GENETIC AND GENETIC BY ENVIRONMENT EFFECTS ON TAR SPOT RESISTANCE AND HYBRID YIELD IN MAIZE

By

Blake Trygestad

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Plant Breeding, Genetics and Biotechnology – Crop and Soil Sciences - Master of Science

2021

ABSTRACT

GENETIC AND GENETIC BY ENVIRONMENT EFFECTS ON TAR SPOT RESISTANCE AND HYBRID YIELD IN MAIZE

By

Blake Trygestad

The phenotype of any plant can be broken down into the three primary sources of variation, genetic (G), environment (E), and genetic by environmental interaction (GxE). Producers and researchers alike will harness repeatable G and GxE effects to maximize their resource efficiency. This study studied the G and GxE effects in the biotic stress of the fungi *Phyllachora maydis* and the environment patterns in advanced yield trial data. In rating 800 genotypes over two seasons, we genetically mapped and identified over 100 significant Single Nucleotide Polymorphisms (SNPs) associated with tar spot resistance using a genome-wide association study. We then conducted genomic prediction, which was 81.5% accurate for predicting tar spot severity within the location and 48% accurate in predicting disease resistance in a new environment. Also, using Genetic and Genotype x Environment (GGE) biplots, we investigated environmental patterns of nine locations in three maturity Zones in the advanced yield trials in the Michigan Yield Performance Trials. First, we identified two locations, one in the late and one in the mid maturity zone, with equal G and GxE effects and should be removed. Then, using a sliding window of year combinations, we analyzed the optimal number of replications needed across the three maturity zones. We determined that an average of three replications are needed to achieve 75% of the maximum repeatability across the zones.

ACKNOWLEDGMENTS

I would first like to acknowledge and thank Dr. Addie Thompson as my advisor and PI. Dr. Thompson took on a prototypical farm kid as a master's student, and I could not have asked for a better person to help guide me in my research and studies. I know without a doubt that I am not only a better scientist but a better person from being under her guidance and wisdom. I also would like to thank Dr. Martin Chilvers and Dr. Maninder Singh for allowing me to bounce ideas off you all and making me more well-rounded in all agricultural sciences.

I would also like to thank all the members of the Thompson Lab for helping me in the field and on the computer keyboard. A special shout to Sidney Sitar for putting up with my terrible corn(y) jokes and being my right hand through most of this project; I know the next steps in the research are in great hands. I also would like to thank Robert Shroute. Not only am I thankful for our "office hours" where I could pick your brilliant mind on the dumbest of questions, but I am also most thankful for your friendship. I wish you the best and hope to stay in contact with you for a long time.

Thirdly, I would like to thank my family. Your love and support throughout my life have grounded me in the best of ways. I know you all are only a quick phone call away and would be on a plane to see me in a heartbeat if need be. I would not be the man I am today without you all.

Finally, and above all, I would like to thank God for his providence. I am not deserving of the chance to study the utterly complex and beautiful world You have created. In the end, my work and I are only temporary and will someday be forgotten. Nevertheless, I will always be remembered and loved by you, God, and I am honored to glorify You in the years ahead.

TABLE OF CONTENTS

LIST OF TABLES ..... **vii**

LIST OF FIGURES ..... **viii**

CHAPTER 1: REVIEW OF GENETIC AND GENETIC ENVIRONMENTAL INTERACTION TOOLS **1**

ABSTRACT 1
INTRODUCTION TO TAR SPOT RESEARCH 2
TAR SPOT SYMPTOMATOLOGY 2
DISEASE CYCLE 3
DISEASE DISTRIBUTION 3
GENETIC HOST RESISTANCE 4
DIVERSITY PANELS 5
GENOME-WIDE ASSOCIATION STUDY (GWAS) 7
GENOMIC PREDICTION 8
AREA UNDER DISEASE PROGRESS CURVE (AUDPC) 9
INTRODUCTION TO ANALYSIS OF CORN PERFORMANCE TRIALS 10
ENVIRONMENTAL ANALYSIS 11
REPLICATION ANALYSIS 13
CONCLUSION 15

CHAPTER 2: GENETIC MAPPING AND PREDICTION OF TAR SPOT (CAUSED BY *PHYLLACHORA MAYDIS*) RESISTANCE IN MAIZE **16**

ABSTRACT 16
INTRODUCTION 17
MATERIALS AND METHODS 19
PLANT MATERIAL 19
EXPERIMENTAL DESIGN 19
PHENOTYPIC DATA ANALYSIS 21
GENOTYPIC ANALYSIS AND GWAS 23
IDENTIFICATION OF CANDIDATE GENES 23
GENOMIC PREDICTION 24
RESULTS 25
MICHIGAN 2019 25
MICHIGAN 2020 27
GWAS 28
GENOMIC PREDICTION 31
DISCUSSION 33

CHAPTER 3: OPTIMIZING USE OF RESOURCES IN CORN PERFORMANCE TRIALS BY ANALYZING GXE INTERACTIONS AND THE NUMBER OF REPLICATION 36
ABSTRACT 36
INTRODUCTION 37
MATERIALS AND METHODS 39
MICHIGAN CORN PERFORMANCE TRIALS (MCPT) 39
STATISTICAL MODELS 41
GGE BIPLOTS 41
REPLICATION ANALYSIS 42
OUTLIER DETECTION 43
RESULTS 43
PEARSON CORRELATION PLOTS 43
GGE BIPLOT ANALYSIS 44
OPTIMAL REPLICATION NUMBER 49
DISCUSSION 50

APPENDIX 53

BIBLIOGRAPHY 72

LIST OF TABLES

Table 1.1: Summary Statistics of WiDiv Final Rating Per Environment 25

Table 1.2: Significant SNPs Per Chromosome Per Trait from MI 2020 GWAS 30

Table 1.3: Significant SNPs Per Chromosome Per Trait from IN 2020 GWAS 30

Table 1.4: Significant SNPs Per Chromosome from WI 2020 GWAS 31

Table 2.1 A-C: Number of Hybrids per Subset 40

Table 2.2 A-B: Average Angle of Location Combinations 45

Table 2.3: Angle between within zone location across years 46

Table 2.4: Angle between zone location across years 46

Table 2.5: Angle Difference between Subsets 47

Table 2.6: Optimal number of replications needed at each location 49

Table 2.7: Optimal number of replications needed at each Zone 50

Table A.1: Inbred Names 54

Table A.2: All Significant SNPs 61

Table A.3: Genes Located Near Significant SNPs 63

Table A.4: Genes That Showed a Change in Expression Due to Disease 66

Table A.5: Genomic Prediction of Trait Per Algorithm 67

Table A.6: Genomic Prediction of Trait Per Algorithm Per SNP Level 67

Table A.7: Genomic Prediction of Observed Genotypes in New Environment / SNP Level 69

Table A.8: Genomic Prediction of Unobserved Genotypes in New Environment / SNP Level ...70

Table B.1: Number of Hybrids at each year and environment combination 71

Table B.2: Number of Hybrids at each year and Zone combination 71

LIST OF FIGURES

Figure 1.1: Disease Rating Scale 20

Figure 1.2: Disease Incidence by Population in MI 2019-2020 26

Figure 1.3: Correlation Heat Map Showing Relationship Among the Traits Collected 27

Figure 1.4: A-B: Distribution of Area Under Disease Progress Curve (AUDPC) And Final Rating by Population In 2020 28

Figure 1.5: A-B: Manhattan Plot for GWAS Result in MI 2020 For AUDPC and Final Rating ..29

Figure 1.6: Manhattan Plot for GWAS Result in IN 2020 For AUDPC 29

Figure 1.7: Genomic Prediction Accuracy of Traits Using Different Algorithms 31

Figure 1.8: Genomic Prediction of Final AUDPC in MI 2020 Using Different Algorithms 32

Figure 1.9: Genomic Prediction of Unobserved IN 2020 AUDPC from MI 2020 AUDPC 33

Figure 2.1: Locations of MCPT Trails.....37

Figure 2.2: Correlation Heatmap of County Combinations 44

Figure 2.3 A-B: GGE Biplot of Between Zones Across All Years 48

Figure A.1: Quantile-Quantile-Plot of AUDPC6 GWAS.....60

CHAPTER 1: REVIEW OF GENETIC AND GENETIC ENVIRONMENTAL INTERACTION TOOLS

ABSTRACT:

The phenotype of any plant can be broken down into the three primary sources of variation, genetic (G), environment (E), and genetic by environmental interaction (GxE). Producers and researchers alike will harness G and GxE repeatable effects to maximize their resource efficiency to get the most out of their resources. This study studied the G and GxE effects in the biotic stress of the fungi *Phyllachora maydis* and the environment patterns in advanced yield trial data. Tar spot is a new and rapidly spreading disease of maize in the United States caused by the Ascomycota fungus *Phyllachora maydis*. The pathogen infects maize leaves, creating black lesions that can lead to the premature death of the plant. This study identified genetic resistance to the fungus using a genome-wide association study and used genomic prediction models to predict the disease severity in new genotypes and environments. Also, using G and GxE (GGE) biplots, we investigated the environmental patterns of nine locations in three maturity zones within the Michigan Corn Performance Trials. Then using a sliding window of year combinations, we analyzed the optimal number of replications needed across the three maturity zones.

INTRODUCTION TO TAR SPOT RESEARCH

Tar spot is a foreign and rapidly spreading disease of maize (*Zea mays L*) in the United States caused by the fungus *Phyllachora maydis*, an ascomycete and obligate plant parasite. While initially identified in Mexico in the early 20th century (Maublanc, 1904), the fungus was constrained to Central and South American countries (Bajet et al. 1994) until 2015 where researchers discovered the fungus in the United States of America (Ruhl, 2016). Since 2015, researchers have confirmed tar spot in ten states and Ontario, Canada (Ruhl, 2016; McCoy et al. 2018; Dalla Lana et al. 2019; Malvick et al. 2020, Tenuta et al. 2020).

TAR SPOT SYMPTOMATOLOGY

The disease tar spot is identified by the stromata, or fruiting bodies, of *P. maydis*. These stromata are where the common name “tar spot” comes from as the stromata are raised hard black lesions that look like tar speckled on both sides of the leaves (Liu, 1973). Often common in Latin America, but not in the United States, a necrotic halo surrounds the stromata known as "fisheye lesions." These fisheye lesions can fuse, causing leaf necrosis and leading to the plant's premature death (Ceballos and Deutsch 1992; Hock et al. 1995; Carson, 1999).

Several studies of Latin American strains have suggested that the pathogenicity of *P. maydis* can be enhanced with another fungus, *Monographella maydis* (Müller and Samuels, 1984; Ceballos and Deutsch 1992; Hock et al., 1991). According to these studies, *M. maydis* by itself will not damage the plant (Müller & Samuels 1984; Hock et al., 1991), but with coinfection with *P. maydis*, *M. maydis* can cause severe necrosis of the plant’s foliage, leading to yield loss (Ceballos and Deutsch 1992 & CIMMYT, 2003). Despite this, in the United States, fields infected with *P. maydis* have not contained *M. maydis* and have yet sustained substantially

damaged plant yields, suggesting that the fungus is unnecessary for fisheye lesions to occur in the United States. (Ruhl et al., 2016; McCoy et al., 2019).

DISEASE CYCLE

While the disease cycle of tar spot is mainly uncharacterized, it is known that the spores of *P. maydis* can overwinter on dead residue from the previous year's crop with no alternative host. (Mottaleb et al., 2018; Groves et al., 2020). In the Upper Midwest, the ascospores of *P. maydis* have survived on residue in winter temperatures below -30^{o}C (Kleczewski et al., 2019; Groves et al., 2020).

After the initial infection, the stomata will form and release spores to infect the new foliage of neighboring plants, exponentially increasing over time. While variable according to the growing degree days and the plant's resistance (Precigout, 2020), symptoms typically show 14 days post-infection, and spores are produced soon after (Hock et al., 1995). Once established, *P. maydis* can infect any exposed foliage (leaves, husks, or sheaths) of any plant age; however, the fungus most commonly appears before the flowering of maize, in early July (Bajet et al., 1994; Hock et al., 1995).

DISEASE DISTRIBUTION

While *P. maydis* is native to parts of Central and South America, in 2015, the fungus was identified in the United States in Indiana and currently has spread to ten states: Illinois, Iowa, Indiana, Minnesota, Michigan, Missouri, Ohio, Wisconsin, Pennsylvania, Florida and in Ontario, Canada (Ruhl, 2016; Ruhl et al., 2016; McCoy et al., 2018; Dalla Lana et al., 2019; Malvick et al. 2020, Tenuta et al. 2020). Researchers debate *P. maydis*'s introductions to the United States, however despite researchers believing that *P. maydis* is not seed-borne, typically, diseases and

pests are accidentally imported by internationally traded plants and plant products (Huber et al., 2002).

GENETIC HOST RESISTANCE

Currently, growers most often manage fungal diseases through fungicide applications and resistant hybrids. Although there are fungicides that affect tar spot, they are expensive to apply and only slow the spread after infection occurs. Conversely, host resistance can prevent infection and is standard for foliar diseases management. For another ascomycete in maize, Northern leaf blight (*Setosphaeria turcica)*, the Ht genes have been providing resistance to specific races of the fungus since their discovery in the 60s and 70s (Hooker, 1963 & 1977) and providing partial polygenic resistance to all races of the fungus (Hooker, 1973). Geneticists have also identified genetic resistance for foliar diseases such as southern corn leaf blight (Kump et al. 2011) and gray leaf spot (Shi et al. 2014; Kuki et al. 2018). Therefore, developing highly resistant temperate lines for tar spot will be crucial to prevent future losses.

Early studies using three segregating bi-parental populations in tar spot resistance established resistance to be highly heritable and dominant (Ceballos and Deutsch, 1992). More recently, however, tar spot resistance has been perceived as a complex multi-gene-controlled resistance trait, with a single-large effect locus and a few minor quantitative trait loci (QTL) (Mahuku et al., 2016; Cao et al., 2017).

A large-effect QTL, named qRtsc8-1, has been detected on chromosome 8 bin three across tropical populations screened in Central and South America (Mahuku et al., 2016; Cao et al., 2017). In these studies, qRtsc8-1 accounted for 18-43% of the observed phenotypic variation (Mahuku et al., 2016; Cao et al., 2017). In addition, this discovery identified several haplotypes that increased resistance to tar spot in tropical materials (Mahuku et al., 2016).

In temperate hybrids, Telenko et al. (2019) assessed current Midwestern United States hybrids for resistance. According to this study, all the hybrids evaluated were susceptible to tar spot, with stromata infection ranging from 1–50% with an estimated 0.32–1.36 bu/A (21.5 to 91.5 kg/ha) loss of yield per 1% increase in tar spot lesion coverage (Telenko et al., 2019).

GENETIC DIVERSITY PANELS

Diversity panels are helpful when assessing natural variation for complex traits such as disease resistance. Large panels such as the CIMMYT panel (Wu et al., 2016) have been trimmed to certain phenologies to increase the panel's utility in specific environments. While maintaining as much diversity as possible, these smaller panels are restricted in specific ways to make more tailored and valuable conclusions on traits of interest.

Wisconsin Diversity Panel

The Wisconsin Diversity panel-942 (WiDiv-942) is a diverse group of 942 inbred lines, from the public sector, privately expired Plant Variety Protection (exPVP), and the Germplasm Enhancement of Maize project (GEM), with restricted phenology to the northern U.S. Corn Belt. Researchers expanded the WiDiv-942 from a smaller panel of 627 inbreds, the WiDiv, to now contain four groups of stiff stalks (B37, B73, B14, and BSSSC0), two groups of non-stiff stalk (Mo17 and Oh43), an Iodent, popcorn, sweet corn, and tropical populations (Mazaheri et al., 2019).

In 2014, Hirsch et al. (2014) enhanced the original WiDiv panel's capability by performing RNA sequencing on 504 seedlings and identified 451,066 Single Nucleotide Polymorphisms (SNPs). Subsequently, using whole seedlings, Mazaheri et al. (2019) conducted RNAseq on the expanded WiDiv-942, identifying 899,784 SNPs in the WiDiv-942 panel. Scientists have also used both the previous panel and its successor in numerous genetic research

projects ranging from flowering time (Hansey et al. 2011), vegetative phase changes (Hirsch et al. 2014), stalk biomass (Mazaheri, 2019), Sugarcane mosaic virus resistance (Gustafson et al., 2018), and dramatic male inflorescence (Gage et al., 2018).

Genetic Enhancement of Maize (GEM)

The Genetic Enhancement of Maize (GEM) project is a collaboration between the United States Department of Agriculture and many public and private institutions. The project's goal is to "effectively increase the diversity of U.S. maize germplasm utilized by producers, global end-users, and consumers" (Pollak, 2003). They hope to accomplish this goal by backcrossing exotic germplasm with temperate material to gain genetic diversity from the world and mature in temperate regions.

To make GEM lines, one private cooperating company crosses an exotic line with a private inbred to make a 50% exotic breeding cross. Then another private cooperator crosses the 50% cross with their own inbred of the same heterotic group to generate a 25% exotic breeding cross (Pollak, 2003). Although these GEM lines will segregate, they carry genetic diversity not usable otherwise. Within the GEM program, double haploid of the backcrossed lines, BGEMS, are used frequently and do not segregate like the backcrossed material.

The GEM lines are popular with geneticists throughout maize research. The GEM program itself studies phenotypic traits of grain composition, starch quality, and oil content. The program also evaluates resistance to various significant maize pests such as European corn borer (Abel et al., 2001), corn rootworm, gray leaf spot, Stewart's wilt, anthracnose stalk rot, fusarium ear rot resistance, virus resistance, among many more (Pollak, 2003).

GENOME-WIDE ASSOCIATION STUDY (GWAS)

The first genome-wide association study (GWAS) was first completed by Ozaki et al. (2002) when finding single nucleotide polymorphisms (SNPs) associated with susceptibility to myocardial infarction in humans. In 2008, Belo et al. used GWAS on 553 maize inbreds to explore the genes affecting fatty acid content in kernels, and this method of genetic mapping became routine after the release of the B73 reference genome (Schnable et al., 2009). With the advances in next-generation sequencing technologies, GWAS using diverse germplasm sets has been an essential tool for researching genetic variation of maize traits (Xiao et al., 2017). For association mapping, geneticists test each maker for an association with a trait of interest. The assumption is that associations will arise because the SNPs will be in linkage disequilibrium with the genetic regions contributing to a trait. (Huang & Han, 2014)

It is essential to avoid confounding effects in GWAS, accounting for population structure such as co-ancestry of families, adaption to local conditions, and inbreeding/genetic drift/admixture. A mixed model approach by Yu et al. (2005) is common to control these factors by forming a kinship matrix from pedigree information (Bernardo, 1993) and using Principal Component Analysis (PCA) to reduce the genotypic data's dimension. This model then can devise a covariate to help control the population structure and reduces random associations (Price et al., 2006).

In order to find causal variation for complex traits, numerous models have been designed to identify the variation held within the population structure. In Fast-LMM-Select (Listgarten et al., 2012) and Settlement of MLM Under Progressively Exclusive Relationship (Wang et al., 2014), the subsetted markers associated with the trait determine kinship. The Multi-Locus Mixed-Model (Segura et al., 2012) uses the markers most associated with the trait of interest, stepwise, as covariates to test multiple markers simultaneously. The Fixed and Random Model

Circulating Probability Unification (FARM-CPU, Liu, et al., 2016) assembles a fixed effect and a random effect model. Then using maximum likelihood, researchers use the markers to remove kinship in the fixed model, and the random model predicts associations until two consecutive iterations leave the number of associations unchanged.

GWAS has been used to inspect the genetic composition of many complex traits in maize, including flowering time (Buckler, 2009), leaf architecture (Tian et al., 2011), stalk biomass (Mazaheri et al., 2019), and disease resistance (Poland et al., 2011).

GENOMIC PREDICTION

In 2001, Meuwissen et al. proposed using all available markers collectively to build a prediction model to predict an individual's genomic estimated breeding value (GEBV) for a population rather than their significance level. This method can establish unbiased and accurate marker effects for early generational testing without phenotypic data in planted field trials. Furthermore, empirical and simulated genomic prediction studies have shown that GEBV prediction accuracies are ample to achieve rapid gains in early selection (Meuwissen et al., 2001; Lorenzana and Bernardo, 2009; Jannink et al., 2010).

Implementing Model

To begin implementing genomic prediction, users must first construct a training population to build the model. This material should be related to the testing population and requires genome-wide marker genotypes and phenotypic values of the trait of interest. Modelers will take the phenotypic and genotypic data and place them in a modeling software program. These software programs will build a prediction model, and researchers then perform cross-validation on the training set.

After cross-validation, genomic marker data of related material is implemented in the prediction model to predict the new lines' GEBVs, which researchers can use to make selections on the material without needing a phenotype.

Genomic Models

While the goal of estimating breeding values for traits using genome-wide marker sets is the same, the assumptions of each model type are different. There are two major types of regression models: Nonparametric (Random Forest etc.) and parametric, which include penalized approaches (rrBLUP, gBLUP, support vector regression, etc.) and also Bayesian approaches (Bayes A Bayes B, BRR, etc.)

