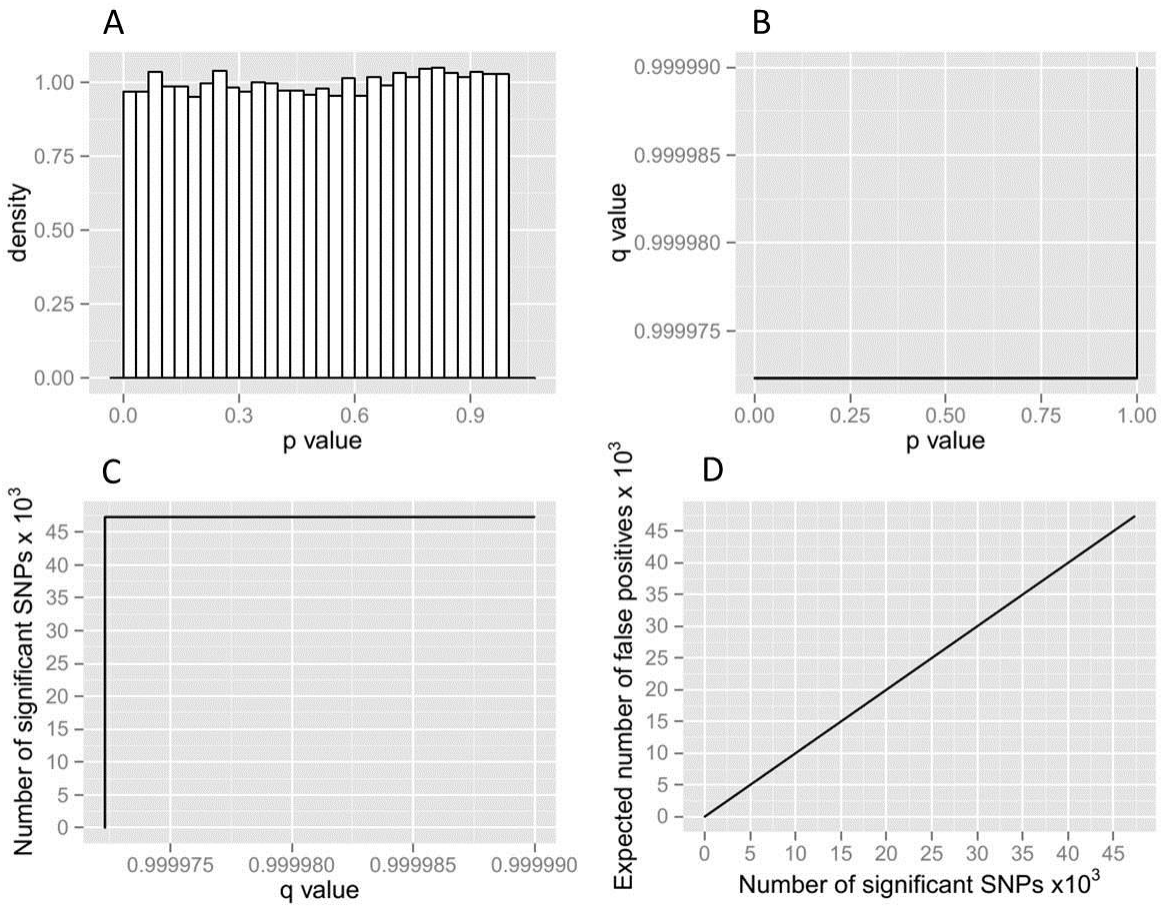


Figure A.4. Estimating the false discovery rate (FDR) for SNP marker association with Fusarium ear rot resistance in the Galicia analysis. (A) A density histogram showing *p*-value distribution of 47,445 SNPs following GWAS. (B) The *q*-values plotted against their respective *p*-values. (C) The number of SNPs plotted against each of the respective *q*-value estimates. (D) The expected number of false positive SNPs versus the total number of significant SNPs given the *q*-values.