The best approach for genomic prediction depends on the genetic architecture of the trait (Bernardo, 2008). Ridge regression best linear unbiased prediction (rrBLUP) assumes that markers have a random nonzero effect with equal variances, which, in general, is best suited for traits controlled by many loci, each with a small effect (Meuwissen et al., 2001; Lorenz et al., 2011). On the other hand, Bayesian models do not assume all markers have a nonzero effect and estimate a separate variance for each marker, following a prior distribution, and therefore are generally better for locating large effect QTLs (Meuwissen et al., 2001). Individually, the Bayes B model allows variances to be zero for prior distribution, while the Bayes A model only allows variances to approach zero (Meuwissen et al., 2001).

AREA UNDER DISEASE PROGRESS CURVE (AUDPC)

The Area Under Disease Progress Curve (AUDPC) is a quantitative summary of disease pressure over time (Shaner & Finney, 1977). This method is standard in pathology resistance studies to compare management tactics on a quantitative scale versus the highest infection rate for that tactic (Jeger & Vilijanen-Rollinson, 2001; Prabhu et al., 2011; Sakr, 2019). The

trapezoidal method (Campbell & Madden, 1990) is most commonly used as it calculates the average disease pressure between each pair of time points using the formula:

$$AUDPC = \sum_{i=1}^{n-1}\left(\frac{y_i+y_{i+1}}{2x(t_{i+1}+t_i)}\right.$$

Where y_i is the percent tar spot severity at the *ith* observation, t_i is the time in days after infection of the *ith* observation, and n is the total number of observations.

INTRODUCTION TO ANALYSIS OF CORN PERFORMANCE TRIALS

Crop variety trials are a common occurrence in variety testing across the world. These trials provide information to a breeder for releasing new varieties and help growers compare current varieties' performance. For example, the Michigan Corn Performance Trials (MCPT) for corn provides unbiased, third-party information on commercial hybrid performance across multiple locations every year. Michigan growers use the data collected from the MCPT to decide which commercial hybrids perform best for their cropping environment.

Though these trials produce invaluable data, they are resource-intensive, requiring many locations and replications to achieve accurate performance data. To counter this cost, researchers have conducted many studies investigating the best allocation of resources by changing the number of locations planted, replications at each location, or years planted (Sprague and Federer, 1951; Wricke and Weber, 1986; Swallow and Wehner, 1989; Zhou et al., 2011).

Weikai Yan et al. has conceptualized and tested two methods of best allocation of resources. One concept, GGE biplots, are graphical representations of the genetic effect and genetic by environmental effect (Yan: et al. 2000, & Kang 2003, & Tinker 2006, et al. 2007, & Fregeau-Reid 2008, & Holand 2010, et al. 2013, et al. 2014). These biplots can compare the environments to visualize similarities and differences. In addition, Yan et al. (2015 & 2021) have

worked on finding the optimal number of replications needed to reach a broad sense heritability level.

With climate change occurring worldwide, checking the integrity of maturity environment zones is critical to maintaining target regions. In addition to checking the accuracy of the maturity environment zones, it is crucial to identify discriminating environments within these zones to match the different environments seen within the maturity zones. These together can identify superior hybrids for regional applications while conserving resources.

It is also apparent that while the number of locations and the years planted are changeable, mature programs will often have a set number of test locations and want to avoid extending the testing period. This reality makes reducing replications at each location an excellent potential target for increasing test efficiency and optimal resource allocation.

To maintain high resource allocation and high-efficiency testing, maintaining non-redundant, discriminative environments along with the optimal number of replications is critical. This research uses Yan et al. methodologies on maize data from the MCPT to maximize testing efficiency. Similarly, GGE biplots are used to compare the environments over the years while using the replication analysis to see how many replications are needed to get the best data.

ENVIRONMENTAL ANALYSIS:

Proper selection of environments for a given crop variety trial is vital. Any trait (such as yield) can be broken down into three main effects of genotype (G), environment (E), and genotype by environment interactions (GxE). Researchers must test identical genotypes in multiple environments and compare their performances to parse out these effects. Optimally, these test environments are representative of a target region while avoiding costly redundancy in the resultant data.

In 2001, Yan et al. set out to biplot the G and GXE effects to compare environments to each other. Since then, GGE biplots have been growing in popularity to compare environments to devise mega-environments and find which cultivars are most productive in each environment type. They have been used in wheat (Thomason & Phillips, 2006), cotton (Blanche, 2006), soybean (Dalló et al. 2019), and breeding and hybrid selection in maize hybrids (Oyekunle et al., 2017; de Oliveira 2019).

Biplots were conceptualized by K.R. Gabriel (1971) as multivariate data shown in two-dimensional space. Biplots are built using the first two principal components of effects, and GGE-biplots are formed when the main environment effect is removed from multi-environmental trial data. As discussed above, a phenotype can be broken into the main effects of genotype (G), environment (E), and the GxE interaction. Removing the not reproducible E effect leaves only the genotype main effect and the GxE interaction effect, which can be graphically displayed in a two-way table (Yan and Kang, 2003). A singular-value decomposition is conducted on environment-centered mean grain yield to obtain the principal components, allowing researchers to focus on the reproducible variation of the trait of interest (Yan, 1999; Yan et al., 2000; Yan and Tinker, 2006).

In GGE biplots specifically, the biplot model proposed by Yan and Kang (2003) was:

$$Y_{ge} - \bar{Y}_e = \lambda_1 \xi_{g1} \eta_{e1} + \lambda_2 \xi_{g2} \eta_{e2} + \varepsilon_{ge}$$

Where $\boldsymbol{Y_{ge}}$ is the mean yield of the $\boldsymbol{g}$th genotype in the $\boldsymbol{e}$th environment; $\boldsymbol{\overline{Y_e}}$ is the mean yield across all genotypes in the $\boldsymbol{e}$th environment; $\boldsymbol{\lambda_1}$ and $\boldsymbol{\lambda_2}$ are the singular values for PC1 and PC2; $\boldsymbol{\xi_{g1}}$ and $\boldsymbol{\xi_{g2}}$ are the PC1 and PC2 eigenvectors for the $\boldsymbol{g}$th genotype; $\boldsymbol{\eta_{e1}}$ and $\boldsymbol{\eta_{e2}}$ are the

PC1 and PC2 eigenvectors for the $\boldsymbol{e}$th environment; and $\boldsymbol{\varepsilon_{ge}}$ is the residual of the model associated with the $\boldsymbol{g}$th genotype in the $\boldsymbol{e}$th environment.

This biplot allows for a comparative analysis between genotypes and environments by comparing the angle between two points on the biplot. An obtuse angle infers a negative correlation between the points, while an acute angle infers a positive correlation between them, and a 90^{o} angle between the points infers no correlation.

REPLICATION ANALYSIS

While the number of locations and the years planted are changeable, mature programs will often have a set number of test locations and want to avoid extending the testing period. This reality makes reducing replications at each location an excellent potential target for increasing test efficiency and optimal resource allocation.

Yan et al. (2015) explored using the breeder's equation to get the optimal number of replications needed to reach a broad sense heritability threshold. Yan et al. (2015) adapted the H equation calculated by DeLacy et al. (1996)

$$H = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_e^2}{r}}$$

moreover, reworked it to get the optimal number of replications at one location:

$$r = \frac{\sigma_e^2}{\sigma_g^2} * (\frac{H}{1-H})$$

Where H is the broad-sense heritability, σ_g^2 is the variance of genotypes, σ_e^2 is the variance of error, and r is the number of replications.

Yan (2021) tested his concept further to account for multi-location and multi-location, and multi-year data.

Single-Year Multi-Location:

$$r = \frac{\sigma^2_{e,ML}}{l * \sigma^2_{g,ML}} * (\frac{H_{ML}}{1 - \frac{H_{ML}}{H_{MML}}})$$

Where $\sigma^2_{g,ML}$ is the genotypic variance, $\sigma^2_{e,ML}$ is the experimental error variance based on the single year, multi-location trial, l is the number of locations, H_{ML} is the heritability threshold, and H_{MML} is the maximum achievable across-location heritability:

$$H_{MML} = \frac{\sigma^2_{g,ML}}{\sigma^2_{g,ML} + \frac{\sigma^2_{gl}}{l}}$$

Where σ^2_{gl} is the variance for the interaction of genotype by location.

Multi-Year Multi-Location:

$$r = \frac{\sigma^2_{e,MLY}}{l * y * \sigma^2_{g,MLY}} * (\frac{H_{MLY}}{1 - \frac{H_{MLY}}{H_{MMLY}}})$$

Where $\sigma^2_{g,MLY}$ is the genotypic variance, $\sigma^2_{e,MLY}$ is the experimental error variance based on the multi-year, multi-location trial, l is the number of locations, y is the number of years, H_{MLY} is the heritability threshold, and H_{MMLY} is the maximum achievable across-location heritability:

$$H_{MMLY} = \frac{\sigma^2_{g,MLY}}{\sigma^2_{g,MLY} + \frac{\sigma^2_{gl}}{l} + \frac{\sigma^2_{gy}}{y} + \frac{\sigma^2_{gly}}{ly}}$$

Where σ^2_{gl} is the variance for the interaction of genotype by location, $\sigma^2_{gy}$is the variance for the genotype by year interaction, and σ^2_{gly} is the variance for the three-way interaction of genotype, location, and year.

Yan et al. concluded that:

1. A goal repeatability level of 75% of the maximum repeatability is ideal to find the optimal number of replications as 75% is the upper limit repeatability can be improved by increasing the number of test environments/replications (Yan et al., 2015).
2. Cross-location analysis should be used to determine the optimal level of replicates (Yan 2014). A single trial basis often overestimates the number of replications needed (Yan 2021).
3. It is inferred that with an increase in test locations, replications needed at each location may decrease; however, excessive replications do not improve cross-location heritability (Yan 2021).

CONCLUSION

The analysis of the G and GxE effects is critical to having plants that have optimal production. In tar spot resistance, the genetic (G) basis of said resistance in temperate material is largely unknown, along with the magnitude of GxE interaction. In crop variety trials, it is the G and GxE effects that growers are most interested in, as these effects are repeatable and therefore controllable. Researchers must fill in these areas of research, as it will not only help growers be more profitable but also feed the world.

CHAPTER 2: GENETIC MAPPING AND PREDICTION OF TAR SPOT (CAUSED BY *PHYLLACHORA MAYDIS*) RESISTANCE IN MAIZE

ABSTRACT:

Tar spot is a new and rapidly spreading disease of maize in the United States caused by the Ascomycota fungus *Phyllachora maydis*. The pathogen infects maize leaves, creating black lesions that can lead to premature death. Although several genetic loci influencing tar spot's susceptibility have been observed in tropical maize genotypes, this is the first study to identify genetic loci contributing to tar spot resistance in temperate materials for U.S. production. Over two seasons in Michigan, 600 genotypes from the Wisconsin Diversity panel and 200 genotypes from Iowa State's Germplasm Enhancement of Maize program were screened. A genome-wide association study was conducted to map resistance, after which the predicted gene regions were used in genomic prediction models. Repeatability for disease resistance ratings ranged from 52.8-67.0% for Michigan fields, and ratings were not associated with flowering time, plant height, or ear height. Over 100 significant SNPs were associated with tar spot resistance, linked to candidate genes that will require further study. None of these SNPs were identified previously in tropical maize germplasm (Cao et al., 2017). Genomic prediction using Bayes B was 81.5% accurate for predicting tar spot severity, and high accuracy (65-75%) was maintained using very small sets of 10 or 20 markers. Using Bayesian ridge regression (BRR), the model was 48% accurate at predicting disease progression in a new environment. Together, these results will help plant breeders develop hybrid maize with lower yield losses due to tar spot infection.

INTRODUCTION

Tar spot is a new and rapidly spreading disease of maize in the United States caused by the fungus *Phyllachora maydis,* an ascomycete and obligate plant parasite. In 2015, maize producers reported lesions caused by the fungus in two counties in Indiana and Illinois (Ruhl 2016). Before 2015, *P. maydis* was restricted to Mexico and Central and South American countries. Since the initial documentation in the U.S., tar spot has been confirmed in ten states and Ontario, Canada (Ruhl 2016; McCoy et al. 2018; Dalla Lana et al. 2019; Malvick et al. 2020, Tenuta et al. 2020).

The tar spot stromata embed in the plant foliage and rapidly kills the plant tissues. A severe infection leads to the rapid blighting of the canopy, early senescence, shriveled kernels, smaller ears, and 50% yield loss per field (Telenko et al. 2019; Mueller et al. 2019, Bajet et al. 1994; Hock et al. 1989). Under favorable conditions for disease, tar spot can progress from only a few stromata present in a field to complete coverage of all the plants in under three weeks (Hock et al. 1992).

Currently, growers can manage fungal diseases through fungicide applications and resistant hybrids. While there are fungicides that affect tar spot, they are expensive and do not prevent the disease but only slow the spread once infected. Host resistance for foliar diseases is also a conventional management practice. While current studies are being done to identify resistant hybrids to tar spot (Telenko et al. 2019), they are primarily uncharacterized and seem only to provide partial protection. Therefore, developing highly resistant lines and hybrids will be crucial to prevent future losses to tar spot.

The genetic basis of disease resistance in plants is typically quantitative, with multiple genetic loci, each potentially contributing only a small effect. For example, for a different

ascomycete in maize, Northern leaf blight (*Setosphaeria turcica*), the Ht genes have been providing resistance to specific races of the fungus since their discovery in the 60s and 70s (Hooker 1963 & 1977) and providing partial polygenic resistance to all races (Hooker 1973). Genetic resistance has also been identified for foliar pathogens such as northern corn leaf blight (Poland et al. 2011; Van Inghelandt 2012; Ding et al. 2015), southern corn leaf blight (Kump et al. 2011), and gray leaf spot (Shi et al. 2014; Kuki et al. 2018).

The International Maize and Wheat Improvement Center (CIMMYT) bred tropical maize lines resistant to tar spot in the early 1990s (Bajet et al. 1994; Ceballos and Deutsch 1992). Initially, the genetic architecture was not known, rendering the use of these lines challenging for breeding varieties in temperate regions. In 2016, Maheku et al. used a tropical line-based genome-wide association study (GWAS) and a tropical quantitative trait loci (QTL) mapping population to identify a major tar spot resistance QTL, qRtsc8-1. In 2017, Cao et al. also mapped loci in tropical material using more single nucleotide polymorphism (SNP) markers. They confirmed the major QTL from Maheku et al., identified a few other minor QTLs present, and performed genomic prediction using ridge regression best linear unbiased prediction (rrBLUP).

Thus far, tar spot research has been conducted in tropical materials, and the resistance status of temperate germplasm is primarily unknown. Identifying temperate resistant donors and the genetic loci linked to resistance will support efforts to incorporate tar spot resistant traits into temperate breeding pipelines. In addition, genomic prediction (Meuwissen et al. 2001, Heslot et al. 2015) can be used to predict tar spot resistance in unobserved related individuals, streamlining the process of generating elite resistant varieties. This study assesses and genetically maps tar spot resistance in temperate maize germplasm and identifies candidate genes associated with

resistance. Genetic mapping is then used to select features in genomic prediction models to demonstrate the predictive ability of tar spot susceptibility from genomic data.

MATERIAL AND METHODS

PLANT MATERIAL

A subset of 600 inbred lines from the Wisconsin Diversity panel-942 (WiDiv-942, Mazaheri 2019) was selected and evaluated over two field seasons in Michigan, USA. WiDiv-942 is an expansion of the 503-line Wisconsin Diversity panel (WiDiv-503; Hirsch et al. 2014). These panels are diverse groups of inbred lines comprised of industry expired plant variety protection material, public breeding programs, and the Germplasm Enhancement of Maize (GEM) project, with constrained phenology to the northern U.S. corn belt. The subset of 600 lines was selected based on grain type (field corn prioritized over sweet corn and popcorn) and potential to attain maturity under Michigan conditions.

Two hundred lines originating from the Germplasm Enhancement of Maize project (GEM; Gardner 2018) were also screened. These included 100 lines derived from backcrosses of tropical germplasm with elite temperate material. The lines are typically selected out of a three-way cross with one tropical donor and two elite parents and therefore are 25% exotic and 75% temperate (United States Department of Agriculture, 2020). The remaining 100 lines are BGEM lines, which are double haploids generated from GEM materials.

EXPERIMENTAL DESIGN AND PHENOTYPIC EVALUATION

In 2019, 362 WiDiv inbreds, 100 GEM lines, and 100 BGEM double haploids lines (Appendix: Table A.1) were planted in a farmers' field with a history of tar spot near Allegan, MI. The trial was planted on 3 Jun. 2019 in two-row plots (6.7 m long, 76.2 cm wide, 15.25 cm plant spacing) in a randomized complete block design with two replications.

Disease ratings were used to assess the average percentage stromal coverage on the ear leaf starting on 26 Aug. 2019 after the first detection of the pathogen. They were then recorded on 30 Aug., 6 Sept., 13 Sept., 20 Sept., and 28 Sept. Raters averaged five ear leaves within the plot to assess the average percentage of stromal coverage per plot using the scale provided (Figure 1.1) by the Crop Protection Network (2020). The percentage was assigned categorically and recorded (percentages of 1, 2.5, 5, 7.5, 10, etc. Figure 1.1). In addition to disease ratings, plant/ear heights, anthesis, and silking were recorded. Anthesis and silking time were recorded with the tar spot ratings, and plant and ear height were recorded at the end of the season by measuring the height of the flag leaf and the ear leaf on a representative plant in each plot.

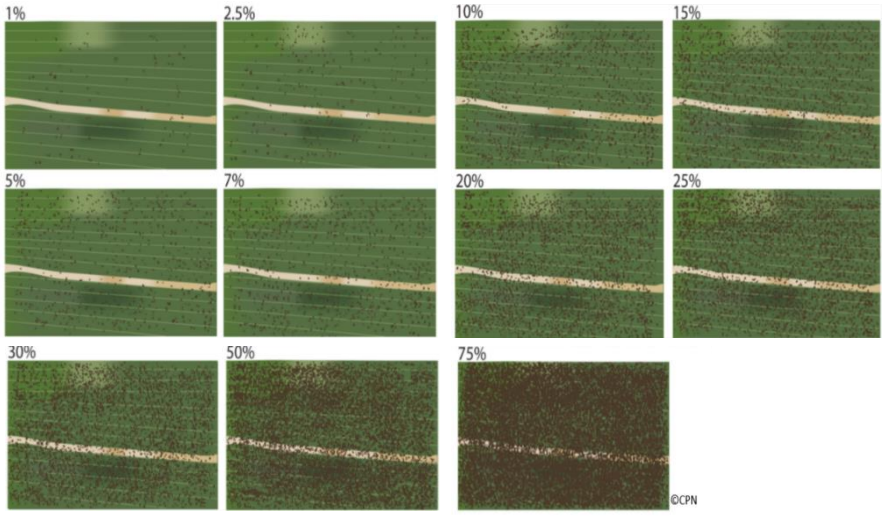

Figure 1: Disease Rating Scale
Computer generated scale used to assess the percent average stromal coverage on the ear leaves. by the crop protection network (Crop Protection Network, 2020) on a per-plot basis.

In 2020, 600 WiDiv, 100 GEMs, and 100 BGEMs inbreds were planted in a farmer's field near Decatur, MI, on 4 May 2020. Three hundred and seven WiDiv lines from 2019 were expanded to 600 inbred lines in 2020. In 2019, the varieties were planted in a randomized complete block design with two replications with the same plot size and plant spacing. The disease was rated starting on 24 Jul. 2020 and recorded on 31 Jul., 7 Aug., 14 Aug., 21 Aug., and

28 Aug. Using the same protocol explained above to assess the average percentage of stromal coverage on the ear leaf. However, numerical percentage values (interpolating in the scale) were used to rate values precisely instead of categorical percentages.

In addition to the Michigan location, collaborators planted trials near West Lafayette, IN (685 inbreds; Appendix: Table A.1) and Madison, WI (691 inbreds; Appendix: Table A.1). Materials grown in all three locations contained a common set of 529 inbred lines. These fields were planted on 15 Jun. in Indiana in 2 row 6-meter plots and on 27 May in Wisconsin in 2 row 3.8-meter plots. However, only one replication was planted at the Indiana location due to a planter issue and space limitations. In Indiana, collaborators rated disease by selecting three ear leaves within each plot, determining percentage stroma coverage, then averaging the plot's three ratings. Ratings were completed on 4 Sept., 17 Sept., and 30 Sept. Due to low disease severity at the Wisconsin location, collaborators only recorded one rating on 16 Sept. using the same method as Indiana with five leaves instead of three.

PHENOTYPIC DATA ANALYSIS

Analysis of variance (ANOVA) was conducted using the linear model function in R to check the significance of genotype and rater. Genotype was significant at $p< 0.01$ in 2020 for percent stroma coverage at each weekly ratings at weeks 2-6 after the initial infection. In 2019, the genotype was significant at all dates. Ratings for each genotype in both field seasons were averaged between the two replications. In 2019, the raw values were averaged; however, tar spot severity was higher in 2020 than in 2019. With the increase in disease pressure, the rater became statistically significant in the ANOVA. To fix the bias, 20 plots were rated by all raters. This data was transformed using a box-cox transformation, and then the fixed effect fi the rater was

subtracted from the values. The data was then untransformed, and these average severity ratings were then used for further analysis.

All statistics were performed in R software (R Core Team 2013). Violin/density plots showing disease distribution across the subpopulations were generated using the ggplot2 package (Wickham 2016). The linear model function in base R software was used for the analysis of variance (ANOVA) and residual analysis using the following model:

$$Y_{\text{ir}} = \upsilon + G_{\text{i}} + r + e$$

Where Y is the phenotypic value of the *i*th genotype (*G*) in the *r*th replicate.

Repeatability (i^2) was calculated for single environments using the formula presented by Webb et al. (2006):

Single environment: $i^2 = \frac{Q_\beta}{Q_\beta + \frac{\sigma_e^2}{r}}$

Where Q_β is a quadratic function of fixed effects, σ_e^2 is error variance, and r is the number of replications in each environment.

Area Under Disease Progress Curve (AUDPC) was used to quantify disease pressure over time in locations with greater than three ratings (Shaner and Finney 1977) using the trapezoidal method (Campell & Madden 1990) and the formula:

$$AUDPC = \sum_{i=1}^{n-1} \left(\frac{y_i + y_{i+1}}{2x(t_{i+1} + t_i)}\right)$$

Where y_i is the percent tar spot severity at the *i*th observation, t_i is the time in days after infection of the *i*th observation, and *n* is the total number of observations. AUDPC was calculated using all three ratings (Indiana 2020; IN_AUDPC), all six ratings (Michigan 2019-2020; AUDPC6), and the first five ratings (Michigan 2019-2020; AUDPC5). This method was

done to compare lines with different maturities, as the sixth ratings in Michigan were recorded very late in the season when some genotypes had already dried down.

GENOTYPIC ANALYSIS AND GWAS

Previously published filtered and imputed SNPs called from WiDiv seedling total RNA-seq data from Mazaheri et al. (2019) were further filtered to remove markers with a minor allele frequency less than 3% and missing data rates greater than 20% for subsets of the population. The number of inbred lines and marker subsets varied by location: Michigan 2019 (Allegan) – 363 inbred lines, 496,845 SNPs; Michigan 2020 (Decatur) – 596 inbred lines, 473,868 SNPs; Indiana – 674 inbred lines, and 476,869 SNPs; Wisconsin – 691 inbred lines and 483,603 SNPs.

The Genome Association and Prediction Integrated Tool (GAPIT) package in R (Lipka et al. 2012) was used to calculate a kinship matrix per the methods of VanRaden (2008). GWAS was then performed using the fixed and random model Circulating Probability Unification (FarmCPU) method in R (Liu et al. 2016) with a significance threshold of FDR 0.05.

GWAS was conducted on all adjusted severity ratings and AUDPC. Some inbreds at the latter rating had desiccated and were not included in the GWAS for those dates.

IDENTIFICATION OF CANDIDATE GENES

Candidate genes were filtered by searching 8000 bp (4000bp on each side) out from the significant SNP reported. Maize GDB (www.maizegdb.org, Andorf et al., 2010) was used to annotate candidate genes or gene models containing the significant SNPs. The interest level was assessed using expression data from Swart et al. (2017) for up and down-regulation of the gene when infected with the fungi *Cercospora zeina* or *Colletotrichum graminicola.*

GENOMIC PREDICTION

rrBLUP (Endelman 2011) and three Bayesian regression models (BGLR: Perez 2014) - Bayesian Ridge Regression, Bayes A, Bayes B - were used in genomic prediction to estimate the Genomic Estimated Breeding Values (GEBV) of all the traits. The Bayesian models had different assumptions regarding how the SNPs affect each other, as described in de Los Campos et al. (2013), and rrBLUP as described in Whittaker (2000).

The top *n* most significant SNPs were taken from the GWAS to predict lines within Michigan. This method was chosen rather than a random subset of SNPs as it was more accurate using fewer SNPs (20,000 random SNPs: 45% accurate; data not shown). Using a 10-fold cross-validation, the 596 inbred lines were divided with subsetted SNP data into ten subsets, where nine sets trained the model while one was used to testing it. The randomization of subsets occurred ten times, and the model measured the accuracy for each run. In addition, the model recorded the Pearson correlation coefficient (r) between the predicted values and the adjusted ratings as the accuracy.

Using the entire Michigan phenotypic dataset to train the model, all four models were evaluated to test their ability to predict the tar spot severity in Indiana. The prediction accuracies for these models were tested using genotypes planted at both locations and the 105 lines that were only planted and rated in Indiana. The Pearson and Spearman correlations between the predicted and the observed values were recorded as the prediction accuracy.

RESULTS

The descriptive statistics for the maize inbred lines' responses over the two seasons are shown in Table 1. Disease expression varied between years and populations (Table 1). However, there was ample differentiation of resistant germplasm each year, and repeatability (i^2) was 0.67 and 0.53 for Michigan (2019 and 2020, respectively) and 0.35 for Wisconsin in 2020.

WiDiv: Final Rating						
	Min	Max	Median	Mean	Std Dev	Repeatability
Michigan 2019	0	25	1	2.08	3.25	67
Michigan 2020	0	38	3	3.95	3.9	52.8
Indiana 2020	0	15.67	1.67	2.05	1.82	Only 1 Rep
Wisconsin 2020	0	0.6	0.02	0.03	0.48	34.9

Table 1.1: Summary Statistics of WiDiv Final Rating Per Environment
Expressed as a percentage stroma coverage. Highest severity occurred in Michigan 2020.

MICHIGAN 2019

In 2019, the first signs of tar spot were recorded on 22 Aug. While most plots showed tar spot symptoms (Figure 1.2), some lines did not exhibit any tar spot lesions. The GEM and BGEM were similar in the distribution of the ratings and AUDPC. However, the WiDiv showed greater variation (standard deviation 3.25 vs. 1.2), containing varieties with no tar spot and one with 25% of the ear leaves covered.

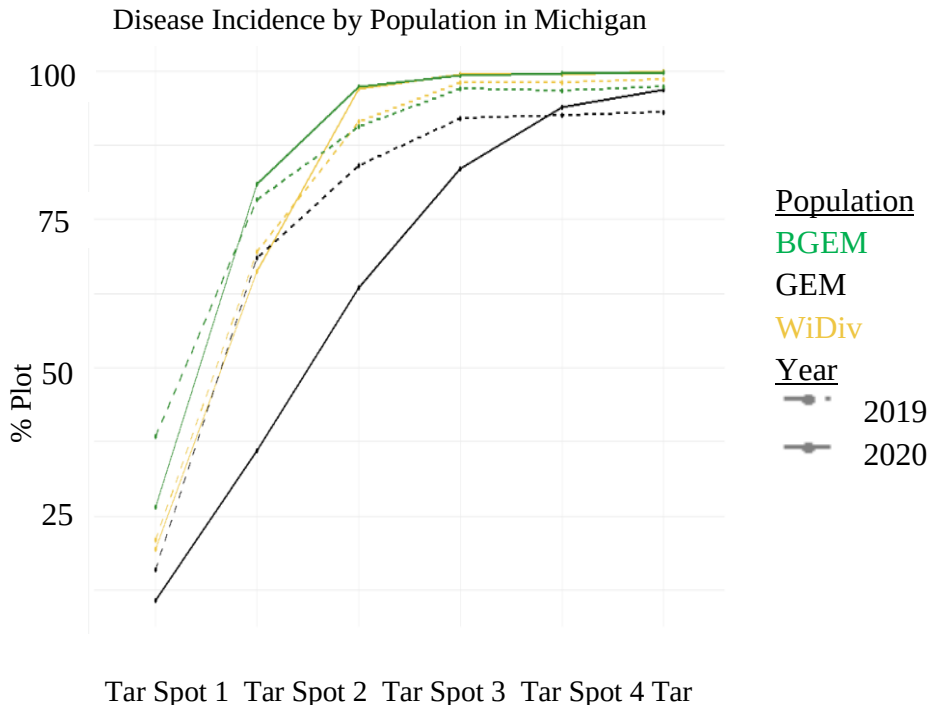

Figure 1.2: Disease Incidence by Population in MI 2019-2020
Shows the percentage of plots that were infected with tar spot at each rating. Green (BGEM), black (GEM), and yellow (Wisconsin Diversity) lines represent population, while dashed (2019) and solid (2020) lines represent year. Final ratings were very similar overall, but GEM population and 2019 both showed slower onset of disease.

Each plot's final plant and ear height, anthesis date, and silking date were also recorded. These traits demonstrated no significant correlation with tar spot rating, as shown with a correlation heat map (Figure 1.3).

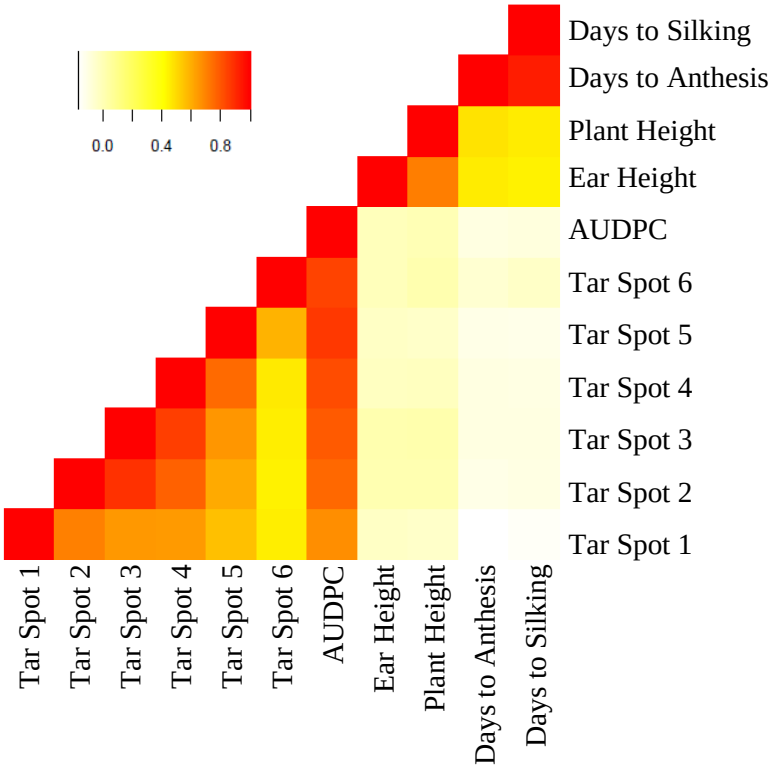

Figure 1.3: Correlation Heat Map Showing Relationship Among the Traits Collected
Tar Spot 1 (TS1) refers to the 1st rating take and goes up to the final rating Tar Spot 6 (TS6). Plant heights and flowering times showed very little correlation with the disease rating traits.

MICHIGAN 2020

In 2020, tar spot was first observed on July 17th. The disease incidence level for 2020 was faster in taking over all the populations than in 2019, and most plots had tar spot symptoms (Figure 1.2). In general, plots in 2020 had higher severity throughout the field compared to 2019. Also, as in 2019, the average plot AUDPC was 14.2 while 30.4 for 2020. As in 2019, the WiDiv had the highest severity overall (38%); however, the BGEM had one line approaching that level (32%). The medians of the BGEM and WiDiv were similar at 3 and 2.1, respectively, while the GEM median was 0.3 (Figure 1.4 A-B).

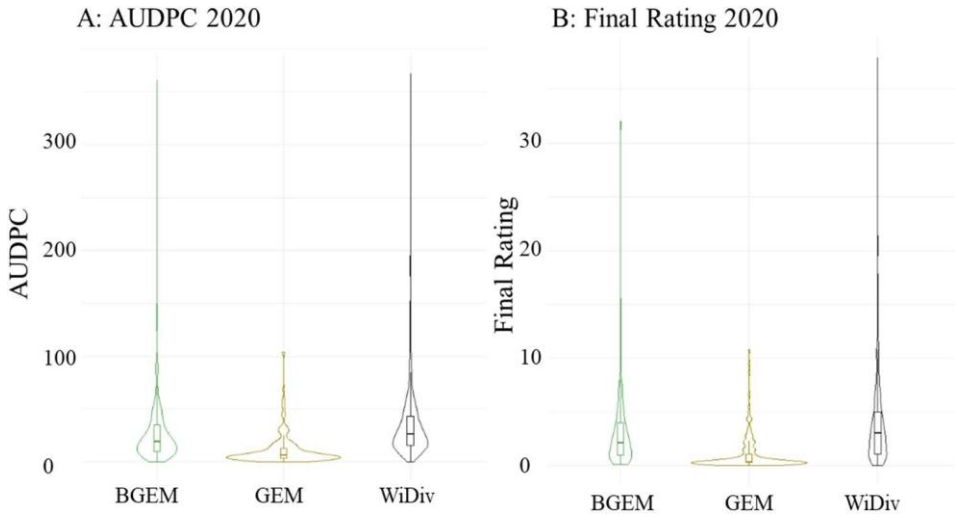

Figure 1.4: A-B: Distribution of AUDPC And Final Rating by Population In 2020
A) Distribution of Area Under Disease Progress Curve (AUDPC), a measure of disease pressure over time) and B) final rating by population in 2020. BGEM and WiDiv populations showed more variation, while GEM lines had the greatest number of resistant lines.

The plant/ear height and flowering time for each genotype were not recorded at the Decatur, MI location; however, these traits were recorded in the field nursery in East Lansing, MI (not included). Like 2019, the traits did not show any significant correlation to tar spot disease severity.

GWAS

A genome-wide association study on the adjusted phenotypic tar spot ratings and the calculated AUDPC for each inbred was used to determine the genetic architecture for tar spot resistance. The first two principal components and a kinship matrix were fitted using GAPIT. The Quantile-quantile plots (Appendix: Figure A.1) showed appropriate control for the population structure and kinship. The GWAS for the Michigan AUDPC, the Michigan final tar spot rating, and the Indiana AUDPC are provided in Figures 1.5A-B & 1.6, respectively. In addition, the number of significant SNPs per adjusted trait are provided in Tables 1.2, 1.3, and 1.4 (total 79: removing overlapped) (Full list: Appendix: Table A.2). There were 110 genes

within 8000 base pairs (4000 on each side) of the significant SNPs identified in the GWAS analysis (Appendix: Table A.3). Candidate genes that respond to pathogen infection in an expression atlas are expressed in Appendix: Table A.4.

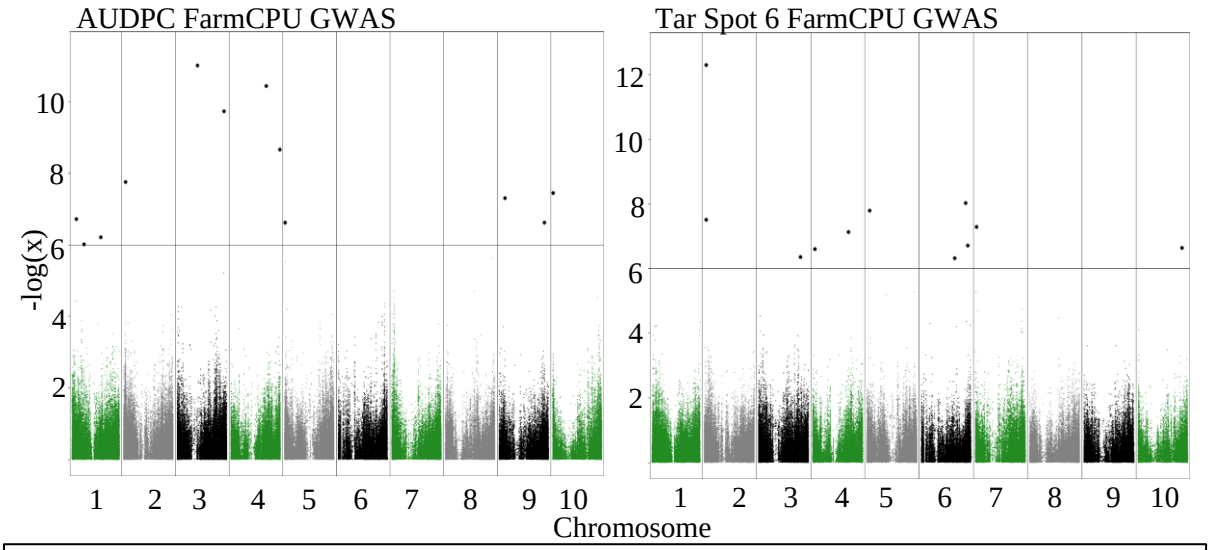

Figure 1.5A-B: Manhattan plots of GWAS in MI 2020 for AUDPC 6 and Final Rating. Michigan 2020 for AUDPC 6 (A) and Final Rating (B; Tar Spot 6). Some significant SNPs were shared between these two traits (Table 1.2), but many were unique, highlighting the unique information obtained from ratings at a single timepoint vs disease progress over time.

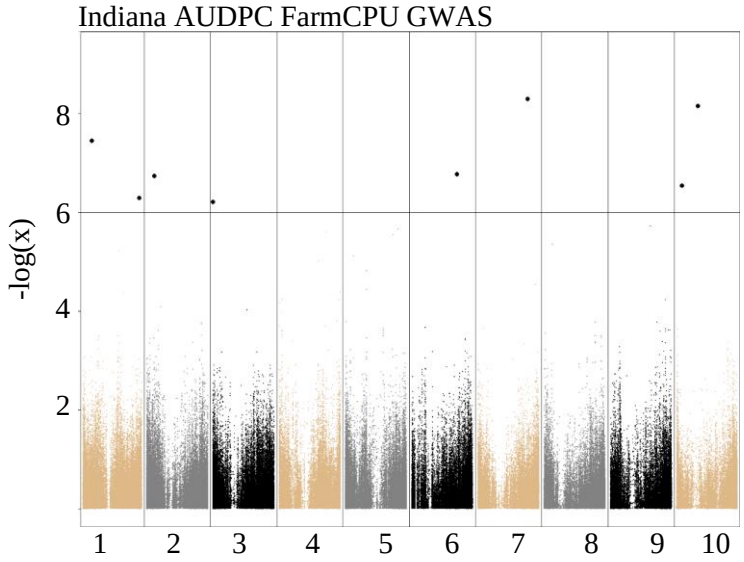

Figure 1.6: Manhattan plot for GWAS result in IN 2020 for AUDPC
There were no SNPs that were shared between the Michigan location and the Indiana locations. This would infer a strong GxE interaction in tar spot resistance.

Significant SNPs - Michigan			
Trait	**# Of Inbreds**	**# Of SNPs**	**Location**
AUDPC6	569	7	2 SNP: Chrom 3 & 4 1 SNP: Chrom 2, 9, 10
Tar Spot 6	571	11	3 SNP: Chrom 6 2 SNP: Chrom 2 & 4 1 SNP: Chrom 3, 4, 5, 7, and 10
Tar Spot 5	588	6	3 SNP: Chrom 1 1 SNP: Chrom 6, 8, 10
Tar Spot 4	593	7	2 SNP: Chrom 1 & 4 1 SNP: Chrom 2, 5, 10
Tar Spot 3	595	5	2 SNP: Chrom 3 1 SNP: Chrom 2, 5, 9
Tar Spot 2	596	6	2 SNP: Chrom 3 & 5 1 SNP: Chrom 7 & Unmapped
Tar Spot 1	596	0	None
AUPDC5	588	8	3 SNP: Chrom 1 2 SNP: Chrom 4 1 SNP: Chrom 3, 5, 6, and 7

Table 1.2: Significant SNPs Per Chromosome Per Trait from MI 2020 GWAS
The distribution of significant SNPs per chromosome per trait from the GWAS of Michigan 2020 data. Tar Spot refers to the 1st rating taken and goes to the final rating, Tar Spot 6. AUPDC5/6 are the AUDPC calculations for ratings 1-5 (AUDPC5) and 1-6 (AUDPC6), respectively.

Significant SNPs - Indiana			
Trait	**# Of Inbreds**	**Num of SNPs**	**Location**
AUDPC	673	8	2 SNP: Chrom 1 & 10 1 SNP: Chrom 2, 3, 6, and 7
Tar Spot 3	673	7	2 SNP: Chrom 1 1 SNP: 3, 5, 6, 7, and unmapped
Tar Spot 2	673	9	3 SNP: Chrom 7 2 SNP: Chrom 2 1 SNP: Chrom 4, 6, and 9
Tar Spot 1	673	0	None

Table 1.3: Significant SNPs Per Chromosome Per Trait from IN 2020 GWAS
The distribution of significant SNPs per chromosome per trait from the GWAS of Indiana 2020.

Significant SNPs Wisconsin			
Trait	**# of Inbreds**	**Num of SNPs**	**Location**
Final Rating	691	14	3 SNP: Chrom 10 2 SNP: Chrom 1, 3, 5, 8, and 9 1 SNP: Chrom 7

Table 1.4: Significant SNPs Per Chromosome from WI 2020 GWAS
The distribution of significant SNPs per chromosome from the GWAS of Wisconsin 2020 data.

GENOMIC PREDICTION

Genomic prediction was conducted using four different methods. Overall, Bayes B was the most effective at predicting all traits averaged across all SNP levels. The next most accurate was Bayes A, followed by a mix of BRR and rrBLUP depending on the trait of interest. End-of-season AUDPC (AUDPC6) was the most predictive trait, at 79.1% accuracy across all SNP levels using Bayes B, followed by the final tar spot rating (Figure 1.7 & Appendix: Table A.5).