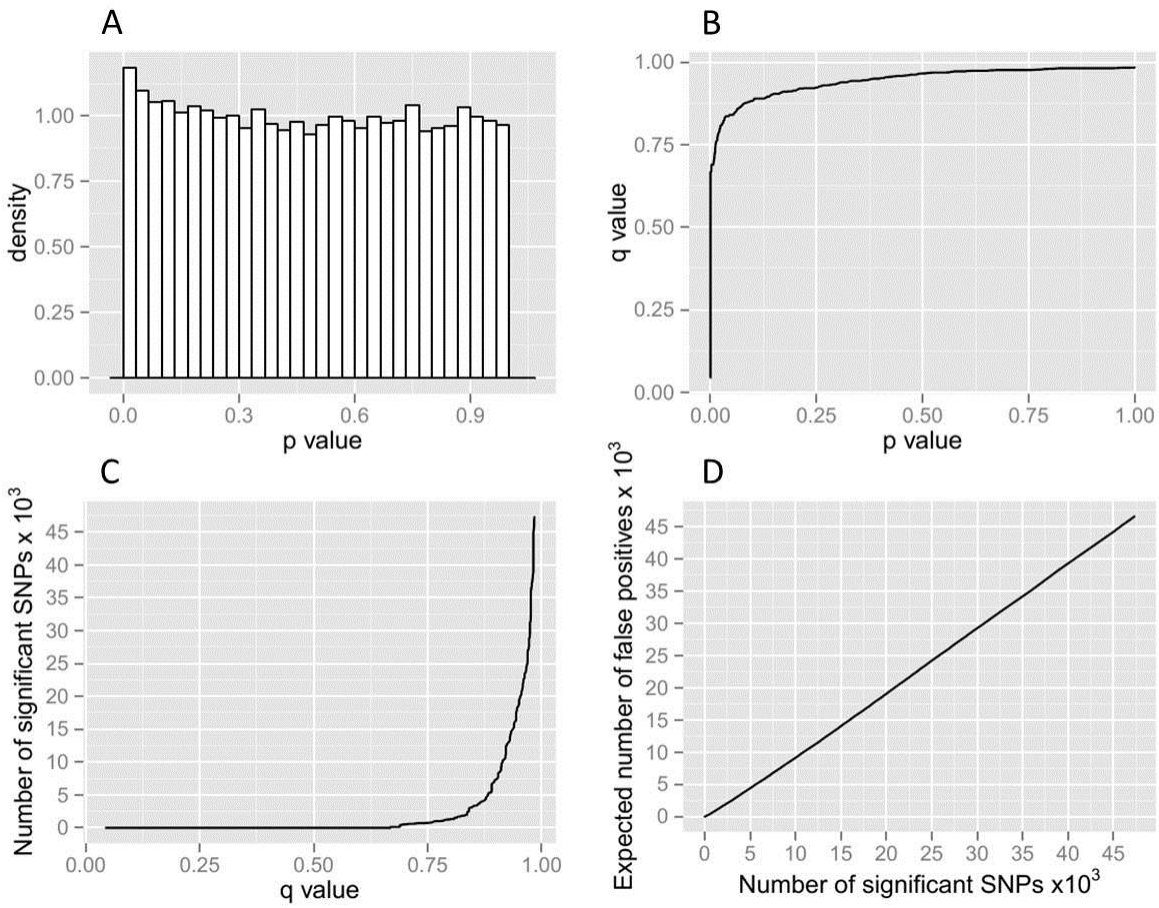


Figure A.5. Estimating the false discovery rate (FDR) for SNP marker association with Fusarium ear rot resistance in the combined analysis. (A) A density histogram showing *p*-value distribution of 47,445 SNPs following GWAS. (B) The *q*-values plotted against their respective *p*-values. (C) The number of SNPs plotted against each of the respective *q*-value estimates. (D) The expected number of false positive SNPs versus the total number of significant SNPs given the *q*-values.

Files A.1-A.3

Supporting data

Available for download at: http://www.g3journal.org/content/3/11/2095/suppl/DC1

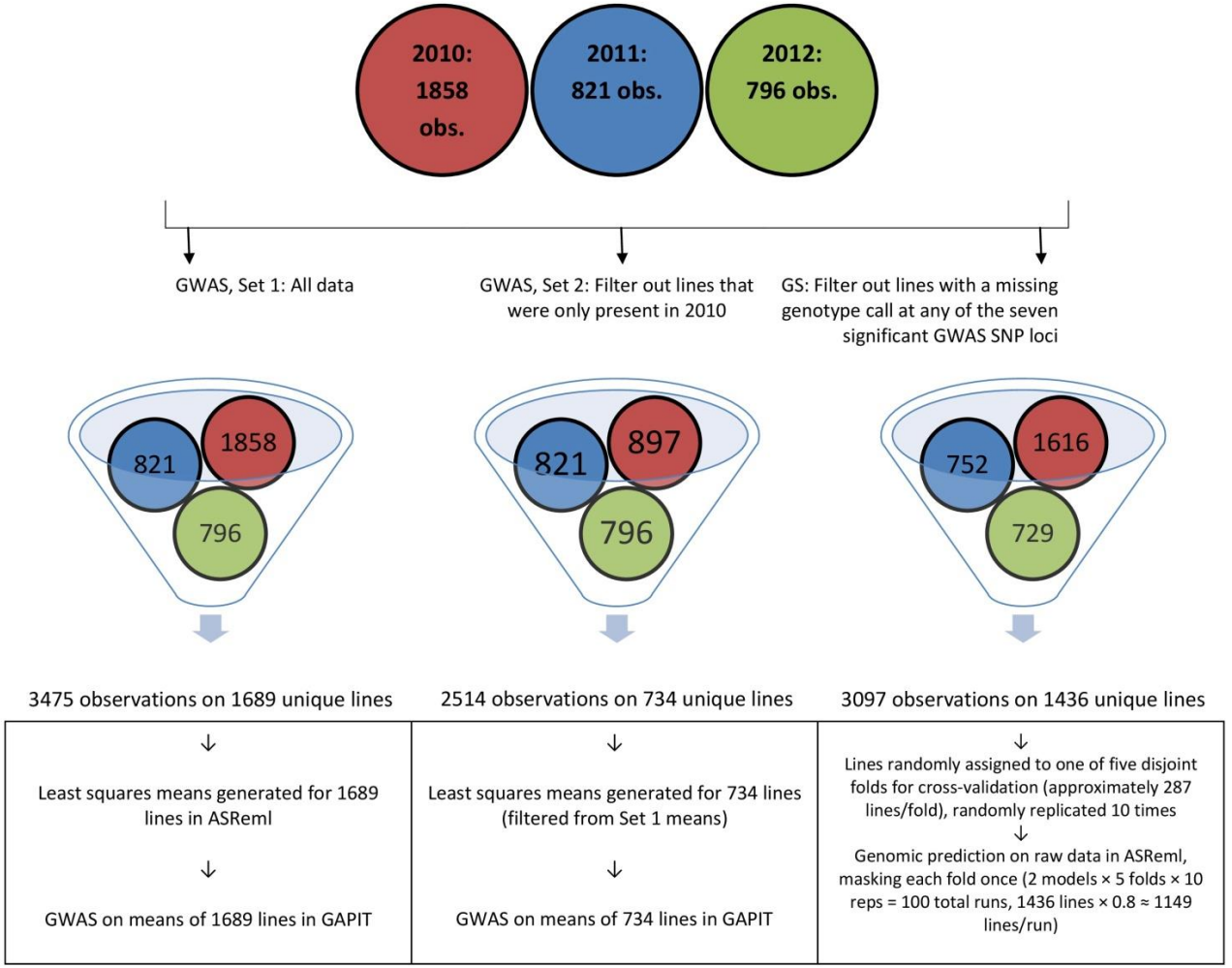
File A.1. Raw phenotypic data from three years in North Carolina and two years in Galicia. Formatted for analysis in ASReml software. Columns in the data file are as follows from left to right: location (Loc, 1=NC, 2=Galicia), environment (Env, a unique combination of location and year), year, row (field position of plot from front of the field to the back), column (field position of plot from left to right), set ("99" is a placeholder in Galicia), rep ("99" is a placeholder in North Carolina), block, plot number, line name (Material), entry number, flowering date (converted to the number of days after planting until flowering), Fusarium ear rot score (rot_AVG, averaged across ears within the plot), number of ears scored within each plot (earno), and the natural log transformation of the average ear rot score (logrot).