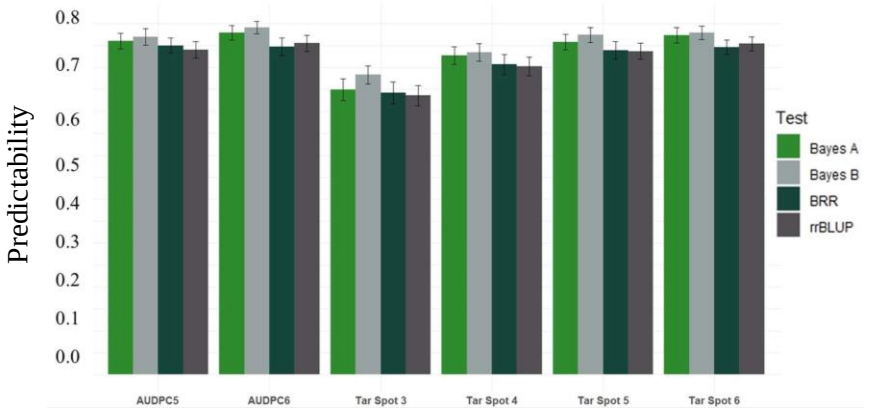

Figure 1.7: Genomic Prediction Accuracy of Traits Using Different Algorithms
Using a 10-fold cross validation of the trait data taken in Michigan, the prediction accuracy of Bayes A, Bayes B, BRR and rrBLUP models are shown. AUDPC6and Bayes B were the most accurate.

Overall, using the 200-400 most significant SNPs led to the highest prediction accuracy without adding a significant number of SNPs to the model, with 300 being the most consistent. As before, AUDPC6 was the most accurately predicted trait at 81.8% using 400 SNPs (Figure 1.8), followed by the final tar spot rating (79.9% at 300 SNPs) (Table A.6 A-D). Surprisingly, using only 20 SNPs was 75% accurate using a Bayes A method.

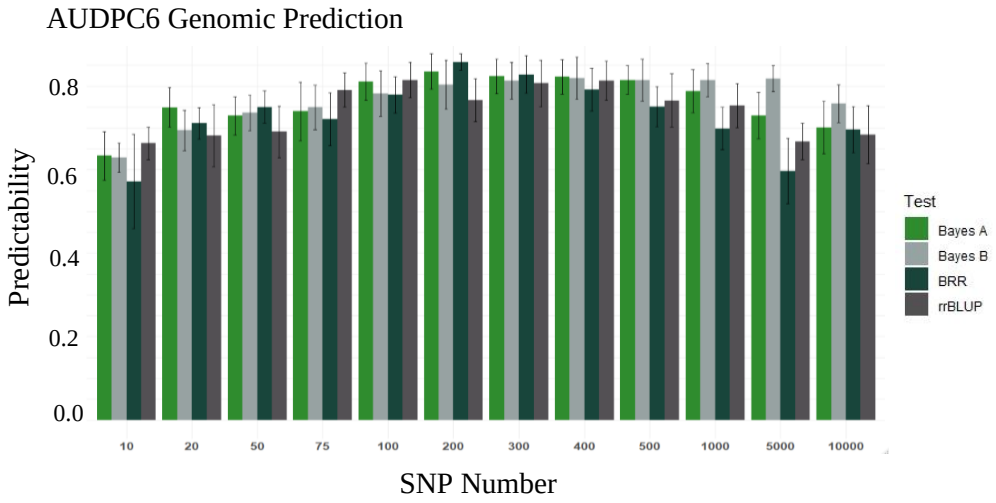

Figure 1.8: Genomic Prediction of Final AUDPC in MI 2020 Using Different Algorithms and SNPs levels
Using a 10-fold cross validation of the Michigan 2020 final AUPDC data, the prediction accuracy of Bayes A, Bayes B, BRR and rrBLUP models are shown at respective SNP numbers. 200-400 SNPs was the most accurate at 81.2-81.8%.

A Bayesian ridge regression model (BBR) using the Michigan 2020 data was used to test the prediction of 576-596 lines planted (dependent on the trait) in the Indiana location (observed genotypes in an unobserved environment), as well as 105 lines planted only in Indiana (unobserved genotypes in an unobserved environment), at multiple SNP levels. Spearman rank correlation was used to indicate this approach's usefulness in selecting the best and worst lines in a breeding program. Once again, AUDPC remained the most accurately predicted trait, at 54.2% and 37.9% (Figure 1.9) correlation for observed and unobserved genotypes, respectively. Indiana's final tar spot rating followed AUDPC Indiana, being 47.9% accurate using Michigan's 5th rating on observed genotypes and 28.6% accurate when using Michigan's last (6th) rating. Accuracy varied slightly with the SNP number but peaked at around 5000 (Table A.7 & A.8).

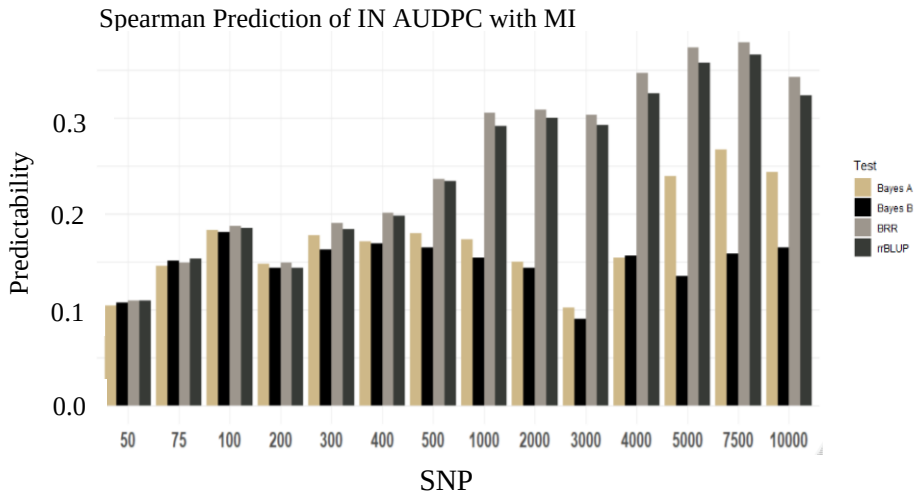

Figure 1.9: Genomic Prediction of Unobserved IN 2020 AUDPC from MI 2020 AUDPC Using AUDPC from Michigan 2020 to train prediction models, the prediction accuracy of Bayes A, Bayes B, BRR and rrBLUP models are shown. BRR and rrBLUP were more accurate at higher SNP levels maxing at 37.6% using 7,500 SNPs.

DISCUSSION

VARIATION IN TAR SPOT RESISTANCE

In this study, resistance to tar spot showed significant variation across both inbred lines and environments. Resistance was moderately repeatable for the Michigan locations, with repeatability at 67.0 and 52.8 in 2019 and 2020, respectively. The disease severity at the Wisconsin location was very low, with repeatability of 34.9%. As expected, the severity of tar spot is environmentally influenced. The data reflected this trend in the genetic mapping results, with no significant SNPs, shared between the Indiana and Michigan results. No evidence was identified for a correlation between tar spot resistance and plant height or maturity. This finding contrasts with Mahuku et al. (2016), who observed a negative correlation between tar spot resistance and maturity in tropical germplasm. A negative association between resistance and maturity may be due to population admixture - more resistant tropical-derived material combined with more susceptible, less tropically derived material - rather than a direct cause-effect

relationship. It is also possible that later-maturing lines could accumulate additional lesions later into the season than early maturing temperate lines in the upper Midwest United States, as hybrids with later maturity have had greater yield losses (Telenko 2019). Having more time for the fungi to reproduce may effectively counteract any negative correlation between the traits.

CANDIDATE GENES

There were 110 genes near the significant SNPs identified in the GWAS analysis (Table A.3). Of these 110 genes, 28 showed a change in expression upon pathogen infection (Table A.4). One interesting gene, Zm00001d041082 (kaurine synthase4/ks4), encodes a key enzyme of diterpene phytoalexin biosynthesis. Phytoalexins are synthesized and accumulate in plants after exposure to microorganisms such as bacteria and fungi. Thus, they are suggested to serve as antimicrobial compounds in plant-induced defense systems in rice (Ono et al. 2001), maize (Block et al. 2019), and other plants (Hammerschmidt 1999).

Another candidate gene, Zm00001d037550 (peroxidase5/px5), is involved in the degradation of baicalein. Baicalein is a flavone that rapidly detoxifies hydrogen peroxide, accumulating in response to pathogen-induced mechanical damage (Mehdy 1994). According to Peng et al. (1992), reactive oxygen species (ROS) inhibit fungal pathogen spore germination, lowering pathogen viability (Keppler and Baker 1986), but also playing a role in abiotic stress tolerance (Gill and Tuteja 2010).

GENOMIC PREDICTION

In 2017, Cao et al. conducted an association study for tar spot resistance using tropical material and performed genomic prediction using rrBLUP. Our candidate gene regions did not overlap with any of their published loci. This result may indicate that temperate and tropical germplasm utilize different sources or pathways to confer resistance to this fungal pathogen. The

rrBLUP model developed in Cao et al. using their 286-line tropical diversity panel resulted in 55% accuracy using 10,000 markers. In this study, rrBLUP was compared to three Bayesian approaches. Bayes A and B were slightly more accurate than rrBLUP (75 vs. 79% using Bayes B for AUDPC6 overall marker sets), and tar spot susceptibility could be predicted at up to 81.2% using 300 markers in Bayes B. This result may convey that the genetic architecture of tar spot resistance in at least the temperate germplasm may involve a finite number of slightly larger-effect genes rather than the infinitesimal model of a large number of genes with a small effect assumed in rrBLUP. This is further supported by the high predictive ability of very small numbers of SNPs (between 65-75% for 10 or 20 SNPs), which may make marker-assisted selection approaches a viable option in breeding for tar spot resistance.

Using the BRR model trained only on disease severity in Michigan, tar spot susceptibility of lines planted in Indiana was predicted with a Spearman rank correlation up to 54% for observed genotypes and up to 37% for unobserved genotypes. Predicting a new environment will cause a significant drop in accuracy, as disease severity is heavily environmentally influenced. In 2020, overall severity in Indiana was lower on average than in Michigan (mean of 3.95 in Michigan vs. 2.05 in Indiana on final rating). Despite this drop, the accuracy is likely high enough for genomic prediction to be successful – that is, the prediction models may enable breeders to estimate the most resistant and susceptible genotypes in breeding or backcross populations without having to test them each cycle under disease pressure. Fine-mapping populations are being developed to validate and narrow down candidate gene regions and enable marker-assisted backcross selection or even gene-editing approaches to confer tar spot resistance to elite lines in the future.

CHAPTER 3: OPTIMIZING USE OF RESOURCES IN CORN PERFORMANCE TRIALS BY ANALYZING GXE INTERACTIONS AND THE NUMBER OF REPLICATION

ABSTRACT:

Crop variety trials, such as the Michigan Corn Performance Trials (MCPT), provide information to producers on which of the tested hybrids perform best in their given environment. Though these trials produce valuable data, they are resource-intensive, requiring many locations and replications to achieve accurate data. To maintain high resource allocation and high-efficiency testing, maintaining non-redundant, discriminative environments along with the optimal number of replications is critical. This study examined nine years of multi-environment yield trial data collected from the MCPT program to determine if any of the nine locations within the three maturity zones produced similar GxE effects. We also investigated the optimal number of replications needed to reach a target level of repeatability (i^2) in each maturity zone. Of the three locations planted in the late-maturing Zone 1, the Branch location was not correlated with the other two locations, Cass, and Washtenaw, which performed similarly to those in the mid-maturing Zone 2. In early maturing Zone 3, we established that the three environments (Montcalm, Mason, and Huron) were discriminating from each other; however, two of those locations (Mason and Huron) seem to act more comparably to the locations in mid maturing Zone 2. Finally, using a sliding window of year combinations, we determined that, while year-dependent, two replications are sufficient in Zone 1 and 2 to get 75% of the maximum repeatability across the two zones, while four replications are needed for Zone 3.

INTRODUCTION

Crop variety trials such as the Michigan Corn Performance Trials (MCPT) for corn (*Zea mays L.*) provide unbiased, third-party information on commercial hybrid performance. Michigan growers use the data collected from the MCPT to decide which commercial hybrids perform best for their cropping environment. The MCPT grows these trials in two to three locations in each of the five Michigan maturity environment zones defined by traditional metrics such as maturity measured in growing degree days (GDDs) and climate factors. Hybrids are planted in these zones at many target locations with several replications, sometimes over multiple years, to obtain accurate data (Figure 2.1).

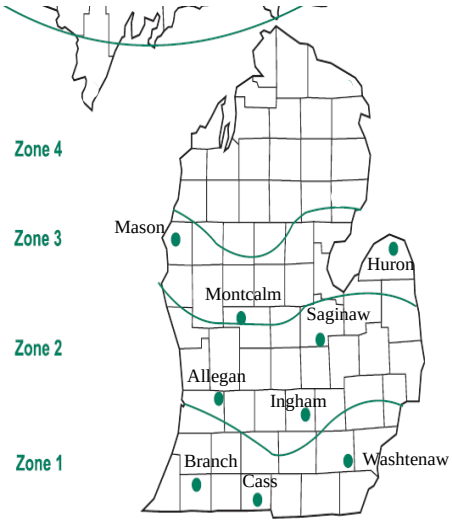

Figure 2.1: Locations of MCPT Trails
Locations used in the MCPT within the five major maturity zones in Michigan in the MCPT. In most zones, the MCPT has three locations per zone. Locations changed per year, but the locations used in this study are those that were most consistently used. The name of the locations coincides with the county's name it is located within. Figure from 2018 Michigan Cron Hybrids Compared (Singh, 2018).

While the MCPT produces valuable data, its integrity depends on the correct establishment of zones across Michigan. Suboptimal zone establishment decreases time- and resource-use efficiency. In addition to maintaining the integrity of MCPT zones, it is crucial to

identify discriminating environments within the zones to match the different cropping environments within the maturity zones. A method to assess MCPT zonal locations will help efficiently identify superior hybrids for regional applications.

GGE biplots can be used to analyze zonal location correlations and identify suboptimal zone groupings. GGE biplots are graphical representations of the genetic effect and genetic by environmental effect (Yan et al., 2000; Yan et al., 2006; Yan et al., 2009; Yan, 2014). A phenotype, such as yield, can be split into three variance components: genotypic (G), environmental (E), and genotype by environment interaction (GxE). Linear modeling can be used to partition these variance components, allowing for the removal of the non-repeatable environmental effect, leaving only the genotypic and GxE interaction effects. The first two principal components derived from the singular-value decomposition of environment-centered mean grain yields graphically display a GxE interaction in a two-way table (Yan & Kang, 2003). GGE biplots have been used in wheat (Thomason & Phillips, 2006), cotton (Blanche, 2006), soybean (Dalló et al., 2019), and in both breeding and hybrid selections in maize (Oyekunle et al., 2017; de Oliveira 2019).

While the number of locations planted is changeable in theory, mature programs will often have a set of accessible test locations. This reality makes reducing replications at each location an excellent potential target for increasing test efficiency and optimal resource allocation.

To find the optimal number of replications, we used a method published by Yan et al. (2015) & Yan (2021). The method reworks the broad sense heritability equation to find the optimal number of replications needed to reach a target broad sense heritability (repeatability) level. Replications help separate noise from the signal as they measure variation, provide an

average of the experimental unit, and control for outliers within the experiment. The more replications in an experiment, the more precise the measurements become; however, replications increase costs to time and resources. Therefore, finding an optimal number of replications to ensure high confidence but conserve resources is crucial to production. For example, for wheat production in Canada, Yan et al. (2015) concluded that instead of planting four replications, in most locations, only three replications were needed to reach a repeatability measure of 75% of the max repeatability.

With the need for growers to have accurate, unbiased yield data, this study takes nine years of MCPT data across three maturity zones and nine environments with the objectives of i) testing GDD zones for similarities to see if they need to be adjusted, ii) testing locations within zones to find differentiating environments for hybrid testing, and iii) finding the optimal number of replications needed for maize yield trials in Michigan.

MATERIAL AND METHODS

MICHIGAN CORN PERFORMANCE TRIALS (MCPT)

MCPT yield data collected between 2011-2019 at the three zones with the most consistently used locations (Zone 1, 2, and 3) were used in this study. Commercial seed companies determined hybrids they wanted to be planted in each maturity zone. This design resulted in a highly unbalanced dataset with few hybrid replications across maturity zones and/or years (Tables 1:A-C).

A

	Zone 1			Zone 2			Zone 3		
Year	**Branch**	**Cass**	**Washtenaw**	**Allegan**	**Ingham**	**Saginaw**	**Huron**	**Montcalm**	**Mason**
2011	115	115	115	122	122	122	116	116	116
2012	103	103	NA	141	141	141	119	119	119
2013	122	121	73	139	139	139	109	NA	109
2014	114	114	114	124	124	124	93	NA	93
2015	89	89	89	108	108	108	75	75	75
2016	103	103	103	130	130	130	84	84	84
2017	94	94	94	126	126	72	77	77	77
2018	88	88	NA	120	67	120	77	77	77
2019	77	77	NA	91	91	NA	66	66	66

B

Year	**Zone 1 & 2**	**Zone 2 & 3**
2011	34	78
2012	38	73
2013	52	72
2014	40	55
2015	32	45
2016	40	53
2017	40	56
2018	31	58
2019	24	39

C

Year	**Zone 1**	**Zone 2**	**Zone 3**	**Zone 1 & 2**	**Zone 2 & 3**
2011-2019	587	902	613	331	529

Tables 2.1 A-C: Number of Hybrids per Subset
The number of hybrids planted in each location year combination (A), multiple zones (B), and overall years. We can see that depending on the year, the hybrid number changed significantly.

Each entry was planted in four replications across the field in four-row plots, and the center two rows were machine-harvested for yield. Following harvest, the yield was adjusted to 15.5% moisture. Additional details such as planting date, spraying, and harvest date are in the reports at https://varietytrials.msu.edu.

STATISTICAL MODELS

GGE BIPLOT

A GGE biplot analysis was conducted on the average yield across the four replications for each genotype within each environment. The GGEBiplots package in R (Dumble et al., 2017) was used to conduct the analysis. We environmentally centered and scaled, but did not transform, the data. We used the biplot model proposed by Yan and Kang (2003):

$$Y_{ge} - \overline{Y_e} = \lambda_1 \xi_{g1} \eta_{e1} + \lambda_2 \xi_{g2} \eta_{e2} + \varepsilon_{ge}$$

Where Y_{ge} is the mean yield of the gth genotype in the eth environment; $\overline{Y_e}$ is the mean yield across all genotypes in the eth environment; λ_1 and λ_2 are the singular values for PC1 and PC2; ξ_{g1} and ξ_{g2} are the PC1 and PC2 eigenvectors for the gth genotype; η_{e1} and η_{e2} are the PC1 and PC2 eigenvectors for the eth environment; and ε_{ge} is the residual of the model associated with the gth genotype in the eth environment.

The angles between environment points indicate the degree to which environments are correlated. For example, an angle greater than 90 degrees indicates that environments are negatively correlated, a 90-degree angle indicates that environments are not correlated, and an angle less than 90 degrees indicates that environments are positively correlated. The angles between points were calculated using the 'angle' function in R's 'matlib' package (Friendly et al., 2020).

REPLICATION ANALYSIS

Yan et al. (2015) explored using the breeder's equation to estimate the optimal number of replications needed to achieve a broad sense heritability threshold. Yan et al. (2015) adapted the H equation calculated by DeLacy et al. (1996) and reworked the equation to get the optimal number of replications at one location:

$$r = \frac{\sigma_e^2}{\sigma_g^2} * (\frac{H}{1-H})$$

Where H is the broad-sense heritability, σ_g^2 is the genotypic variance, σ_e^2 is the error variance, and r is the number of replications.

Yan (2021) tested his concept further to account for a single-year and multi-location trial by using:

$$r = \frac{\sigma_{e,ML}^2}{l * \sigma_{g,ML}^2} * (\frac{H_{ML}}{1-\frac{H_{ML}}{H_{MML}}})$$

Where $\sigma_{g,ML}^2$ is the genotypic variance, $\sigma_{e,ML}^2$ is the experimental error variance based on the single year, multi-location trial, l is the number of locations, H_{ML} is the heritability threshold, and H_{MML} is the maximum achievable across-location heritability:

$$H_{MML} = \frac{\sigma_{g,ML}^2}{\sigma_{g,ML}^2 + \frac{\sigma_{gl}^2}{l}}$$

Where σ_{gl}^2 is the variance for location by genotype interaction.

Yan (2021) also tested a multi-location and multi-year equation:

$$r = \frac{\sigma_{e,MLY}^2}{l * y * \sigma_{g,MLY}^2} * (\frac{H_{MLY}}{1-\frac{H_{MLY}}{H_{MMLY}}})$$

Where $\sigma^2_{g,MLY}$ is the genotypic variance, $\sigma^2_{e,MLY}$ is the experimental error variance based on the multi-year, multi-location trial, l is the number of locations, y is the number of years, H_{MLY} is the heritability threshold, and H_{MMLY} is the maximum achievable across-location heritability:

$$H_{MMLY} = \frac{\sigma^2_{g,MLY}}{\sigma^2_{g,MLY} + \frac{\sigma^2_{gl}}{l} + \frac{\sigma^2_{gy}}{y} + \frac{\sigma^2_{gly}}{ly}}$$

Where σ^2_{gl} is the variance for location by genotype interaction, $\sigma^2_{gy}$is the variance for the genotype by year interaction, and σ^2_{gly} is the variance for the three-way interaction of genotype, location, and year.