File A.2. Least square means for 267 inbred lines estimated within each experiment (North Carolina and Galicia) and across experiments. Formatted for analysis in Tassel software. Column <trait> (Tassel nomenclature) contains the line names corresponding to File A.1, and the other columns are as follows: least square means based on North Carolina data (NC_BLUE), means based on Galicia data (ES_BLUE), and means based on combined data (Overall_BLUE).

File A.3. A 279 ×279 genetic kinship matrix (**K**) based on Van Raden (2008). Formatted for analysis in Tassel software. The first column contains line names, and all other columns contain the pair-wise kinship coefficients between lines.

APPENDIX B: Supplemental Material for Chapter 4

Figure B.1. Diagram showing the relationships between the full inbred line data set and the subsets used for GWAS and genomic prediction. The total number of phenotypic observations collected within the three years of field experiments is shown in the top three circles (2010 in red, 2011 in blue, and 2012 in green). The number of observations from each year that is present in each data subset (GWAS Set 1, GWAS Set 2, and GS) after filtering is shown in the three funnels. The type of analysis conducted on each subset is shown in the bottom row of text boxes.


2010: 1858 obs.
2011: 821 obs.
2012: 796 obs.
GWAS, Set 1: All data
GWAS, Set 2: Filter out lines that were only present in 2010
GS: Filter out lines with a missing genotype call at any of the seven significant GWAS SNP loci
821
1858
796
821
897
796
752
1616
729
3475 observations on 1689 unique lines
2514 observations on 734 unique lines
3097 observations on 1436 unique lines
Least squares means generated for 1689 lines in ASReml
GWAS on means of 1689 lines in GAPIT
Least squares means generated for 734 lines (filtered from Set 1 means)
GWAS on means of 734 lines in GAPIT
Lines randomly assigned to one of five disjoint folds for cross-validation (approximately 287 lines/fold), randomly replicated 10 times
Genomic prediction on raw data in ASReml, masking each fold once (2 models × 5 folds × 10 reps = 100 total runs, 1436 lines × 0.8 ≈ 1149 lines/run)