OUTLIER DETECTION

Both GGE biplots and optimal replication analysis rely on the genotypic variance associated with the environment. Abnormal, uncontrolled errors such as flooding, animal & irrigation wheel damage can occur on certain replications. To maximize the usefulness of this analysis, we implemented a Dixon Q test to remove any replications over the .05 threshold from the replication grouping (Dean & Dixon, 1951). Observations with a studentized residual > 3.25 or < -3.25 were removed to maintain similar normalization levels in each maturity zone subset and prevent extrapolations. After detection and removal, across all years, there were 1987 hybrids with 29,222 vs. 2,169 hybrids with 34,576 replications in the original data.

RESULTS:

PEARSON CORRELATION PLOTS:

We calculated the pair-wise Pearson correlations of hybrid yields across locations in all years (Figure 2.2). The correlation varied substantially between location combinations. These correlations infer what we would expect in the GGE biplots but do not parse all the variances

separately. Zones 1 and 3 had very few hybrids in common, so we discarded this pairing for all analyses.

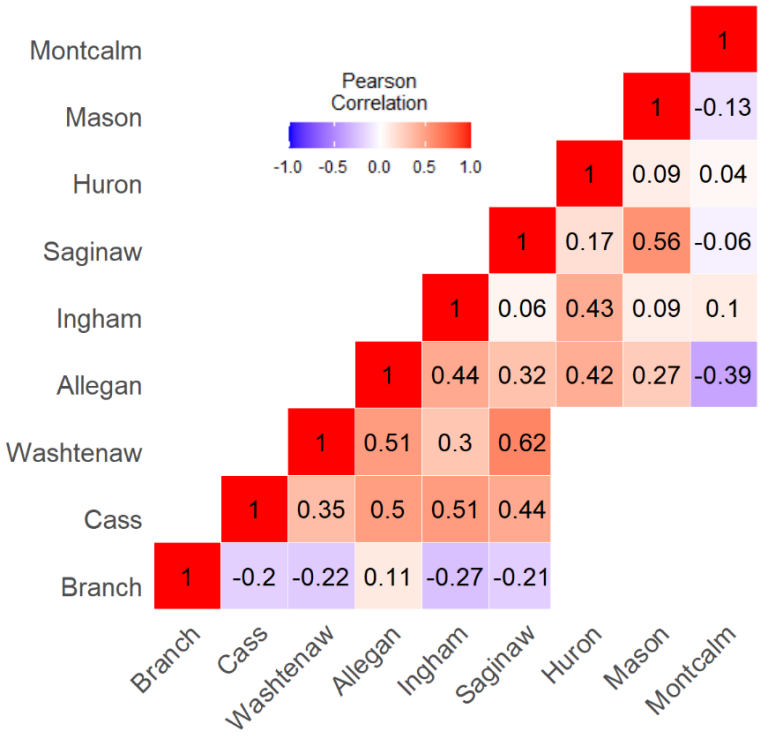

Figure 2.2: Correlation Heatmap of County Combinations
Pearson Correlation plots using the hybrids planted across the locations and years. There are some trends such as Branch and Montcalm having negative or no correlation for most locations while Allegan is nearly all positive with the exception of Montcalm.

GGE BIPLOT ANALYSIS

SINGLE YEAR

We constructed GGE biplots on a per-year basis using all the hybrids planted within and across zones. The average angle, the standard deviation, and 95% confidence intervals were calculated and are shown in Table 2: A-B. While helpful in identifying patterns in the data, as previously established by Yan et al. (2001), year-to-year interactions or single-year plots are not as meaningful or repeatable as multi-year GGE biplots.

County Combo	Average	Stdev	CI
Allegan: Huron	66.9	26.4	17.2
Allegan: Ingham	39.9	24.2	15.8
Allegan: Mason	23.2	18.3	12.0
Allegan: Montcalm	35.3	21.5	14.0
Allegan: Saginaw	31.6	19.1	13.3
Huron: Ingham	55.1	37.8	24.7
Huron: Mason	56.8	29.0	18.9
Huron: Montcalm	66.7	30.3	22.5
Huron: Saginaw	63.2	37.2	25.8
Ingham: Mason	39.0	19.2	12.5
Ingham: Montcalm	31.7	20.1	14.9
Ingham: Saginaw	52.4	35.6	24.7
Mason: Montcalm	33.1	28.2	20.9
Mason: Saginaw	48.7	22.5	15.6
Montcalm: Saginaw	42.3	42.2	33.7

County Combo	Average	Stdev	CI
Allegan: Branch	68.9	42.1	29.2
Allegan: Cass	42.9	34.8	24.1
Allegan: Ingham	38.2	34.9	24.2
Allegan: Saginaw	67.8	34.9	24.2
Allegan: Washtenaw	58.9	39.4	31.6
Branch: Cass	49.4	35.4	24.5
Branch: Ingham	68.8	56.9	39.4
Branch: Saginaw	62.8	49.2	34.1
Branch: Wash	33.1	34.1	27.3
Cass: Ingham	48.2	32.1	22.2
Cass: Saginaw	53.6	27.3	18.9
Cass: Washtenaw	64.8	29.5	23.6
Ingham: Saginaw	52.8	41.1	28.5
Ingham: Washtenaw	66.6	55.4	44.4
Saginaw: Washtenaw	65.2	56.7	45.4

Table 2.2 A-B: Average Angle of Location Combinations
The average angle, standard deviation, and confidence interval for each location combination using single year data. As expected, variation is high.

MULTI-YEAR

We generated GGE biplots by combining all the years, estimating the overall G and GxE effects across the nine years. The angles within/between zones were calculated and placed in Table 2.3 & 2.4. We assume that with additional hybrids available, within-zone variation will be more accurate than between zones. Assuming this, the angle will not change in a data subset, and we can therefore verify the environment's location by comparing the within-zone locations by themselves with the within-zone locations on the between-zone plots (Table 2.5).

Zone 1		Zone 2		Zone 3	
County	Angle	County	Angle	County	Angle
Branch-Cass	87.73	Allegan-Ingham	35.1	Huron-Montcalm	147.5
Branch-Washtenaw	116.76	Allegan-Saginaw	58.08	Huron-Mason	96.7
Cass-Washtenaw	29.03	Ingham-Saginaw	93.18	Montcalm-Mason	115.7

Table 2.3: Angle between within zone location across years
Angles of correlation between location combination only using within-zone hybrids. These will be more accurate than between-zone, as there are more hybrids tested. A color key for within zone combinations. green equates to a Zone 1 by Zone 1 location, orange equates to a Zone 2 by Zone 2 location, and blue equals a Zone 3 by Zone 3 location.

Zone 1/2		Zone 2/3	
County	Angle		Angle
Branch-Cass	75.3	Allegan-Ingham	17.1
Branch-Washtenaw	125.0	Allegan-Saginaw	70.8
Branch-Allegan	84.8	Allegan-Huron	61.2
Branch-Ingham	84.2	Allegan-Montcalm	162.2
Branch-Saginaw	118.6	Allegan- Mason	60.0
Cass-Washtenaw	49.7	Ingham-Saginaw	88.0
Cass-Allegan	9.5	Ingham-Huron	44.0
Cass-Ingham	8.8	Ingham-Montcalm	145.0
Cass-Saginaw	43.3	Ingham-Mason	77.1
Washtenaw-Allegan	40.2	Saginaw-Huron	131.9
Washtenaw-Ingham	27.6	Saginaw-Montcalm	127.0
Washtenaw-Saginaw	40.8	Saginaw-Mason	10.7
Allegan-Ingham	0.6	Huron-Moncalm	101.0
Allegan-Saginaw	33.8	Huron-Mason	121.0
Ingham-Saginaw	34.4	Montcalm-Mason	138.0

Table 2.4: Angle between zone location across years
Angles of correlation between location combinations using only between-zone hybrids. There are several correlations across zone boundaries, indicating that the current zones are not optimally defined. A color key for within zone combinations. green equates to a Zone 1 by Zone 1 location, orange equates to a Zone 2 by Zone 2 location, and blue equals a Zone 3 by Zone 3 location.

County Combo	Angle Difference
Branch-Cass	-12.4
Branch-Washtenaw	8.2
Cass-Washtenaw	20.7
Allegan-Ingham	34.5
Allegan-Saginaw	24.3
Ingham-Saginaw	58.8
Huron-Montcalm	46.5
Huron-Mason	-24.3
Montcalm-Mason	-22.3
Allegan-Ingham	18.0
Allegan-Saginaw	-12.7
Ingham-Saginaw	5.2

Table 2.5: Angle Difference between Subsets
Difference between angles from within-zone and between-zone estimates using only shared hybrids. With the exception of Zone 2 and Zone 1 vs 2, angles calculated from between-zones show similar trends to within-zone estimates, bolstering confidence in their accuracy. A color key for within zone combinations. green equates to a Zone 1 by Zone 1 location, orange equates to a Zone 2 by Zone 2 location, and blue equals a Zone 3 by Zone 3 location.

In this study, in Zone 1, test sites in Branch and Washtenaw counties have a minimally negative correlation (116.8°), while Cass and Washtenaw are positively correlated, around 29° (Table 2.1). Conversely, Cass and Branch had no correlation at a value of 87.7° (Table 2.3). In Zone 2, the Ingham and Saginaw locations did not correlate (93.2°); however, they positively correlated with the Allegan location (35° and 58°, respectively) (Table 2.3). Finally, Zone 3 contained the most diverse environments, having no or negative correlations between all environments (Table 2.3).

When comparing hybrids planted in Zone 1 and 2 (Table 2.3 & Figure 2.3A), we concluded that the subset of hybrids planted in Zone 1 contained a similar trend of GGE interactions, but those in Zone 2 did not. Cass County, therefore, is more correlated with the Zone 2 locations than any of the locations in Zone 1 (Cass-Allegan: 9.5^{o}, Cass-Ingham: 8.8^{o}, Cass-Saginaw: 43.3^{o} vs. Cass-Washtenaw 49.7^{o} and Cass-Branch 75.3^{o}). In addition,

Washtenaw County positively correlates with other Zone 2 locations; however, Branch County did not positively correlate with any Zone 2 locations.

When comparing hybrids planted in Zone 2 and 3 (Table 2.4 & Figure 2.3B), the trend of GGE interactions within the subsets of both Zones was stable. We identified that the Mason location was highly positively correlated with the Saginaw location, and the Huron location positively correlates with the Ingham location. We also concluded that the Allegan location positively correlates with locations in both Huron and Mason counties. The Montcalm location negatively correlates with all locations in Zone 2 and Zone 3.

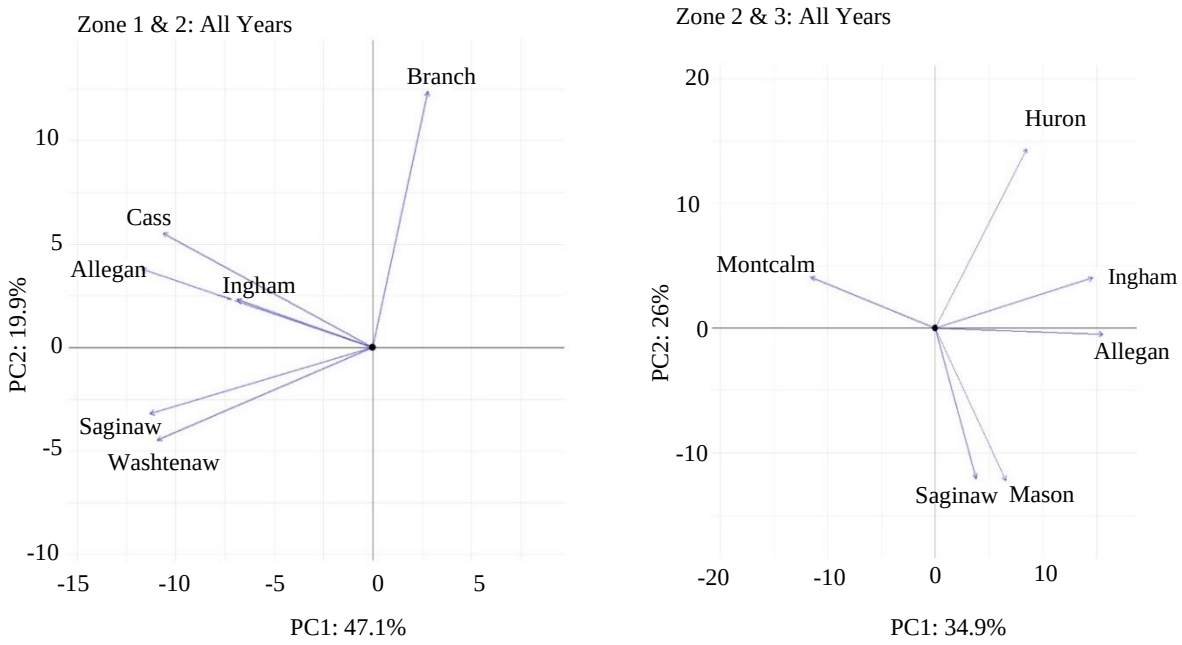

Figure 2.3 A-B: GGE Biplot of Between Zones Across All Years
Within Zone 1, Branch is distinct from Washtenaw and Cass. In Zone 2, Allegan and Ingham are similar. All locations are distinct in Zone 3. Between Zones 1 and 2, Cass behaves more like Zone 2, as does Washtenaw to a lesser extent. When examining Zones 2 & 3 together, Saginaw and Mason are highly similar, while Ingham and Huron are positively correlated. Allegan trends towards Huron and Mason, while Montcalm is completely unique among tested locations.

OPTIMAL REPLICATION NUMBER

Because of the unbalanced nature of MCPT design, it is impossible to calculate the variance components across all years and locations. To counteract this, hybrids were subsetted into two to three-year increments to generate complete datasets for analysis. A goal repeatability level of 75% of the maximum was used to find the optimal number of replications; as 75% of the max is the upper limit, repeatability can be improved by increasing the number of test environments/replications (Yan et al., 2015).

The replication needed at each location was first calculated separately using the year variable as the 'environment' (Table 2.6 & Appendix: Table B.1). In all cases, the median optimal location replications were 2.9-5.7. However, Yan (2021) discovered that this methodology was less accurate than the multi-year and location model.

Year	Allegan	Branch	Cass	Huron	Ingham	Mason	Saginaw	Washtenaw	Montcalm
2011-2012	2.91	2.98	NA	4.87	2.93	4.7	3.72	5.55	NA
2012-2013	4.35	4.3	4.82	1.99	NA	3.23	NA	NA	NA
2012-2014	2.04	1.55	1.89	NA	17.04	7.2	NA	3.87	NA
2013-2014	4.33	4.48	7.53	2.89	3.22	3.35	NA	3.67	6.7
2013-2015	2.73	4.54	5.74	6.9	2.1	8.75	NA	5.4	5.05
2014-2015	2.41	3.57	7.54	NA	2.75	4.88	NA	5.23	2.14
2014-2016	1.39	3.15	6.25	NA	3.14	5.46	NA	6.68	4.12
2015-2016	3.18	4.74	9.62	5.75	3.36	5.06	12.45	5.94	4.88
2015-2017	3.19	NA	15.75	NA	5.45	4.82	NA	8.05	3.87
2016-2017	3.61	2.85	NA	2.46	6.37	10.34	NA	NA	3.56
2016-2018	NA	1.58	4.75	6.01	7.12	1.35	8.77	NA	NA
2017-2018	3.78	1.94	3.98	7.16	7.05	2.41	4.6	NA	NA
2017-2019	2.93	1.35	1.94	NA	NA	2.73	5.06	NA	NA
2018-2019	2.53	2.82	2.49	6.11	NA	3.33	NA	NA	NA
Average	3.03	3.07	6.03	4.90	5.50	4.83	6.92	5.55	4.33
Median	2.93	2.98	5.28	5.75	3.36	4.76	5.06	5.48	4.12

Table 2.6: Optimal number of replications needed at each location
0.75 repeatability per year combination. The variation in the number of replications was high as expected. A value of N.A. was assigned when there were not enough hybrids in the trial to generate enough degrees of freedom for the linear model to parse out all the variance components.

The replications needed in each zone across years were then calculated, allowing for G x E, G x Y, and G x E x Y interactions (Table 2.7 & Appendix: Table B.2). The average for all zones was three replications; however, in Zone 1, only 1.8 replications were needed to reach the desired threshold, while 4.4 replications were needed in Zone 3.

Year	Z1	Z2	Z3
2011-2012	3.13	2.00	2.20
2012-2014	1.08	1.99	NA
2013-2014	1.91	1.62	2.12
2013-2015	1.57	1.18	5.91
2014-2015	1.70	1.43	NA
2014-2016	1.77	1.22	5.01
2015-2016	NA	1.60	3.16
2015-2017	2.34	1.94	9.59
2016-2017	1.86	2.92	4.92
2016-2018	2.21	3.19	2.82
2017-2018	2.28	4.59	3.87
2017-2019	1.00	NA	NA
2018-2019	1.69	NA	NA
Averages	**1.88**	**2.15**	**4.40**

Table 2.7: Optimal number of replications needed at each Zone
0.75 repeatability per year combination. The variation in the number of replications was lower than the single location Four replications are currently used in data collection, but it would seem that three would be sufficient at least in Zones 1 and 2. A value of N.A. was assigned when there were not enough hybrids in the trial to generate enough degrees of freedom for the linear model to parse

DISCUSSION:

GGE BIPLOTS

Every performance trial program aims to test hybrids in a range of similar and different environments to depict hybrid yield accurately. To reach this goal, programs need to keep locations similar enough to be compared, however different enough not to be redundant. Knowing this, we would hypothesize that the locations within-zone GGE angles would differ; however, they are more positively correlated than locations outside these zones. We tested this theory with the GGE biplots and got mixed results.

Based on analysis within zones, we can infer:

- Zone 1: Branch County is significantly different from Washtenaw and Cass Counties.
- Zone 2: Allegan and Ingham Counties are similar.
- Zone 3: All locations are distinct.

Based on the between-zone tests, the analysis is less conclusive. The subset of hybrids planted in Zones 1 and 3 have a similar trend of GGE compared to the whole set; however, that is not the case for Zone 2 locations. Based on this, we are less confident about the definition of the boundaries of Zone 2 relative to the neighboring zones. However, if they are confidently accurate, we can assume:

- Zone 1 & 2: Cass reacts like Zone 2, and Washtenaw trends in that direction.
- Zone 2 & 3: Saginaw and Mason are highly similar, while Ingham and Huron positively correlate. We also can infer that Allegan is more similar to Huron and Mason. We also infer that Montcalm is not comparable to any location tested.

These results suggest that an optimal allocation of resources maximizing differences between zones involves the following changes to each zone:

- Zone 1: Cass is removed due to redundancy with Washtenaw, Allegan, and Ingham
- Zone 2: Allegan or Ingham is removed as they are similar
- Zone 3: Mason is removed as it is similar to Saginaw

OPTIMAL REPLICATION NUMBER

Overall, the average optimal number of replications needed to obtain the target repeatability measure of 75% of the maximum at all individual locations was more than four. However, this number varied significantly by year within each zone. For instance, at Branch and Cass County locations in 2017-2019, less than two replications were needed to reach the

threshold; however, more than 4.5 replications were needed at those exact locations in 2013-2014. This result confirms what Yan (2019) reported: models containing only a genotype and environment effect would overpredict the number of replications needed to obtain the repeatability measure.

The average number of replications required across a maturity Zone in all trials was 2.8. While year-dependent, the average number of replications for Zones 1 and 2 were less than 3, at 1.88 and 2.15, respectively, while 4.40 were needed in Zone 3. Based on the GGE biplots, we know that Montcalm County is unlike all other locations. Therefore, if Montcalm is removed from the replication, the average number of replications needed in Zone 3 shifts to 2.56, bringing the average replications needed across the trial to 2.2 vs. 2.8 replications.

One year-zone combination had abnormally high optimal replication values: Zone 3: 2015-2017. This anomaly occurred because nearly all the variance in this combination was in the location or year (not both), leaving little variance in the interaction terms and genotype. This result leads to a reasonable maximum achievable across-location heritability (H_{mmly}) but a proportionally higher Q term ($\frac{\sigma_{Error}}{\sigma_{Gen*\#ofLoc*\#Years}}$) than expected which in turn increased the required number of replications.