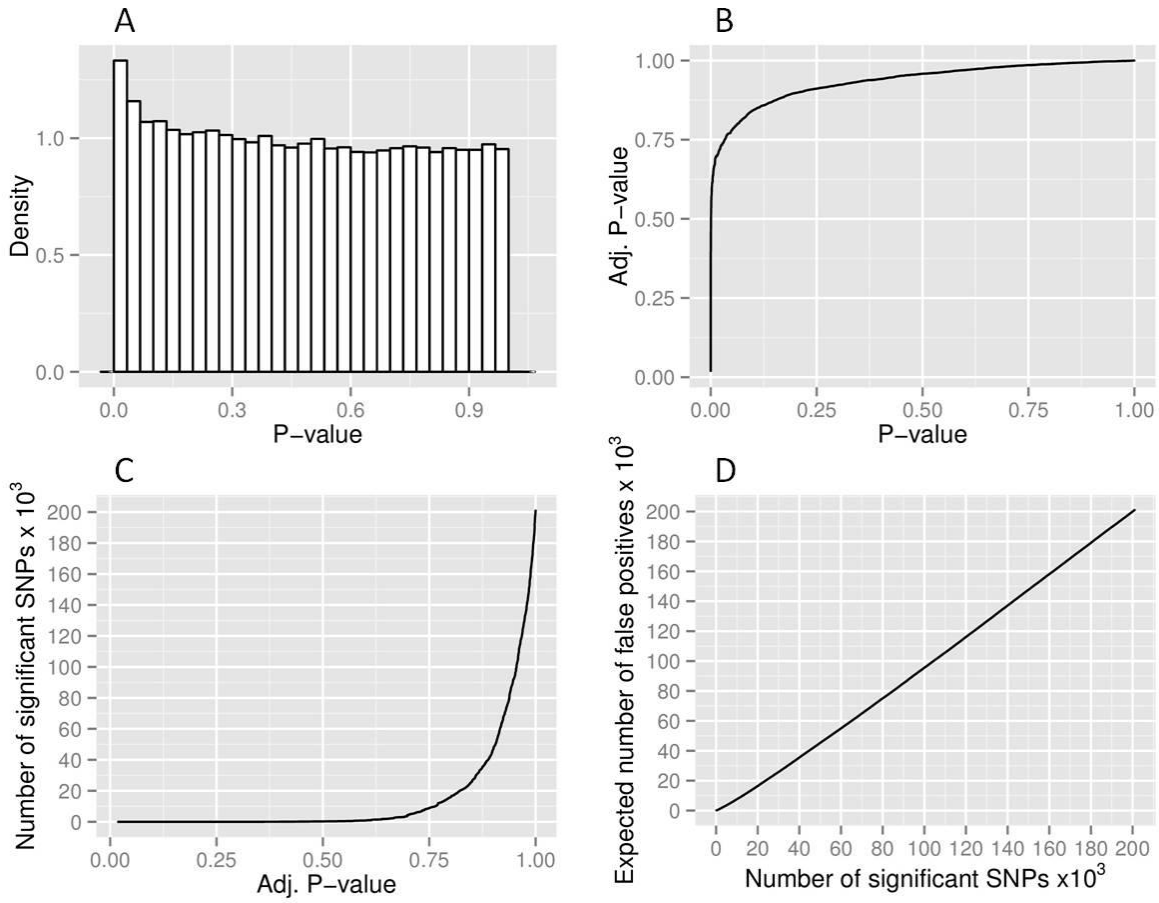


Figure B.2. Estimating the false discovery rate (FDR) for SNP marker association with Fusarium ear rot resistance in the full inbred association panel analysis. (A) A density histogram showing raw *P*-value distribution of 200,978 SNPs following GWAS. (B) The FDR-adjusted *P*-values plotted against their respective raw *P*-values. (C) The number of SNPs plotted against each of the respective FDR-adjusted *P*-value estimates. (D) The expected number of false positive SNPs versus the total number of significant SNPs given the FDR-adjusted *P*-values.

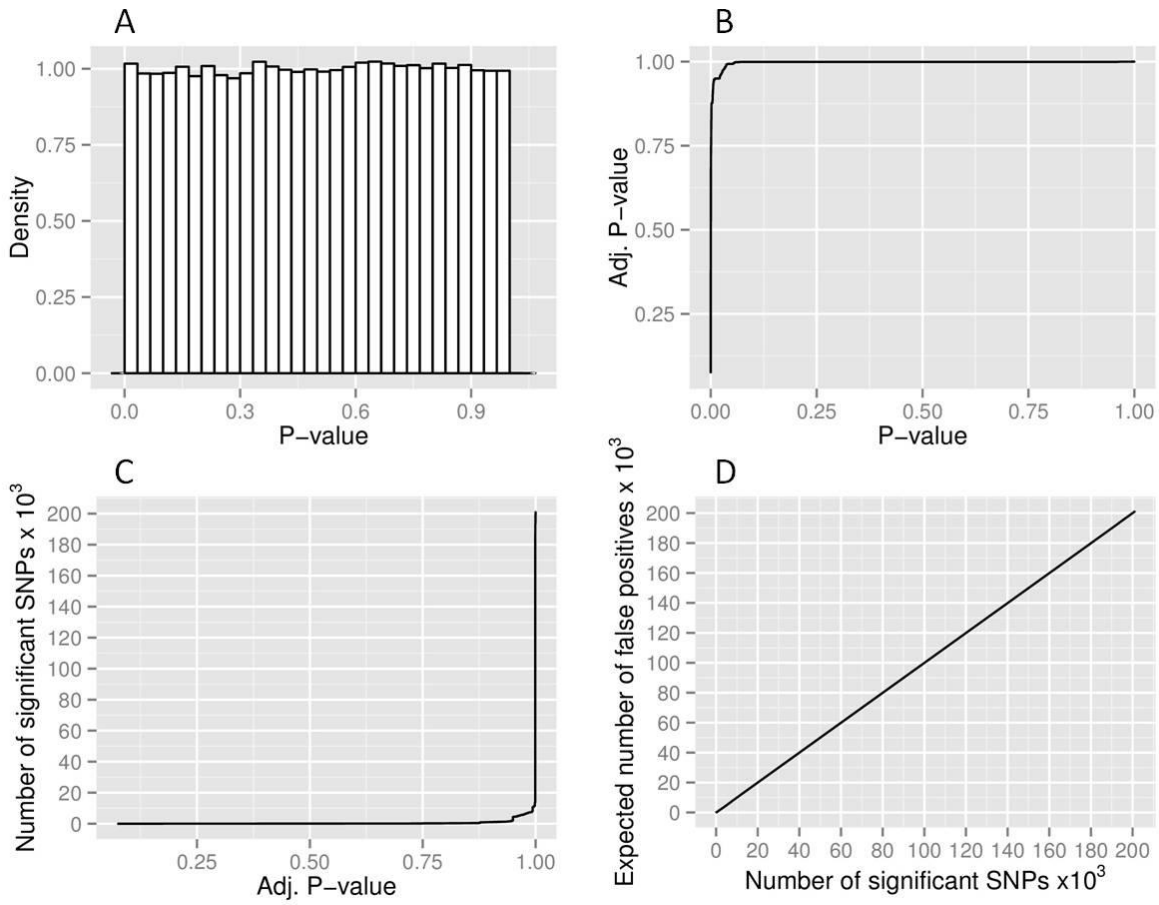


Figure B.3. Estimating the false discovery rate (FDR) for SNP marker association with Fusarium ear rot resistance in the filtered inbred association panel analysis. (A) A density histogram showing raw *P*-value distribution of 200,978 SNPs following GWAS. (B) The FDR-adjusted *P*-values plotted against their respective raw *P*-values. (C) The number of SNPs plotted against each of the respective FDR-adjusted *P*-value estimates. (D) The expected number of false positive SNPs versus the total number of significant SNPs given the FDR-adjusted *P*-values.

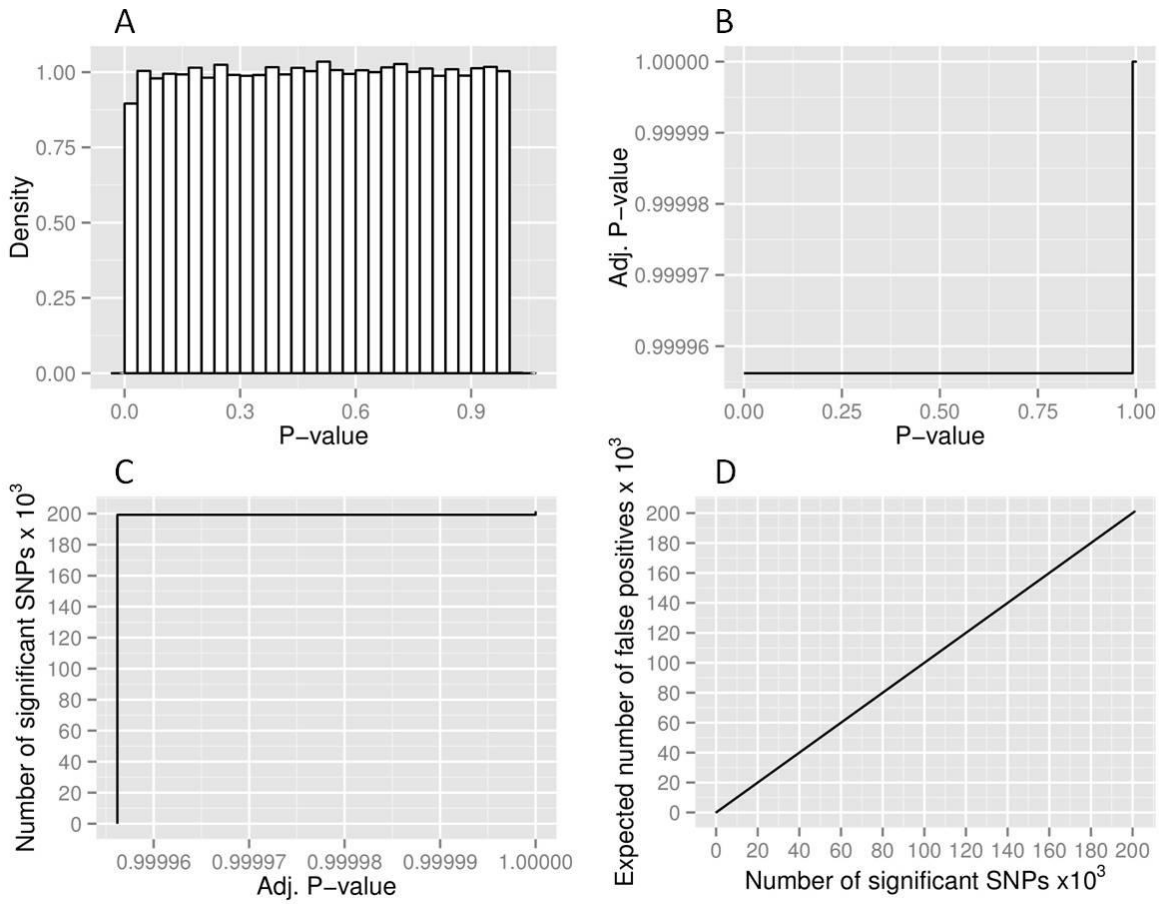


Figure B.4. Estimating the false discovery rate (FDR) for SNP marker association with Fusarium ear rot resistance in the B47 topcross analysis. (A) A density histogram showing raw *P*-value distribution of 200,978 SNPs following GWAS. (B) The FDR-adjusted *P*-values plotted against their respective raw *P*-values. (C) The number of SNPs plotted against each of the respective FDR-adjusted *P*-value estimates. (D) The expected number of false positive SNPs versus the total number of significant SNPs given the FDR-adjusted *P*-values.