APPENDIX

Table A.1: Inbred Names
All inbred and all GEM line names used subset by year and environment used

MI19 & 20

- CO192
- CML 228
- W812G
- R177
- ND167
- T232
- DK3IBZ2
- BGEM-0018-S
- BGEM-0019-S
- BGEM-0022-S
- BGEM-0023-S
- BGEM-0025-S
- BGEM-0026-S
- BGEM-0027-S
- BGEM-0028-S
- BGEM-0029-S
- BGEM-0030-S
- BGEM-0031-S
- BGEM-0032-S
- BGEM-0033-S
- BGEM-0034-S
- BGEM-0036-S
- BGEM-0037-S
- BGEM-0039-N
- BGEM-0040-N
- BGEM-0041-S
- BGEM-0042-S
- BGEM-0059-S
- BGEM-0063-N
- BGEM-0070-S
- BGEM-0071-S
- BGEM-0072-S
- BGEM-0073-S
- BGEM-0083-S
- BGEM-0088-N
- BGEM-0089-N
- BGEM-0090-N
- BGEM-0094-S
- BGEM-0095-S
- BGEM-0097-S
- BGEM-0099-S
- BGEM-0100-S
- BGEM-0102-N
- BGEM-0110-N
- BGEM-0120-N
- BGEM-0121-N
- BGEM-0122-N
- BGEM-0123-N
- BGEM-0125-N
- BGEM-0127-N
- BGEM-0129-N
- BGEM-0130-N
- BGEM-0134-S
- BGEM-0136-S
- BGEM-0137-S
- BGEM-0138-S
- BGEM-0162-S
- BGEM-0164-S
- BGEM-0165-S
- BGEM-0166-S
- BGEM-0167-S
- BGEM-0169-S
- BGEM-0170-S
- BGEM-0178-S
- BGEM-0179-S
- BGEM-0182-N
- BGEM-0184-N
- BGEM-0186-S
- BGEM-0187-S
- BGEM-0188-S
- BGEM-0200-S
- BGEM-0201-N
- BGEM-0202-N
- BGEM-0215-N
- BGEM-0216-N
- BGEM-0218-S
- BGEM-0221-S
- BGEM-0222-S
- BGEM-0226-S
- BGEM-0228-N
- BGEM-0233-S
- BGEM-0235-N
- BGEM-0236-S
- BGEM-0237-N
- BGEM-0239-N
- BGEM-0240-N
- BGEM-0242-N
- BGEM-0243-S
- BGEM-0246-N
- BGEM-0247-N
- BGEM-0248-N
- BGEM-0250-S
- BGEM-0252-S
- BGEM-0253-N
- BGEM-0254-S
- BGEM-0255-S
- BGEM-0256-N
- BGEM-0259-N
- BGEM-0260-N
- BGEM-0261-S
- BGEM-0262-S
- BGEM-0263-S
- BGEM-0264-S
- BGEM-0266-S
- BGEM-0269-S
- BGEM-0272-S
- GEMN-0048
- GEMN-005
- GEMN-0077
- GEMN-0083
- GEMN-0094
- GEMN-0095
- GEMN-0096
- GEMN-0110
- GEMN-0117
- GEMN-0140
- GEMN-0141
- GEMN-0144
- GEMN-0145
- GEMN-0156
- GEMN-0157
- GEMN-0186
- GEMN-0187
- GEMN-0190
- GEMN-0191
- GEMN-0192
- GEMN-0193
- GEMN-0202
- GEMN-0221
- GEMN-0225
- GEMN-0229
- GEMN-0249
- GEMN-0252
- GEMN-0285
- GEMN-0286
- GEMN-0302
- GEMN-0309
- GEMS-0050
- GEMS-0051
- GEMS-0052
- GEMS-0053
- GEMS-0063
- GEMS-0064
- GEMS-0066
- GEMS-0072
- GEMS-0073
- GEMS-0074
- GEMS-0075
- GEMS-0084
- GEMS-0085
- GEMS-0086
- GEMS-0093
- GEMS-0100
- GEMS-0113
- GEMS-0115
- GEMS-0118
- GEMS-0142
- GEMS-0143
- GEMS-0149
- GEMS-0150
- GEMS-0160
- GEMS-0161
- GEMS-0162
- GEMS-0163
- GEMS-0175
- GEMS-0176
- GEMS-0180
- GEMS-0181
- GEMS-0182
- GEMS-0183
- GEMS-0184
- GEMS-0185
- GEMS-0188
- GEMS-0189
- GEMS-0200
- GEMS-0201
- GEMS-0202
- GEMS-0203
- GEMS-0222
- GEMS-0223
- GEMS-0224
- GEMS-0226
- GEMS-0235
- GEMS-0237
- GEMS-0240
- GEMS-0241
- GEMS-0250
- GEMS-0251
- GEMS-0263
- GEMS-0265
- GEMS-0275
- GEMS-0276
- GEMS-0277
- GEMS-0278
- GEMS-0279
- GEMS-0280
- GEMS-0281
- GEMS-028

Table A.1 (cont'd)

- GEMS-0283
- GEMS-0290
- GEMS-0299
- GEMS-0307
- GEMS-0308

MI19, MI20, IN20, & WI20

- NC230
- NC232
- Oh43
- H95
- N28
- K55
- Yong 28
- YE 4
- A401
- A674
- A680
- MS72
- MS223
- MS225
- Va99
- WXB6
- 33-16
- H14
- H121
- CO257
- F2834T
- H91
- Ky21
- M37W
- N6
- N28Ht
- Pa762
- R4
- T234
- W603S
- W809G
- W810G
- W814G
- W817G
- W818G
- W819G
- CI 21E
- CI 28A
- K150
- K155
- Oh33
- K4
- W23
- W24
- W9
- A
- R113
- R134
- R197
- B8
- B10
- A258
- A659
- A415-1-3 INBRED
- KUNG-70
- YING-55
- TZU-CHIAO-HSI-WU 105
- YE-CHI-HUNG
- 4578 INBRED
- Chi-tan 120
- Pa392
- Pa468
- Pa880
- NC264
- SD44
- Pa891
- R227
- SD101
- SD102
- LH195
- LH204
- LH205
- LH211
- LH127
- LH163
- LH206
- LH190
- LH194
- LH202
- LH191
- LH192
- PHK46
- PHK56
- PHN46
- PHP38
- PHP76
- PHW51
- PHW86
- LH208
- Lp215D
- PHJ89
- PHJ90
- PHK93
- PHM81
- PHN66
- PHR03
- PHR58
- PHR61
- PHT11
- PHW30
- B66
- B68
- B73
- DE811
- LH164
- LH214
- 911
- 912
- LH199
- LH216
- Mo17
- ICI 193
- ICI 441
- ICI 986
- CS405
- MQ305
- OS602
- PHBA6
- PHBW8
- PHK74
- PHN18
- PHP85
- PHPR5
- PHR31
- PHT69
- PHV53
- PHVA9
- PHWG5
- 904
- LH172
- LH223
- LH217
- LH200
- LH167
- B97
- B101
- PHVJ4
- PHAW6
- PHEM9
- PHEW7
- PHHB9
- PHHH9
- PHJR5
- PHKE6
- PHMK0
- PHV57
- PHW80
- CQ806
- LH218
- LH169Ht
- LH185
- PHAA0
- PHTE4
- PHTD5
- AM0776
- OQ601
- LH189Ht
- LH231
- CM105
- A632
- A679
- A682
- W64A
- W182BN
- ZS1791
- Hi26
- B104
- B106
- N501
- N534
- N538
- N545
- N209
- B107
- B109
- Seagull Seventeen
- FR19
- LH39
- LH51
- PH207
- PHB47
- F42
- AS6103
- LH93
- DJ7
- DKIB014
- LH150
- DK4676A
- LH57
- LH52
- NK794
- LP5
- LH146Ht
- LH60
- NS701
- DK78371A
- DKPB80

Table A.1 (cont'd)

- PHK29
- PHR25
- 793
- PHK42
- PHK76
- PHN11
- PHV63
- W8304
- 11430
- PHT10
- PHT60
- OQ603
- NKH8431
- S8324
- S8326
- 1538
- CR14
- WIL901
- WIL903
- J8606
- L 127
- L 135
- L 139
- W8555
- PHJ33
- PHJ75
- PHM10
- PHN73
- PHN82
- PHR63
- PHT22
- PHV37
- PHW03
- SD40
- N7A
- A661
- A662
- B77
- B79
- B87
- B75
- NY6371
- DE3
- DE4
- LH299
- LH143 (Maintainer)

IN20 & WI20

- W182B
- Mt42
- CM37
- A96
- A155
- A305
- A321
- A334
- A340
- A344
- A508
- A548
- A572
- MS24A
- MS116
- MS153
- MS1334
- CM99
- CO117
- C49A
- CM7
- CO125
- MEF156-55-2
- Mo44
- W802G
- W117HT
- W37A
- R53
- ND230
- 4F-306 108
- W552
- B90
- PHW06
- B46
- ND245
- ND246
- ND249
- ND251
- ND259
- ND260
- Mo16W
- A554
- A654
- ZS01250
- OC19
- LH74
- PHB09
- CR1HT
- PHK05
- PHP02
- ND287

MI19, WI20 & IN20

- CH157
- A635
- Oh7B
- CM48
- A427
- A649
- A673
- CG10
- MS142
- Tr
- Ill.Hy
- CO158
- W604S
- W605S
- W811G
- WR3
- R181B
- SD15
- B14
- YANG
- LH160
- LH162
- RS 710
- OQ403
- PHFA5
- PHRE1
- PHKM5
- LH175
- LH149
- A15
- A651
- MS67
- Eng-Li Chih
- ND262
- LH220Ht
- ND265
- L 155
- LH222
- OQ101
- PHT73
- PHTM9
- LH165
- LH145
- PHJ40
- LH61

MI19

- BGEM-0270-S
- GEMN-0060
- GEMS-0065
- GEMS-0091

MI19, MI20 & IN20

- Ky226
- U267Y
- W815G
- LH196
- NKBCC03
- DK6F629
- DK6M502A
- DKNL001
- DK29MIBZ2
- DKF118
- DK83IBI3
- Mo3
- ICI 581
- DKMBWZ
- DKAQA3
- DK91IFC2
- DK2FADB
- DKMM501D
- DK3IJI1
- DK3IIH6
- DK8M129
- DK2MCDB
- N199
- NKNP901
- DK84QAB1
- DKMBZA
- DKWDAD1
- DK8F196
- LH260
- Oh7
- KO679Y
- N215
- PHG39
- DKHBA1
- DK78010
- DKIBB15
- DKIBC2
- LH59
- PHV78
- PHN47
- WIL900
- PHP60
- DK2FACC
- NK792

MI19 & WI20

- CSJ3
- M162W
- 3IBZ2
- 779
- LH85

Table A.1 (cont'd)

MI20, IN20, & WI20

- GE54
- GE129
- NC13
- Va17
- Va52
- C103
- Wf9
- A634
- C123
- Va22
- Mo23W
- A73
- H114
- Pa405
- NORTH 7
- Bei 10 = North 10
- 52220
- Huanyao
- Huobai
- A239
- A322
- A374
- A556
- A627
- A672
- MS71
- MS106
- MS132
- MS200
- MS221
- MS222
- MS224
- MS226
- CI 540
- CI 3A
- CI 40H
- CI 187-2
- H5
- H49
- H113
- H122w
- H124w
- CH753-4
- CO256
- CO258
- B73Htrhm
- B164
- CH701-30
- K64
- K148
- MoG
- NC294
- NC302
- NC306
- NC310
- NC318
- NC326
- NC328
- NC342
- NC358
- NC368
- Oh43E
- Va14
- Va85
- Yu796 NS
- CI 91B Goodman-Buckler
- Mo28W
- Mo39
- W601S
- W602S
- W813G
- W816G
- W821G
- Fe
- INB 101LFY/LFY (A632 X M16 S5)
- K41
- J47
- Oh40B
- W22
- W32
- W182E
- M14
- R30
- R71
- R78
- R101
- A71
- AusTRCF 306238
- B7
- L 289
- L317
- Os420
- A648
- INBRED 109
- INBRED 141
- INBRED 2-687
- INBRED 305
- INBRED 309
- INBRED 321
- 4F-234 BX 4
- NY 159 (Neveh Year)
- NY 166 (Neveh Year)
- T141
- WU-TAN-TZAO
- A797NW
- SD42
- B91
- R225
- R226
- LH210
- LH193
- PHN34
- PHV07
- LH128
- LH213
- PHR55
- B64
- B14A
- B54
- B57
- B76
- H110
- NC250
- B88
- N192
- N193
- LH215
- LH197
- LH198
- Mo1W
- Mo5
- ICI 740
- ICI 893
- NQ508
- PHT47
- Pa778
- CS608
- LH224
- LH166
- LH184
- LH183
- PHN41
- PHW53
- CQ702RC
- NQ402
- N200
- N201
- LH225
- ZS635
- LH186
- PHBB3
- PHEG9
- PHHB4
- Mo45
- Mo46
- Mo47
- LH250
- L222
- B99
- LH188
- H105W
- H84
- H99
- A619
- Mo24W
- Pa91
- Va26
- W153R
- R229
- Hi28
- B105
- N523
- N540
- N542
- N217
- N218
- LP1 NR Ht
- LH38
- LH119
- LH132
- PHG50
- PHG80
- LH123HT
- PHG71
- LH82
- PHG83
- AS5707
- PHG29
- DK78002A
- LH54
- PHG47
- PHZ51
- PHR36
- PHW17
- 764
- NK807

Table A.1 (cont'd)

- DKFBHJ
- PHG86
- NK740
- 787
- PHT55
- PHH93
- PHR32
- PHW52
- NS501
- 4N506
- PHJ31
- PHJ70
- PHK35
- PHM57
- PHP55
- PHW43
- LH284
- B42
- N527
- DE2
- A663
- B84
- B85
- LH1

MI19, MI20, & WI20

- BCC03
- 6F629
- 6M502A
- NL001
- 29MIBZ2
- F118
- MBWZ
- AQA3
- 91IFC2
- 2FADB
- MM501D
- 3IJI1
- 3IIH6
- 8M129
- 2MCDB
- NP901
- 84QAB1
- MBZA
- WDAD1
- 8F196
- IBB15
- IBC2
- 2FACC
- 792

WI20

- 78004
- 78010
- 04033V
- NC236
- Va59
- A641
- R168
- C68
- Ia 453
- A188
- N197
- Os426
- W59E
- EAST 028
- A208
- C42
- MS12
- MS211
- B2
- H71
- H96
- CH711-10
- CL17
- CL18
- CL27
- CMV3
- CO216
- CO236
- CO237
- CO245
- A441.5
- CH9
- CO106
- CO255
- E2558W
- EP1
- Ia5125B
- Il14H
- Il 101T
- Ki11
- NC344
- WD
- Il778d
- W803G
- WD456
- B112
- BSSSC0001
- BSSSC0002
- BSSSC0003
- BSSSC0005
- BSSSC0006
- BSSSC0007
- BSSSC0008
- BSSSC0009
- BSSSC0012
- BSSSC0013
- BSSSC0015
- BSSSC0016
- BSSSC0018
- BSSSC0019
- BSSSC0020
- BSSSC0021
- BSSSC0022
- BSSSC0023
- BSSSC0024
- BSSSC0025
- BSSSC0026
- BSSSC0028
- BSSSC0029
- BSSSC0030
- BSSSC0031
- BSSSC0033
- BSSSC0034
- BSSSC0036
- BSSSC0037
- BSSSC0038
- BSSSC0039
- BSSSC0040
- BSSSC0041
- BSSSC0042
- BSSSC0043
- BSSSC0044
- BSSSC0045
- BSSSC0046
- BSSSC0048
- BSSSC0050
- BSSSC0051
- BSSSC0052
- BSSSC0053
- BSSSC0054
- BSSSC0056
- BSSSC0057
- BSSSC0058
- BSSSC0060
- BSSSC0061
- BSSSC0062
- CG106
- CG108
- CG65
- CI 64
- F431
- I224
- ICI581
- Ill. 12E
- NC412
- NC472
- Tr 9-1-1-6
- W100010003
- W100010007
- W100010009
- W100010010
- W100010012
- W100010016
- W100010018
- W100010030
- W100010031
- W100010040
- W100020004
- K47
- W703
- ND283
- A12
- A171
- 4226
- 4F-35 BK
- 4F-403 JV 15
- F2
- F7
- FC46
- A385
- CR 22 INBRED
- NO. 380
- T9
- G22 T122
- T146
- T242
- CA-4
- B-18
- INBR.FR.SUPERG
- B-28
- 80-2
- U 123
- 4554 INBRED
- NC258
- NC262
- S 56
- SD107
- NP87
- HP72-11
- B65
- B70
- ND247
- T8
- LH209

Table A.1 (cont'd)

- Mo7
- PHDD6
- PHGG7
- PHGW7
- AR228
- B98
- 907
- PHEM7
- ML606
- 4722
- HP301
- Sg 1533
- P39
- IA2132
- Va35
- Va102
- N211
- PHG72
- IB02
- 790
- PHW65
- PHM49
- B110
- B111
- B113
- B114
- B115
- B118
- B119
- B120
- B121
- PHWRZ
- Ny821
- I29
- IDS28
- IDS69
- IDS91
- SG 18
- SG 30A
- NC290A
- NC314
- NC362
- NC364
- CI31A
- W 7151
- CI 82B

MI20 & IN20

- VaW6
- Va38
- Tx303
- 38-11
- H52
- CML 91
- CML 154Q
- CML 218
- CML 220
- CML 322
- NC260
- NC324
- NC338
- NC340
- NC348
- NC356
- NC298
- Mo30W
- A3G-3-3-1-313
- F44
- K201
- C102
- INBRED 100
- PHJ65
- DKFBLL
- DKFBLA
- LH181
- LH212Ht
- DKMBUB
- B37
- DKMM402A
- DKLIBC 4
- Mo15W
- Mo13
- PHGV6
- LH159
- Oh603
- NKNP899
- DK6F545
- DK84BRQ4
- DKF274
- DK8M116
- LH252
- Ky228
- B108
- DKMDF-13D
- PHG35
- DKMB
- PHG84
- LH156
- DKMBPM
- PHT77
- 2369
- DK2MA22
- DK6M502
- DK78551S
- DK87916W
- DKHB8229
- DKIBB14
- DKMBST
- WIL500
- E8501
- PHR62
- PHW20
- DE1
- PH5HK

MI20 & WI20

- FBLL
- FBLA
- MBUB
- MM402A
- LIBC 4
- NP899
- 6F545
- 84BRQ4
- F274
- 8M116
- MDF-13D
- DKMBNA
- MBPM
- 2MA22
- 6M502
- 87916W
- HB8229
- IBB14
- MBST

IN20

- NC350
- CML 395
- F115
- Mp339
- B52co
- NK907
- DKIB02

Figure A.1: Quantile-Quantile-Plot of AUDPC6 GWAS
Quantile plot of FarmCPU GWAS. The graph shows great control of the trait with a tail only at the end of the line

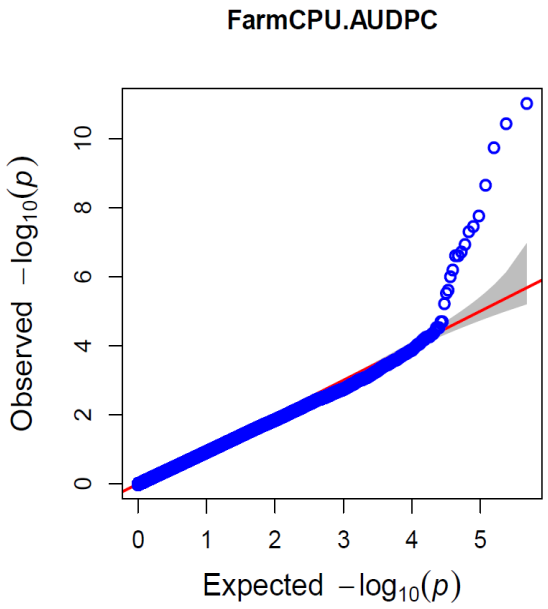

Table A2: All Significant SNPs

Table of all significant SNPS identified by the FarmCPU model of GWAS. The chromosome, location on se chromosomes, trait SNP was significant for, effect said SNP had on trait, and minor allele frequency are provided for the 80 SNPs

Chr #	SNP Loc	Effect	Trait	MAF
1	27701677	-1.613	AUDPC5	0.34
1	27820181	-0.542	TS5	0.11
1	42465699	0.484	TS5	0.04
1	43291770	0.384	IN_TS3	0.32
1	46189301	-6.457	IN_AUDPC	0.41
1	78350595	-3.381	AUDPC6	0
1	185912407	-4.908	AUDPC6	0
1	190274005	-1.956	AUDPC5	0.27
1	197395431	0.004	WI_TS	0.33
1	197546543	-0.065	TS4	0.34
1	204751114	0.008	WI_TS	0.07
1	209550543	-3.00 & -0. 47	AUDPC5 & TS5	0.04
1	254167519	-0.624	IN_TS3	0.09
1	293010669	-10.537	IN_AUDPC	0.06
2	10322736	-0.962	TS6	0.11
2	10597616	0.715 & 3.803	TS6 & AUDPC6	0.38
2	22932891	0.067	TS3	0.15
2	31137464	12.565 & 0.58	IN_ AUDPC & IN2	0.08
2	217976609	0.158	TS4	0.03
2	228718047	-0.555	IN_TS2	0.09
3	699520	0.009	WI_TS	0.09
3	1009710	-7.125	IN_AUDPC	0.25
3	5489073	0.116	TS3	0.05
3	6842175	-0.017	TS2	0.16
3	54002361	-0.019	TS2	0.27
3	96079038	2.954	AUDPC5 & AUDPC6	0.07
3	96079096	2.954	AUDPC5 & AUDPC6	0.07
3	160732338	-0.463	IN_TS3	0.13
3	198494090	-0.458	TS6	0.48
3	214384773	0.005	WI_TS	0.45
3	223075613	-11.349	AUDPC6	0.04
4	11907047	0.544	TS6	0.34
4	164542828	-0.413	IN_TS2	0.17
4	175609393	-1.83, -6.03 & -0.66	AUDPC5, AUDPC6 & TS6	0.2
4	184920580	0.094 & -1.89	TS4 & AUDPC5	0.09
4	244571379	4.959	AUDPC6	0.37
4	244833803	0.081	TS4	0.17

Table A.2: (cont'd)

5	2833037	-5.808	AUPDC6	0
5	3606546	-0.03	TS2	0.04
5	11962287	-0.581	TS6	0.32
5	12112021	0.043	TS3	0.24
5	16524103	-1.239	AUDPC5	0.46
5	18227999	0.006	WI_TS	0.13
5	31711295	0.264	IN_TS3	0.47
5	73931011	0.01	WI_TS	0.04
5	206965116	0.156	TS4	0.04
5	209550543	0.03	TS2	0.04
6	22923136	-2.139	AUDPC5	0.13
6	100474708	0.57	TS5	0.05
6	116217531	0.454	TS6	0.48
6	129103197	-8.70 & -0. 486	IN_ AUDPC & IN2	0.1
6	154791110	0.68	TS6	0.27
6	161068264	-1.093	TS6	0.06
6	169737760	0.327	IN_TS3	0.41
7	4195485	-2.257	AUDPC5	0.14
7	5345453	-0.852	TS6	0.08
7	5373948	0.029	TS2	0.05
7	10798365	0.005	WI_TS	0.35
7	123563152	-0.733	IN_TS2	0.03
7	147634792	11.85 & 0. 58	IN_ AUDPC & IN2	0.09
7	151058142	-0.332	IN_TS3	0.48
7	171051141	-0.306	IN_TS2	0.41
8	17976817	-0.012	WI_TS	0.03
8	140400534	-0.435	TS5	0.04
8	178112585	0.01	WI_TS	0.04
9	19826467	-6.966	AUDPC6	0.1
9	24407558	-0.77	IN_TS2	0.09
9	72367542	0.007	WI_TS	0.1
9	146547949	-4.428	AUDPC6	0
9	154449335	0.01	WI_TS	0.06
10	1299595	4.203	AUDPC6	0.2
10	1500433	-0.375	TS5	0.1
10	13341469	-7.878	IN_AUDPC	0.2
10	53432205	15.294 & 1.017	IN_ AUDPC & IN2	0.04
10	77543586	-0.005	WI_TS	0.3
10	133097101	-0.533	TS6	0.28
10	137296645	-0.007	WI_TS	0.07
10	141042230	-0.093	TS4	0.12
10	143309595	-0.006	WI_TS	0.1

Table A3: Genes Located Near Significant SNPs
110 genes within 8000 base pairs (4000 on each side) of the significant SNPs identified in the GWAS analysis.