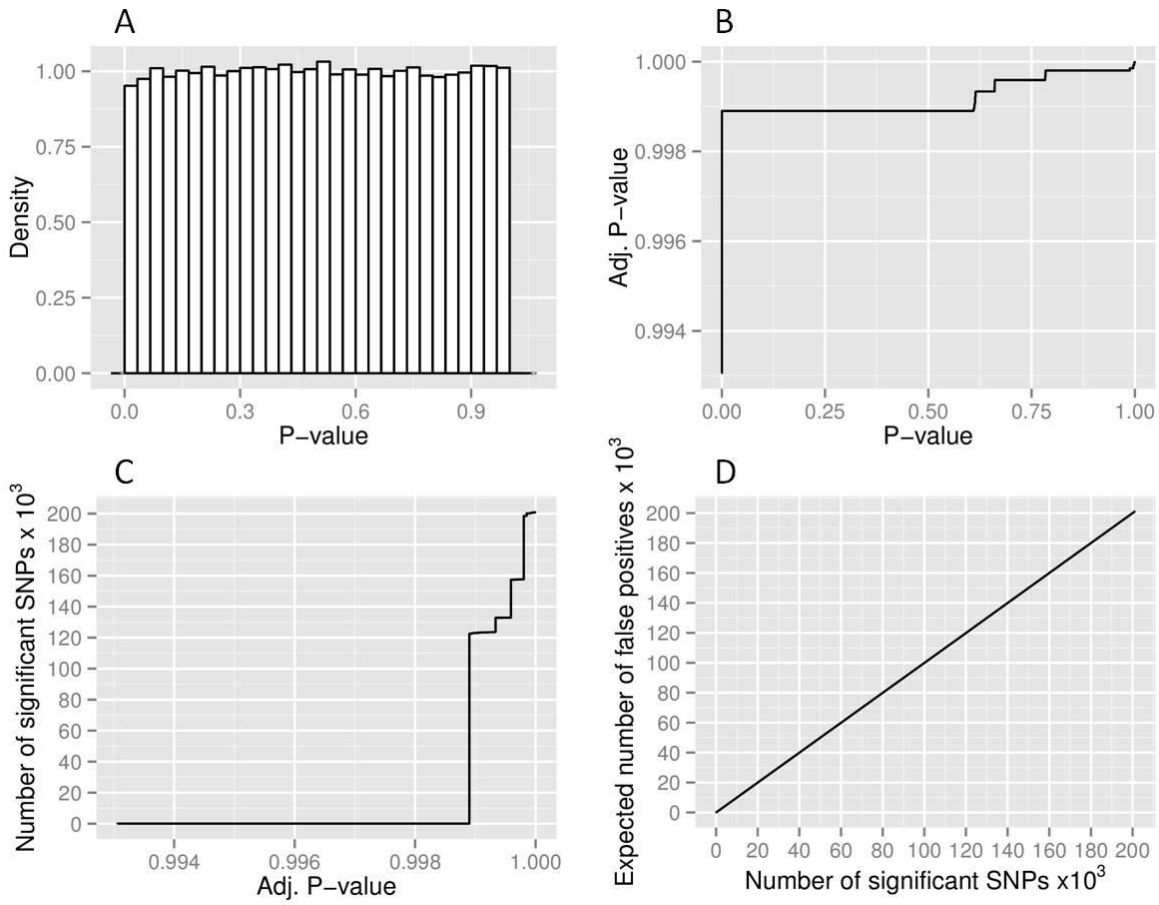


Figure B.5. Estimating the false discovery rate (FDR) for SNP marker association with Fusarium ear rot resistance in the PHZ51 topcross analysis. (A) A density histogram showing raw *P*-value distribution of 200,978 SNPs following GWAS. (B) The FDR-adjusted *P*-values plotted against their respective raw *P*-values. (C) The number of SNPs plotted against each of the respective FDR-adjusted *P*-value estimates. (D) The expected number of false positive SNPs versus the total number of significant SNPs given the FDR-adjusted *P*-values.

Files B.1-B.5

Supporting data

Will be available for download at date of publication from
http://www.panzea.org/lit/publication.html#2014

File B.1. Raw phenotypic data from the inbred association panel experiments in 2010-2012, formatted for spatial analysis in ASReml software. Columns in the data file are as follows from left to right: year (1=2010, 2=2011, 3=2012), row (field position of plot from front of field to the back), column (field position of plot from left to right), set, block, plot, inbred names (Material), anthesis date (DTA, converted to the number of days after planting until anthesis), silking date (DTS, converted to the number of days after planting until silking), Fusarium ear rot score (rot_AVG, averaged across ears within the plot), number of ears scored within each plot (earno), the natural log transformation of the average ear rot score (logrot), first through fourth order polynomial row trend effects (R1-R4), and first through fourth order polynomial column trend effects (C1-C4). Observations with the "MISSING" qualifier in the Material column are placeholders for the purposes of spatial analysis in ASReml.