Name	Chr	Trait	SNP Loc
Zm00001d028240	1	AUDPC5	27701677
Zm00001d028241	1	AUDPC5	27701677
Zm00001d028243	1	TS5	27820181
Zm00001d028671	1	TS5	42465699
Zm00001d028690	1	IN_TS3	43291770
Zm00001d028776	1	IN_AUDPC	46189301
Zm00001d028778	1	IN_AUDPC	46189301
Zm00001d028777	1	IN_AUDPC	46189301
Zm00001d029595	1	AUDPC6	78350595
Zm00001d031317	1	AUDPC6	185912407
Zm00001d031445	1	AUDPC5	190274005
Zm00001d031651	1	WI_TS	197395431
Zm00001d031655	1	TS4	197546543
Zm00001d031871	1	WI_TS	204751114
Zm00001d032016	1	AUDPC5 & TS5	209550543
Zm00001d017869	1	TS2	209550543
Zm00001d033204	1	IN_TS3	254167519
Zm00001d033205	1	IN_TS3	254167519
Zm00001d034440	1	IN_AUDPC	293010669
Zm00001d034441	1	IN_AUDPC	293010669
Zm00001d002338	2	TS6 & AUDPC6	10597616
Zm00001d002339	2	TS6 & AUDPC6	10597616
Zm00001d002340	2	TS6 & AUDPC6	10597616
Zm00001d035356	2	AUDPC5	22923136
Zm00001d002797	2	TS3	22932891
Zm00001d032188	2	TS4	217976609
Zm00001d006856	2	TS4	217976609
Zm00001d007328	2	IN_TS2	228718047
Zm00001d039259	3	WI_TS	699520
Zm00001d039258	3	WI_TS	699520
Zm00001d039284	3	IN_AUDPC	1009710
Zm00001d039480	3	TS3	5489073
Zm00001d039481	3	TS3	5489073
Zm00001d039522	3	TS2	6842175

Table A.3: (cont'd)

Zm00001d040614	3	TS2	54002361
Zm00001d041082	3	AUDPC5 & AUDPC6	96079038
Zm00001d041082	3	AUDPC5 & AUDPC6	96079096
Zm00001d042317	3	IN_TS3	160732338
Zm00001d043389	3	TS6	198494090
Zm00001d043388	3	TS6	198494090
Zm00001d043946	3	WI_TS	214384773
Zm00001d043945	3	WI_TS	214384773
Zm00001d043944	3	WI_TS	214384773
Zm00001d044253	3	AUDPC6	223075613
Zm00001d044251	3	AUDPC6	223075613
Zm00001d048993	4	TS6	11907047
Zm00001d051612	4	IN_TS2	164542828
Zm00001d051611	4	IN_TS2	164542828
Zm00001d051610	4	IN_TS2	164542828
Zm00001d051967	4	AUDPC5 & AUDPC6 & TS6	175609393
Zm00001d052256	4	TS4 & AUDPC5	184920580
Zm00001d053981	4	AUDPC6	244571379
Zm00001d053982	4	AUDPC6	244571379
Zm00001d053997	4	TS4	244833803
Zm00001d053998	4	TS4	244833803
Zm00001d013039	5	TS2	3606546
Zm00001d013040	5	TS2	3606546
Zm00001d013452	5	TS6	11962287
Zm00001d013453	5	TS6	11962287
Zm00001d013459	5	TS3	12112021
Zm00001d013463	5	TS3	12112021
Zm00001d013461	5	TS3	12112021
Zm00001d013658	5	AUDPC5	16524103
Zm00001d013709	5	WI_TS	18227999
Zm00001d014078	5	IN_TS3	31711295
Zm00001d015064	5	WI_TS	73931011
Zm00001d015065	5	WI_TS	73931011
Zm00001d017791	5	TS4	206965116
Zm00001d017793	5	TS4	206965116
Zm00001d002325	6	TS6	10322736
Zm00001d036776	6	TS5	100474708
Zm00001d036775	6	TS5	100474708
Zm00001d038622	6	TS6	116217531
Zm00001d037215	6	TS6	116217531

Table A.3: (cont'd)

Zm00001d037550	6	IN_ AUDPC & IN_TS2	129103197
Zm00001d038334	6	TS6	154791110
Zm00001d039086	6	IN_TS3	169737760
Zm00001d018751	7	AUDPC5	4195485
Zm00001d048775	7	TS6	5329584
Zm00001d018792	7	TS6	5345453
Zm00001d018795	7	TS2	5373948
Zm00001d018961	7	WI_TS	10798365
Zm00001d020578	7	IN_TS2	123563152
Zm00001d021278	7	IN_ AUDPC & IN_TS2	147634792
Zm00001d021401	7	IN_TS3	151058142
Zm00001d021400	7	IN_TS3	151058142
Zm00001d022139	7	IN_TS2	171051141
Zm00001d008731	8	WI_TS	17976817
Zm00001d011144	8	TS5	140400534
Zm00001d011145	8	TS5	140400534
Zm00001d012660	8	WI_TS	178112585
Zm00001d045366	9	AUDPC6	19826467
Zm00001d045489	9	IN_TS2	24407558
Zm00001d045490	9	IN_TS2	24407558
Zm00001d046214	9	WI_TS	72367542
Zm00001d048315	9	WI_TS	154449335
Zm00001d048314	9	WI_TS	154449335
Zm00001d023243	10	AUDPC6	1299595
Zm00001d023258	10	TS5	1500433
Zm00001d025888	10	TS6	133097101
Zm00001d023640	10	IN_AUDPC	13341469
Zm00001d023641	10	IN_AUDPC	13341469
Zm00001d024178	10	IN2 & IN_AUDPC	53432205
Zm00001d024544	10	WI_TS	77543586
Zm00001d024545	10	WI_TS	77543586
Zm00001d025887	10	TS6	133097101
Zm00001d026060	10	WI_TS	137296645
Zm00001d026213	10	TS4	141042230
Zm00001d026308	10	WI_TS	143309595
Zm00001d026307	10	WI_TS	143309595

Table A4: Genes That Showed a Change in Expression Due to Disease
Table of genes located within 8000 bp of the significant SNPS identified by the FarmCPU model of GWAS that showed a change in expression due to an infection. The interest level was assessed using expression data from Swart et al. (2017) for up and down-regulation of the gene when infected with the fungi *Cercospora zeina* or *Colletotrichum graminicola*.

Name	Chr#	Gene Start	Gene End	Trait	SNP Loc
Zm00001d041082	3	96077271	96081198	AUDPC5 & AUDPC6	96079038
Zm00001d037550	6	129101873	129103497	IN_ AUDPC,IN2	129103197
Zm00001d042317	3	160732228	160732479	IN_TS3	160732338
Zm00001d013709	5	18220021	18227041	WI_TS	18227999
Zm00001d031655	1	197545631	197549052	TS4	197546543
Zm00001d053997	4	244821758	244829813	TS4	244833803
Zm00001d021401	7	151061639	151067413	IN_TS3	151058142
Zm00001d032188	1	216280294	216286285	TS4	216283940
Zm00001d048314	9	154446529	154450415	WI_TS	154449335
Zm00001d037215	6	116217062	116229724	TS6	116217531
Zm00001d025887	10	133094068	133097996	TS6	133097101
Zm00001d039284	3	1009359	1014295	IN_AUDPC	1009710
Zm00001d039259	3	701136	705353	WI_TS	699520
Zm00001d046214	9	72368059	72391811	WI_TS	72367542
Zm00001d026060	10	137296427	137304511	WI_TS	137296645
Zm00001d006856	2	217976403	217978586	TS4	217976609
Zm00001d047968	9	146641666	146642304	TS3	146640446
Zm00001d013039	5	3603517	3604227	TS2	3606546
Zm00001d022139	7	171049645	171052026	IN_TS2	171051141
Zm00001d023243	10	1295378	1301629	AUDPC6	1299595
Zm00001d028240	1	27695825	27697683	AUDPC5	27701677
Zm00001d048775	4	5328661	5330696	TS6	5329584
Zm00001d017791	5	206956208	206962117	TS4	206965116
Zm00001d043946	3	214388200	214392378	WI_TS	214384773
Zm00001d031651	1	197393450	197395554	WI_TS	197395431
Zm00001d031871	1	204748411	204755635	WI_TS	204751114
Zm00001d023640	10	13316511	13341579	IN_AUDPC	13341469

Table A.5: Genomic Prediction of Trait Per Algorithm
Bayes B and AUDPC6 are the best in both cases

Genomic Prediction Accuracy - Michigan - by Test						
	AUDPC6	Tar Spot 6	Tar Spot 5	Tar Spot 4	Tar Spot 3	AUDPC5
Bayes A	0.779	0.773	0.758	0.727	0.649	0.76
Bayes B	0.791	0.779	0.774	0.734	0.683	0.77
BRR	0.747	0.746	0.739	0.707	0.642	0.75
rrBLUP	0.755	0.754	0.737	0.702	0.636	0.74

Table A.6 A-D: Genomic Prediction of Trait Per Algorithm Per SNP Level

Genomic Prediction Bayes A - Michigan - by SNP Level						
A	AUDPC6	Tar Spot 6	Tar Spot 5	Tar Spot 4	Tar Spot 3	AUDPC5
# Of SNPs	Acc	Acc	Acc	Acc	Acc	Acc
10	0.633	0.524	0.431	0.392	0.428	0.569
20	0.749	0.600	0.479	0.446	0.552	0.645
50	0.729	0.714	0.740	0.686	0.634	0.710
75	0.740	0.746	0.729	0.730	0.652	0.754
100	0.811	0.778	0.717	0.756	0.639	0.750
200	0.835	0.791	0.757	0.780	0.727	0.803
300	0.824	0.847	0.811	0.760	0.742	0.815
400	0.823	0.819	0.813	0.794	0.691	0.817
500	0.815	0.834	0.820	0.796	0.729	0.801
1000	0.788	0.841	0.796	0.735	0.690	0.814
5000	0.730	0.678	0.719	0.606	0.514	0.656
10000	0.701	0.687	0.676	0.630	0.476	0.657

Table A.6 A-D: (cont'd)

Genomic Prediction Bayes B - Michigan - by SNP Level						
B	AUDPC6	Tar Spot 6	Tar Spot 5	Tar Spot 4	Tar Spot 3	AUDPC5
# Of SNPs	Acc	Acc	Acc	Acc	Acc	Acc
10	0.629	0.495	0.366	0.360	0.443	0.509
20	0.694	0.618	0.564	0.383	0.494	0.612
50	0.736	0.726	0.694	0.670	0.577	0.689
75	0.749	0.693	0.773	0.698	0.633	0.715
100	0.783	0.762	0.745	0.669	0.643	0.723
200	0.804	0.809	0.817	0.782	0.754	0.767
300	0.813	0.799	0.783	0.787	0.754	0.802
400	0.819	0.792	0.802	0.719	0.738	0.801
500	0.814	0.818	0.774	0.803	0.755	0.788
1000	0.814	0.820	0.801	0.805	0.707	0.834
5000	0.818	0.816	0.814	0.741	0.712	0.797
10000	0.758	0.757	0.739	0.670	0.556	0.818

Genomic Prediction BBR - Michigan - by SNP Level						
C	AUDPC6	Tar Spot 6	Tar Spot 5	Tar Spot 4	Tar Spot 3	AUDPC5
# Of SNPs	Acc	Acc	Acc	Acc	Acc	Acc
10	0.572	0.495	0.348	0.412	0.476	0.507
20	0.711	0.615	0.426	0.513	0.561	0.657
50	0.750	0.719	0.709	0.686	0.624	0.746
75	0.721	0.741	0.742	0.755	0.638	0.789
100	0.779	0.749	0.786	0.734	0.659	0.756
200	0.858	0.802	0.768	0.752	0.739	0.710
300	0.828	0.802	0.768	0.755	0.714	0.785
400	0.791	0.784	0.817	0.749	0.752	0.790
500	0.751	0.790	0.773	0.756	0.715	0.767
1000	0.699	0.777	0.718	0.686	0.606	0.814
5000	0.597	0.664	0.668	0.599	0.518	0.673
10000	0.696	0.630	0.643	0.599	0.453	0.671

Table A.6 A-D: (cont'd)

Genomic Prediction rrBLUP - Michigan - by SNP Level						
D	AUDPC6	Tar Spot 6	Tar Spot 5	Tar Spot 4	Tar Spot 3	AUDPC5
# Of SNPs	Acc	Acc	Acc	Acc	Acc	Acc
10	0.663	0.509	0.449	0.391	0.379	0.504
20	0.682	0.592	0.513	0.493	0.572	0.667
50	0.691	0.703	0.701	0.634	0.583	0.707
75	0.791	0.732	0.730	0.711	0.613	0.765
100	0.815	0.764	0.774	0.721	0.672	0.742
200	0.766	0.810	0.745	0.766	0.723	0.739
300	0.807	0.799	0.797	0.783	0.730	0.804
400	0.813	0.823	0.780	0.778	0.726	0.789
500	0.766	0.801	0.809	0.737	0.707	0.778
1000	0.753	0.806	0.710	0.712	0.633	0.747
5000	0.668	0.661	0.663	0.563	0.538	0.697
10000	0.684	0.637	0.660	0.612	0.439	0.664

Table A.7: Genomic Prediction of Observed Genotypes in New Environment / SNP Level
Traits were tested at multiple SNP levels and with multiple trait combinations

G.P. of Observed Genotypes in IN by SNP level			
# Of SNPs	TS6 x TS3	TS5 x TS3	AUD6 x AUD
50	0.393	0.319	0.428
75	0.396	0.356	0.449
100	0.396	0.363	0.477
200	0.413	0.393	0.488
300	0.430	0.426	0.512
400	0.432	0.448	0.512
500	0.427	0.447	0.516
1000	0.439	0.468	0.511
2000	0.429	0.466	0.525
3000	0.440	0.460	0.527
4000	0.450	0.466	0.534
5000	0.455	0.476	0.543
7500	0.454	0.477	0.537
10000	0.451	0.479	0.542

Table A.8: Genomic Prediction of Unobserved Genotypes in New Environment / SNP Level
Traits were tested at multiple SNP levels and with multiple trait combinations

G.P. of Unobserved Genotypes in IN by SNP level			
# Of SNPs	TS6 x TS3	TS5 x TS3	AUD6 x AUD
50	0.136	0.014	0.110
75	0.149	0.099	0.148
100	0.140	0.117	0.187
200	0.196	0.092	0.148
300	0.179	0.209	0.190
400	0.242	0.210	0.201
500	0.228	0.229	0.236
1000	0.215	0.236	0.305
2000	0.226	0.212	0.308
3000	0.239	0.159	0.303
4000	0.286	0.159	0.346
5000	0.269	0.167	0.373
7500	0.254	0.134	0.379
10000	0.242	0.149	0.342

Table B.1: Number of Hybrids at each year and environment combination
The variation is high, and the three-year combinations have much less than the 2-year combinations

Year	Allegan	Branch	Cass	Huron	Ingham	Mason	Montcalm	Saginaw	Washtenaw
2011-2012	34	35	35	31	34	31	31	34	0
2012-2013	19	16	15	17	19	17	0	19	0
2012-2014	8	6	5	7	8	7	0	8	0
2013-2014	38	38	37	35	38	35	0	38	25
2013-2015	15	18	18	11	15	11	0	15	10
2014-2015	43	40	40	26	43	26	0	43	40
2014-2016	20	13	13	14	20	14	0	20	13
2015-2016	44	33	33	28	44	28	28	44	33
2015-2017	16	12	12	12	16	12	12	10	12
2016-2017	43	37	37	25	43	25	25	29	37
2016-2018	15	15	15	9	12	9	9	13	0
2017-2018	34	34	34	21	22	21	21	23	0
2017-2019	11	12	12	5	4	5	5	0	0
2018-2019	27	26	26	18	14	18	18	0	0

Table B.2: Number of Hybrids at each year and zone combination.
The variation is high, and the three-year combinations have much less than the 2-year combinations.

Year	Z1	Z2	Z3
2011-2012	35	34	31
2012-2013	15	19	17
2012-2014	5	8	7
2013-2014	25	38	35
2013-2015	10	15	11
2014-2015	40	43	26
2014-2016	13	20	14
2015-2016	33	44	28
2015-2017	12	10	12
2016-2017	37	29	25
2016-2018	15	12	9
2017-2018	34	22	21
2017-2019	12	4	5
2018-2019	26	14	18

BIBLIOGRAPHY

BIBLIOGRAPHY

Abel, C. A., Pollak, l. M., Salhuana, W., Widrlechner, M. P., & Wilson, R. L. (2001). Registration of GEMS-0001 Maize Germplasm Resistant to Leaf Blade, Leaf Sheath, and Collar Feeding by European Corn Borer. Crop Science, 41(5), 1651.

Andorf, CM, Lawrence, CJ, Harper, LC, Schaeffer, ML, Campbell, DA, Sen, TZ. (2010). The Locus Lookup tool at MaizeGDB: identification of genomic regions in maize by integrating sequence information with physical and genetic maps Bioinformatics. 2010 26: 434-436.

Bajet, N. B., Renfro, B. L., & Carrasco, J. M. V. (1994). Control of tar spot of maize and its effect on yield. International Journal of Pest Management, 40(2), 121–125. DOI: 10.1080/09670879409371868

Belo, A., Zheng, P. Luck, S. Shen, B., Meyer, D., Li, B. Tingey, S. and Ragalski, A. (2008) Whole-genome scan detects an allelic variant of *fad2* associated with increased oleic acid levels in maize. Molucular Genetics Genomics 279: 1-10. https://Doi.org/10.1007/s00438-007-0289-y

Bernardo, R. (1993). Estimation of coefficient of co-ancestry using molecular markers in maize. Theoretical and Applied Genetics, 85(8), 1055–1062. DOI: 10.1007/bf00215047

Bernardo, R. (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Science 48: 1649-1664.

Blanche, S.B., and G.O. Myers. (2006) Identifying discriminating locations for cultivar selection in Louisiana. Crop Science 46: 946–949.

Block, A.K., Vaughan, M.M., Schmelz, E.A., et al. (2019). Biosynthesis and function of terpenoid defense compounds in maize (*Zea mays*). Planta 249, 21–30. https://Doi.org/10.1007/s00425-018-2999-2

Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., … Mcmullen, M. D. (2009). The Genetic Architecture of Maize Flowering Time. Science, 325(5941), 714–718. DOI: 10.1126/science.1174276

Cao, S., Loladze, A., Yuan, Y., Wu, Y., Zhang, A., Chen, J., … Zhang, X. (2017). Genome-Wide Analysis of Tar Spot Complex Resistance in Maize Using genotyping-by-sequencing SNPs and Whole-Genome Prediction. The Plant Genome, 10(2). DOI: 10.3835/plantgenome2016.10.0099

Campbell, C. L., and Madden, L. V. (1990). Introduction to Plant Disease Epidemiology. John Wiley & Sons, New York.

Carson, M. L. (1999). Diseases of minor importance or limited occurrence. In: Compendium of Corn Diseases [ed. by White, \D. G.]. St Paul, Minnesota, USA: American Phytopathological Society, 23-25.

Ceballos, H. and Deutsch, J. A. (1992). Inheritance of resistance to tar spot complex in maize. Phytopathology, 82:505-512

CIMMYT. (2003). Maize Diseases: A guide for field identification. 4th Edition. Mexico, D.F., Mexico: International Maize and Wheat Improvement Center,119 pp. http://www.cimmyt.org/english/docs/field_guides/maize/pdf/mzDis_foliar.pdf

Crop Protection Network. (2020). Tar spot disease severity scale. https://cropprotectionnetwork.org/

Dalla Lana, F., Plewa, D.E., Phillippi, E.S., Garzonio, D., Hesterman, R., Kleczewski, N. M., Paul, P.A. (2019). First report of tar spot of Maize (*Zea mays*), caused by *Phyllachora maydis*, in Ohio. Plant Disease, 103 (7): 1780.