File B.2. Raw phenotypic data from the topcross experiments in 2011 and 2012, formatted for spatial analysis in ASReml software. Columns in the data file are as follows from left to right: year (1=2011, 2=2012), row (field position of plot from front of field to the back), column (field position of plot from left to right), set, group, block, tester (1=PHZ51, 2=B47, 3=placeholder for check Pioneer 31G66, 4=placeholder for check NC478×GE440), maturity (1=early, 2=late), plot, hybrid names (Pedigree), inbred association panel parent of hybrid (Parent), anthesis date (DTA, converted to the number of days after planting until anthesis), silking date (DTS, converted to the number of days after planting until silking), Fusarium ear rot score (rot_AVG, averaged across ears within the plot), number of ears scored within each plot (earno), the natural log transformation of the average ear rot score (logrot), first through fourth order polynomial row trend effects (R1-R4), and first through fourth order polynomial column trend effects (C1-C4). Observations with the "MISSING" qualifier in the Material column are placeholders for the purposes of spatial analysis in ASReml.

File B.3. Genotypic data in HapMap format consisting of 200,978 SNP markers on the 2480 genotyped entries from across the inbred and hybrid experiments, compressed in .zip format due to size. Data have been filtered to remove SNPs with greater than 20% missing data and minor allele frequencies less than 0.05.

File B.4. Fusarium ear rot least square means for all 2480 genotyped entries from across the inbred and hybrid experiments, formatted for analysis in the R software package GAPIT. The Material column contains the line names, and the other columns are as follows: least square means for the full inbred association panel (Inbred_full), means for the filtered inbred

association panel (Inbred_filt), means for the B47 topcrosses (B47), and means for the PHZ51 topcrosses (PHZ51). Missing means are denoted by the "NA" qualifier.

File B.5. A 2480×2480 genetic kinship matrix (K) based on VanRaden (2008), formatted for analysis in the R software package GAPIT. The first column contains line names, and all other columns contain the pair-wise kinship coefficients between lines.

APPENDIX C: Protocols for Preparing Inoculum for Fusarium Ear Rot Experiments

Protocol for Preparing *Fusarium verticillioides* Cultures

Starter plates

1. Calculate the total number of Fusarium plates needed to inoculate all summer disease experiments using the "Inoculum Calculation Example.xlsx" data sheet (M:\File_Server\FieldBooks). This calculation ensures that enough inoculum is made for all experiments while accounting for multiple days of inoculation trips. If this data sheet is missing, the calculation is as follows (rounding all numbers up to the nearest whole number):
    a) Determine the total number of long rows (10-12 ears/plot) to be inoculated. Divide this number by two for short rows (5-6 ears/plot).
    b) Divide a) by 11.12 to get the total number of plates needed across all isolates and inoculations (assuming two inoculation trips per experiment).
    c) Divide b) by two to get the number of plates needed per inoculation trip across all isolates.
    d) Divide c) by six (if using six isolates) to get the number of plates needed per isolate per inoculation trip. It is important that any zeros are rounded up to one.
    e) Summarize results. Multiply d) by two to get the total number of plates needed per isolate overall, and then multiply that number by six to get the grand total number of plates needed overall.

2. One flask of media will pour between 12-15 plates. You can make all of the plates needed for both the starter plates and inoculation plates at once, or you can make a small batch for the starter plates and wait to make the larger batch for the inoculation plates until the starter plates are ready to transfer. If you opt for the former, unused PDA plates can be stored in a refrigerator sealed with Parafilm until needed. For each flask needed, dissolve 39 g of potato dextrose agar (PDA) in 1 L of ddH_2O in a 2000 mL Erlenmeyer flask. It may be helpful to microwave the flask for 5 minutes after shaking vigorously by hand. It is acceptable for a few clumps of undissolved PDA to be present before autoclaving.
3. Cover the mouth of the flask with a piece of aluminum foil. Autoclave on a liquid cycle for 30 minutes.
4. Disinfect all surfaces inside the safety hood with 70% ethanol. Using the silicon mitt, carefully pour the autoclaved PDA into 150 mm × 15 mm Petri plates. Only pour enough PDA to just completely cover the bottom of each plate.
5. Allow covered plates to cool in the hood overnight. It is preferred to turn off the hood fan to avoid the plates drying out.
6. Disinfect all surfaces inside the safety hood with 70% ethanol. Using the sterile method, streak plates with *F. verticillioides* isolates from glycerol stocks. There should be 2-3 starter plates per isolate.
7. Allow covered plates to grow until white mycelia completely cover the plate (1-2 weeks; Figure C.1).

Inoculation plates

1. Autoclave six spatulas/spoonulas in an aluminum foil packet on a short gravity cycle. There should be at least one sterile spatula per isolate.
2. Disinfect all surfaces inside the safety hood with 70% ethanol. Using a sterile spatula, cut the starter plates into 1 cm^2 sections.
3. Transfer three sections of the starter plate to a fresh PDA plate, arranging the sections opposite from one another near the edges of the plate.
4. Allow plates to grow for at least 2 weeks. Plates can be moved from the hood to a bench top to save space. Plates should be completely covered with white mycelia before using. Discard any plates that have any visual contamination. All discarded plates should be placed in the orange biohazard waste.