Dalló, S.C., Zdziarski, A.D., Woyann, L.G. et al. (2019). Across year and year-by-year GGE biplot analysis to evaluate soybean performance and stability in multi-environment trials. Euphytica 215, 113

DeLacy, I. H., Basford, K. E., Cooper, M., Bull, J. K., and McLaren, C. G. (1996). "Analysis of multi-environment trials — A historical perspective," in Plant Adaptation and Crop Improvement, eds M. Cooper and G. L. Hammer (Wallingford: CAB International), 39–124.

Dean R. & Dixon, W. (1951) "Simplified Statistics for Small Numbers of Observations." Anal. Chem., 1951, 23 (4), 636–638

Ding, J. et al. (2015) Genome-wide association mapping reveals novel sources of resistance to northern corn leaf blight in maize. BMC Plant Biology. 15

Dumble S. (2017). GGEBiplots: GGE Biplots with 'ggplot2'. R package version 0.1.1. https://CRAN.R-project.org/package=GGEBiplots

Endelman JB (2011). "Ridge regression and other kernels for genomic selection with R package rrBLUP." Plant Genome, 4, 250-255.

Friendly M., Fox J., and Chalmers P. (2020). matlib: Matrix Functions for Teaching and Learning Linear Algebra and Multivariate Statistics. R package version 0.9.4. https://CRAN.R-project.org/package=matlib

Gage, J. L., White, M. R., Edwards, J. W., Kaeppler, S., & Leon, N. D. (2018). Selection Signatures Underlying Dramatic Male Inflorescence Transformation During Modern Hybrid Maize Breeding. Genetics, 210(3), 1125–1138. DOI: 10.1534/genetics.118.301487

Gabriel, K.R., (1971) The biplot graphic display of matrices with application to principal component analysis. Biometrika, Volume 58, Issue 3, December 1971, Pages 453–467, https://Doi.org/10.1093/biomet/58.3.453

Gardner, C.A., Krakowsky, M.D., Peters, D.W. (2018). Germplasm Enhancement of Maize (GEM) - 24 Years of Public-Private Sector Collaboration to Increase Maize Genetic Diversity [abstract]. Plant and Animal Genome Conference. Abstract # W527.

Gill S. S. Tuteja N. (2010) Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. Plant Physiol Biochem. 48(12): 909-30.

Groves, C. L., Kleczewski, N. M., Telenko, D. E. P., Chilvers, M. I., and Smith, D. L. (2020) *Phyllachora maydis* ascospore release and germination from overwintered corn residue. Plant Health Program. 21:26-30. https://Doi.org/10.1094/PHP-10-19-0077-RS

Gustafson, T. J., Leon, N., Kaeppler, S. M., & Tracy, W. F. (2018). Genetic Analysis of Sugarcane mosaic virus Resistance in the Wisconsin Diversity Panel of Maize. Crop Science, 58(5), 1853–1865. DOI: 10.2135/cropsci2017.11.0675

Hammerschmidt, R. (1999). PHYTOALEXINS: What Have We Learned After 60 Years?. Annu. Rev. Phytopathol, 306(37). DOI: 10.1146/annurev.phyto.37.1.285

Hansey, C. N., Johnson, J. M., Sekhon, R. S., Kaeppler, S. M., & Leon, N. D. (2011). Genetic Diversity of a Maize Association Population with Restricted Phenology. Crop Science, 51(2), 704–715. DOI: 10.2135/cropsci2010.03.0178

Heslot, N., Jannink, J., Sorrells, M. (2015) Perspectives for Genomic Selection Applications and Research in Plants. Crop Sciences 55:1-12. DOI: 10.2135/cropsci2014.03.0249

Hirsch, C. N., Foerster, J. M., Johnson, J. M., Sekhon, R. S., Muttoni, G., Vaillancourt, B., … Buell, C. R. (2014). Insights into the Maize Pan-Genome and Pan-Transcriptome. The Plant Cell, 26(1), 121–135. DOI: 10.1105/tpc.113.119982

Hock, J, Kranz, J. Renfro, B. (1989) El complejo “mancha de asfalto” de maiz: Su distribuccion geografica, requisites ambientales e importancia econmica en Mexico. Rev. Mex. Feiopatol. 7:129-135

Hock, J. (1991). Requisitos ambientales para el desarrollo del “complejo mancha de asfalto” que ataca al maíz en Mexico. Phytopathology, 81:693.

Hock, J., Dittrich, U., Renfro, B. L., and Kranz, J. (1992). Sequential development of pathogens in the maize tarspot disease complex. Mycopathologia, 117(3), 157–161.

Hock, J., Kranz, J., & Renfro, B. L. (1995). Studies on the epidemiology of the tar spot disease complex of maize in Mexico. Plant Pathology, 44(3), 490–502. DOI: 10.1111/j.1365-3059.1995.tb01671.

Hooker, A. L. (1963) Inheritance of chlorotic-lesion resistance to *Helminthosporium turcicum* in seedling corn. Phytopathology 53:660-662. 12.

Hooker, A. L. (1977) A second major gene locus in corn for chlorotic lesion resistance to *Helminthosporium turcicum*. Crop Science. 17:132-135.

Hooker, A. L., and Kim, S. K. (1973) Monogenic and multigenic resistance to *Helminthosporium turcicum* in corn. Plant Diseases Rep. 57:586-589.

Huang, X., & Han, B. (2014). Natural Variations and Genome-Wide Association Studies in Crop Plants. Annual Review of Plant Biology, 65(1), 531–551. DOI: 10.1146/annual-arplant-050213-035715

Huber, D. Hugh-Jones, M., Rust, M., Sheffield S., Simberloff D., and Taylor, C.R. (2002). Invasive Pest Species: Impacts on Agricultural Production, Natural Resources, and the Environment. Council for Agricultural Science and Technology (CAST). 20. https://www.iatp.org/sites/default/files/Invasive_Pest_Species_Impacts_on_Agricultural_.htm

Jannink, J.-L. (2010) Dynamics of long-term genomic selection. Genet. Select. Evol. 42: 35.

Jeger, M., Viljanen-Rollinson, S. (2001) The use of the area under the disease-progress curve (AUDPC) to assess quantitative disease resistance in crop cultivars. Theoretical Applied Genetics 102, 32–40. https://Doi.org/10.1007/s001220051615

Keppler L. D. Baker C. J. (1989) O2--Initiated Lipid Peroxidation in a Bacteria-Induced Hypersensitive Reaction in Tobacco Cell Suspensions. Phytopathology 79: 555-562.

Kleczewski, N. M., Chilvers, M., Mueller, D. S., Plewa, D., Robertson, A. E., Smith, D. L., and Telenko, D. E. (2019). Corn disease management: Tar spot. Crop protection network CPN 2012-W.

Kuki, M. C. et al. (2018). Genome-wide association study for gray leaf spot resistance in tropical maize core. PLoS ONE 13, 1–13

Kump, K. L. et al. (2011). Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat. Genet. 43, 163–168

Lana, F. D., Plewa, D. E., Phillippi, E. S., Garzonio, D., Hesterman, R., Kleczewski, N. M., & Paul, P. A. (2019). First Report of Tar Spot of Maize (*Zea mays*), Caused by *Phyllachora maydis*, in Ohio. Plant Disease, 103(7), 1780. DOI: 10.1094/pdis-01-19-0070-pdn

Lorenzana, R.E., and R. Bernardo. (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theoretical Applied Genetics 120: 151-161.

Lipka, A. E. et al. (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28, 2397–2399

Listgarten, J., Lippert, C., Kadie, C. M., Davidson, R. I., Eskin, E., & Heckerman, D. (2012). Improved linear mixed models for genome-wide association studies. Nature Methods, 9(6), 525–526. DOI: 10.1038/nmeth.2037

Liu, L-J. (1973). Incidence of tar spot disease of corn in Puerto Rico. J Agric Univ Puerto Rico 42:211–216

Liu, X., Huang, M., Fan, B., Buckler, E. S., & Zhang, Z. (2016). Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLOS Genetics, 12(2). DOI: 10.1371/journal.pgen.1005767

Lorenz, A.J., S. Chao, F.G. Asoro, E.L. Heffner, T. Hayashi, H. Iwata, K.P. Smith, M.E. Sorrells, and J.-L. Jannink. (2011) Genomic selection in plant breeding: knowledge and prospects. Adv. Agron. 110: 77-123. DOI:10.1016/B978-0-12-385531-2.00002-5

Mahuku, G., Chen, J., Shrestha, R., Narro, L. A., Guerrero, K. V. O., Arcos, A. L., & Xu, Y. (2016). Combined linkage and association mapping identifies a major QTL (qRtsc8-1), conferring tar spot complex resistance in maize. Theoretical and Applied Genetics, 129(6), 1217–1229. DOI: 10.1007/s00122-016-2698-y

Malvick, D. K., Plewa, D. E., Lara, D., Kleczewski, N. M., Floyd, C. M., & Arenz, B. E. (2020). First Report of Tar Spot of Corn Caused by *Phyllachora maydis* in Minnesota. Plant Disease: The American Phytopathological Society. DOI: 10.1094/pdis-10-19-2167-pdn

Maublanc, A. (1904). Espéces nouvelles de champignons inferieurs. Bull la Soc Mycol Fr 20(2):72–74

Mazaheri, M., Heckwolf, M., Vaillancourt, B., Gage, J. L., Burdo, B., Heckwolf, S., … Kaeppler, S. M. (2019). Genome-wide association analysis of stalk biomass and anatomical traits in maize. BMC Plant Biology, 19(1). DOI: 10.1186/s12870-019-1653-x

Mccoy, A. G., Romberg, M. K., Zaworski, E. R., Robertson, A. E., Phibbs, A., Hudelson, B. D., … Chilvers, M. I. (2018). First Report of Tar Spot on Corn (*Zea mays*) Caused by *Phyllachora maydis* in Florida, Iowa, Michigan, and Wisconsin. Plant Disease, 102(9), 1851. DOI: 10.1094/pdis-02-18-0271-pdn

Mccoy, A. G., Roth, M. G., Shay, R., Noel, Z. A., Jayawardana, M. A., Longley, R. W., … Chilvers, M. I. (2019). Identification of Fungal Communities Within the Tar Spot Complex of Corn in Michigan via Next-Generation Sequencing. Phytobiomes Journal, 3(3), 235–243. DOI: 10.1094/pbiomes-03-19-0017-r

Mehdy MC. (1994). Active Oxygen Species in Plant Defense against Pathogens. Plant Physiol.105(2):467-472. DOI: 10.1104/pp.105.2.467. PMID: 12232215; PMCID: PMC159383.

Meuwissen, T.H.E., Hyes, B.J., Goddard, M.E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157:1819-1829. PMID: 11290733; PMCID: PMC1461589.

Mottaleb, K. A., Loladze, A., Sonder, K., Kruseman, G., & Vicente, F. S. (2018). Threats of Tar Spot Complex disease of maize in the United States of America and its global consequences. Mitigation and Adaptation Strategies for Global Change, 24(2), 281–300. DOI: 10.1007/s11027-018-9812-1

Mueller, D., Wise, K., and Sisson, A. (2019). Corn disease management: Corn disease loss estimates from the United States and Ontario, Canada – 2018. Crop Protection Network.

Müller, E. and Samuels, J. G. (1984). *Monographella maydis* and its connection to the tar–spot disease of *Zea mays*. Nova Hedwigia 40: 113–121.

Oliveira de, Tâmara Rebecca, Carvalho, Hélio, Oliveira, Gustavo, Costa, Emiliano, Gravina, Geraldo, Santos, Rafael, Carvalho Filho, José Luiz. (2019) Hybrid maize selection through GGE biplot analysis. Bragantia, 78. 10.1590/1678-4499.20170438

Ono E, Wong HL, Kawasaki T, Hasegawa M, Kodama O, Shimamoto K. (2001) Essential role of the small GTPase Rac in disease resistance of rice. Proc Natl Acad Sci U S A. 98(2):759-64. DOI: 10.1073/pnas.021273498. PMID: 11149940; PMCID: PMC14661.

Oyekunle, M., Haruna, A., Badu-Apraku, B., Usman, I.S., Mani, H., Ado, S.G., Olaoye, G., Obeng-Antwi, K., Abdulmalik, R.O. and Ahmed, H.O. (2017), Assessment of Early-Maturing Maize Hybrids and Testing Sites Using GGE Biplot Analysis. Crop Science, 57: 2942-2950. https://Doi.org/10.2135/cropsci2016.12.1014

Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, et al. (2002). Functional SNPs in the lymphotoxin-alpha gene are associated with susceptibility to myocardial infarction. *Nat Genet.*; 32:650–654.

Peng, M., Kuc, J. (1992). Peroxidase-generated hydrogen peroxide as a source of antifungal activity in vitro and on tobacco leaf disks. Phytopathology 82:696-699

Perez, P., de Los Campos, G., (2014) Genome-Wide Regression and Prediction with the BGLR Statistical Package. Genetics 198 (2): 483-495.

Poland, J. A., Bradbury, P. J., Buckler, E. S., & Nelson, R. J. (2011). Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proceedings of the National Academy of Sciences, 108(17), 6893–6898. DOI: 10.1073/pnas.1010894108

Pollak, L. M. (2003). The History and Success of the public-private project on germplasm enhancement of maize (GEM). Advances in Agronomy, 73, 45–87. DOI: 10.1016/s0065-2113(02)78002-4

Précigout, P. A., Claessen, D., Makowski, D., and Robert, C. (2020) Does the latent period of leaf fungal pathogens reflect their trophic type? A meta-analysis of biotrophs, hemibiotrophs, and necrotrophs. Phytopathology 110:345-361. https://Doi.org/10.1094/PHYTO-04-19-0144-R Link, ISI, Google Scholar

Prabhu Dayal Meena, Chirantan Chattopadhyay, Syam Sunder Meena & Arvind Kumar (2011) Area under disease progress curve and apparent infection rate of Alternaria blight disease of Indian mustard (*Brassica juncea*) at different plant age, Archives of Phytopathology and Plant Protection, 44:7, 684-693, DOI: 10.1080/03235400903345281

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38(8), 904–909. DOI: 10.1038/ng1847

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.URL http://www.R-project.org/.

Rife, Trevor, Poland, Jesse. (2014) Field Book: An Open-Source Application for Field Data Collection on Android. Crop Science, 54(4), 1624-1627. DOI: https://Doi.org/10.2135/cropsci2013.08.0579

Ruhl, G., Romberg, M. K., Bissonnette, S., Plewa, D., Creswell, T., & Wise, K. A. (2016). First Report of Tar Spot on Corn Caused by *Phyllachora maydis* in the United States. Plant Disease, 100(7), 1496. DOI: 10.1094/pdis-12-15-1506-pdn

Sakr, N. (2019) In Vitro Quantitative Resistance Components in Wheat Plants to Fusarium Head Blight. The Open Agriculture Journal, 13, 9-18. DOI: 10.2174/1874331501913010009

Schnable, P., Ware, D., Fulton, R., Stein, J., Wei, Fusheng., Pasternak, S., Liang, C., Shang, J., Fulton, L., Graves, T. et al. (2009). The B73 Maize Genome: Complexity, Diversity, and Dynamics. Science. 326:1112-1115. DOI: 10.1126/science.1178534

Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q., & Nordborg, M. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nature Genetics, 44(7), 825–830. DOI: 10.1038/ng.2314

Shaner, G. & Finney, R., (1977). The Effect of Nitrogen Fertilization of Expression of Slow-Mildewing Resistance in Knox Wheat. Phytopathology 67: 1051–1056.

Shi, L. et al. (2014). Genetic characterization and linkage disequilibrium mapping of resistance to gray leaf spot in maize (*Zea mays L.*). Crop J. 2, 132–143

Singh, M. P., Widdicombe, W.D., and Williams L.A. (2018) Michigan Corn Hybrids Compared. Extension Bulletin E-431. https://varietytrials.msu.edu/corn/

Sprague, G. F., and Federer, W. T. (1951). A comparison of variance components in corn yield trials. II. Error, year _ variety, location _ variety, and variety components. Agron J. 43, 535–541. DOI: 10.2134/agronj1951.00021962004300110003x

Swallow, W. H., and Wehner, T. C. (1989). Optimum allocation of plots to years, seasons, locations, and replications, and its application to once-over-harvest cucumber trials. Euphytica 43, 59–68. DOI: 10.1007/bf00037897

Swart V, Crampton BG, Ridenour JB, Bluhm BH, Olivier NA, Meyer JJM, Berger DK. (2017) Complementation of CTB7 in the Maize Pathogen *Cercospora zeina* Overcomes the Lack of In Vitro Cercosporin Production. Mol Plant-Microbe Interact. (9):710-724. DOI: 10.1094/MPMI-03-17-0054-R. PMID: 28535078.

Telenko, D., Chilvers, M., Kleczewski, N., Smith, D., Byrne, A., Devillez, P., … Williams, L. (2019). How Tar Spot of Corn Impacted Hybrid Yields During the 2018 Midwest Epidemic. Crop Protection Network. DOI: 10.31274/cpn-20190729-002

Tenuta, A. (2020) Corn School: Tackling tar spot in Ontario. https://www.realagriculture.com/2020/10/corn-school-tackling-tar-spot-in-ontario

Tian, F., Bradbury, P. J., Brown, P. J., Hung, H., Sun, Q., Flint-Garcia, S., … Buckler, E. S. (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. Nature Genetics, 43(2), 159–162. DOI: 10.1038/ng.746

Thomason, W.E., and S.B. Phillips. (2006) Methods to evaluate wheat cultivar testing environments and improve cultivar selection protocols. Field Crops Res. 99:87–95.

United States Department of Agriculture. (2020) "GEM Protocol (Modified Pedigree Method)." www.ars.usda.gov/midwest-area/ames/plant-introduction-research/home/germplasm-enhancement-of-maize/germplasm-development-gem-lines/gem-line-summary/.

Van Inghelandt, D., Melchinger, A. E., Martinant, J.-P., Stich, B. (2012) Genome-wide association mapping of flowering time and northern corn leaf blight (*Setosphaeria turcica*) resistance in a vast commercial maize germplasm set. BMC Plant Biol. 12

Venables WN, Ripley BD (2002) Modern Applied Statistics with S, Fourth Edition. Springer, New York. ISBN 0-387-95457-0, http://www.stats.ox.ac.uk/pub/MASS4/.

Wang, Q., Tian, F., Pan, Y., Buckler, E. S., & Zhang, Z. (2014). A SUPER Powerful Method for Genome-Wide Association Study. PLoS ONE, 9(9). DOI: 10.1371/journal.pone.0107684

Webb, N.M. R.J. Shavelson, E.H. Haertel. (2006) Reliability coefficients and generalizability theory. Handbook Stat. 26:81-124

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

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Wricke, G., and Weber, E. (1986). Quantitative Genetics and Selection in Plant Breeding. Berlin: Walter de Gruyter.

Wu, Y., Vicente, F. S., Huang, K., Dhliwayo, T., Costich, D. E., Semagn, K., … Babu, R. (2016). Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. Theoretical and Applied Genetics, 129(4), 753–765. DOI: 10.1007/s00122-016-2664-8

Xiao, Y., Liu, H., Wu, L., Warburton, M., & Yan, J. (2017). Genome-wide Association Studies in Maize: Praise and Stargaze. Molecular Plant, 10(3), 359–374. DOI: 10.1016/j.molp.2016.12.008

Yan, W. et al. (2000) "Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot.' Crop Science, 40:597-605

Yan, W., Cornelius, P.L., Crossa, J. and Hunt, L. (2001), Two Types of GGE Biplots for Analyzing Multi-Environment Trial Data. Crop Science, 41: 656-663. DOI: 10.2135

Yan W, Kang M (2003). GGE Biplot Analysis: A Graphical Tool for Breeders, Geneticists, and Agronomists. CRC Press.

Yan W, Tinker NA (2006) Biplot analysis of multi-environment trial data: principles and applications. Can J Plant Sci 86:623–645.

Yan, Weikai, and James B. Holland. (2009) "A Heritability-Adjusted GGE Biplot for Test Environment Evaluation." Euphytica, vol. 171, no. 3, pp. 355–369., DOI:10.1007/s10681-009-0030-5.

Yan, W. (2014). Crop Variety Trials: Data Management and Analysis. Hoboken, NJ: John Wiley & Sons.

Yan, W., Frégeau-Reid, J., Martin, R. et al. (2015). How many test locations and replications are needed in crop variety trials for a target region?. Euphytica 202, 361–372DOI: 10.1007/s10681-014-1253-7

Yan, Weikai. (2021) "Estimation of the Optimal Number of Replicates in Crop Variety Trials." Frontiers in Plant Science, vol. 11, DOI:10.3389/fpls.2020.590762.

Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., … Buckler, E. S. (2005). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics, 38(2), 203–208. DOI: 10.1038/ng1702

Zhou, M., Chihana, A., and Parfitt, R. (2011). "Trends in variance components and optimum replications and crop-years for variety trials at Dwangwa sugar estate in Malawi," in Proceedings of the South African Sugar Technologists' Association, 84, Durban, 363–374.