Protocol for Preparing a Live Suspension of *Fusarium verticillioides* Inoculum

Preparing the conidia suspension

1. Based on the "Inoculum Calculation Example.xlsx" data sheet, retrieve the needed number of inoculation plates per isolate.
2. Cut a square of cheesecloth large enough to cover the mouth of a plastic 1 L container. Secure the cheesecloth with a rubber band, leaving some slack in the middle of cheesecloth.
3. Copiously wet each plate with ddH_2O and scrape all mycelia onto the cheesecloth using a paintbrush. Use a squirt bottle of ddH_2O to rinse loose debris from the plate and

paintbrush onto the cheesecloth. All isolates should be combined together into a single suspension.

4. Remove the rubber band from the container, carefully gather the cheesecloth into a ball (making sure no agar debris to passes into the suspension), and wring all water from the plate debris into the suspension. Discard the cheesecloth/debris ball and spent inoculation plate in the biohazard waste.
5. Transfer inoculum suspension to a 1 L glass Pyrex jug (orange cap). Bring container to a volume of 1 L with additional ddH_2O. Shake thoroughly.

Standardization of conidia suspension

1. Draw off a 10 µL sample of the conidia suspension into a 1.5 mL microtubule. Add 90 µL of ddH_2O to the microtubule for a 1:10 dilution.
2. Load approximately 15 µL of the diluted conidia suspension onto a hemocytometer. Using a microscope, count the number of conidia in the four corner squares and the center square of the counting area (Figure C.2). Wipe the hemocytometer and cover slip with a Kimwipe both before and after counting. *Note: the Holland Lab is not equipped with a microscope. However, the Balint-Kurti Lab (Thomas Hall, room 2574) has a suitable microscope. Do not throw away cover slip – clean and reuse.*
3. Sum the count of the five squares and divide into 2000 (2000/(sum condia count over 5 squares)). This is the amount of the stock conidia suspension (in mL) needed to make a 1 L suspension at a concentration of 2×10^6 spores/mL.
4. Inoculum may be made the day before use. Store in a refrigerator overnight.

5. For use, prepare a 5 L volume of a 2×10^6 spores/mL in a modified Solo backpack sprayer. Add two drops of Tween-20 to every 5 L to help the condia disperse in solution. A 5 L volume will inoculate approximately 1000 ears before needing refilled. Plots should be inoculated approximately one week after flowering down the silk channel and one week after that near the base of the ear.

Figure C.1. Example of a PDA plate completely covered by mycelia of *Fusarium verticillioides*.

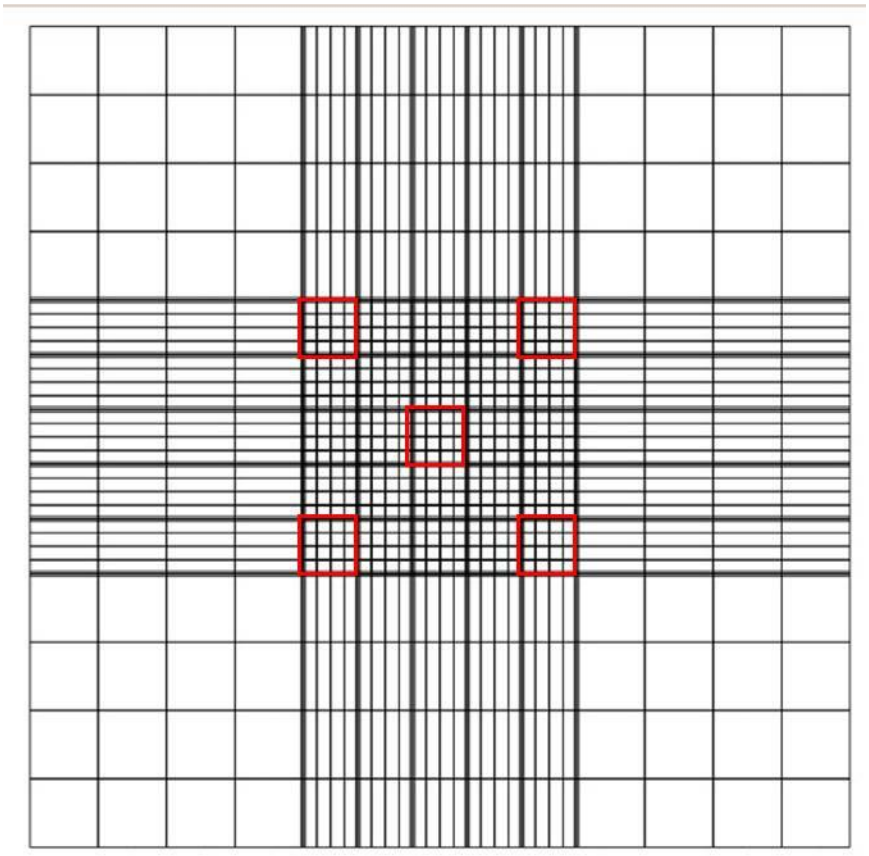


Figure C.2. Diagram of the grid area of a hemocytometer as viewed under a microscope. Conidia within the five grid areas highlighted in red are those that should be counted when making a stock solution of *F. verticillioides* inoculum.