MICRO AND MACRO VIEWS OF THE MAIZE-*SETOSPHAERIA TURCICA* PATHOSYSTEM

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Tyr Wiesner-Hanks

May 2020

Micro and macro views of the maize-*Setosphaeria turcica* pathosystem

Tyr Wiesner-Hanks, Ph. D.

Cornell University 2020

Interactions between host and pathogen can be understood at many different spatial scales, from nanometers to kilometers. In this dissertation, I explored multiple diseases at two very different spatial scales, focusing chiefly on the economically important disease Northern Leaf Blight (NLB) and the components of its pathosystem- the host, maize, and the fungal pathogen, *Setosphaeria turcica*.

At the “micro” scale, I reviewed the genetics of multiple disease resistance (MDR), what is known about the biological mechanisms thereof, and the ways in which MDR can be improved in plants. I mapped genetic loci conditioning MDR in order to understand the genetic architecture thereof, finding that the high degree of MDR observed in a maize line derived from recurrent selection was mostly attributable to independent loci for resistance to individual diseases, rather than pleiotropic loci conditioning MDR. I used RNA-seq to explore the transcriptomic aspects of infection, with a focus on the pathogen’s transition from biotrophy to necrotrophy and the impacts of pathogen virulence/avirulence in the presence or absence of the host *Ht2* R gene. Gene

expression in both host and pathogen shifted dramatically between biotrophy and necrotrophy, with specific trends demonstrating the different molecular mechanisms of infection and host defense during each phase. Pathogen avirulence, due to R-gene mediated resistance, led to an apparent arrest of the pathogen in the biotrophic phase. The importance of gene-sparse regions of the *S. turcica* genome for pathogenesis was shown for the first time.

At the “macro” scale, I combined crowdsourcing and machine learning to develop a new method for aerial detection of disease symptoms in the field. The task of annotating thousands of disease lesions in order to train a machine learning model was split in two, with experts annotating lesions in low resolution and numerous non-experts performing the more time-consuming task of outlining lesions, using the expert annotations as a base. This method allowed us to generate a large amount of reliable training data very quickly and at low cost. These data were used to train a convolutional neural network (CNN) to high accuracy, and a fully-connected conditional random field (CRF) was used to segment images into lesion and non-lesion areas using the CNN output. The final model was able to delineate lesions in aerial images down to the millimeter level, a finer spatial scale than any previously reported method. It also outperformed human experts by identifying lesions that they had missed. Though the techniques, findings, and impacts involved in work at these two very different scales are accordingly varied, they all contribute to a holistic understanding of the pathosystem and our ability to make practical change.

BIOGRAPHICAL SKETCH

Tyr was raised in Mequon, Wisconsin, where he developed a love for mathematics from an early age and a passion for plants from a slightly older age. He attended Northwestern University and graduated *cum laude* with a B.A. in Biology, concentrating in Plant Biology as part of a fairly new joint program between Northwestern and the Chicago Botanic Garden. While at Northwestern, he conducted undergraduate research with Nyree Zerega on germplasm characterization and population genetics in breadfruit and related *Artocarpus* species.

In the fall of 2013, he joined Rebecca Nelson's lab at Cornell University as a Ph.D. student, studying the maize-*Setosphaeria turcica* pathosystem and related topics in plant and fungal genetics. He spent his summers conducting field research in Aurora, NY and the rest of the year in the greenhouse or growth chamber and in front of the computer. The first half of his Ph.D. was financed by an NSF grant to study maize disease with transcriptomic, mutagenic, and microscopic approaches, while the second half was funded by an NSF National Robotics Grant submitted by himself and others. He was an active member of the Synapsis Graduate Student Organization, serving as treasurer for two years and social chair for two years.

In June 2019, he moved to Rhinelander, WI to work for PepsiCo as an oat and potato discovery scientist, and in November 2019, he became the oat breeder for PepsiCo/Quaker Oats.

Dedicated to my mother, father, and brother.

ACKNOWLEDGMENTS

All of the work herein only possible thanks to the guidance and wisdom of many others. Thanks first to my committee- Rebecca Nelson, Gillian Turgeon, and Michael Gore- for their guidance and mentorship. Many thanks are due to the fellow members of the Nelson Lab, first among them Judy Kolkman, my closest collaborator and comrade during my time at Cornell. I have learned a great deal from my fellow graduate students- Laura Morales, Jenny Cornell, Anthony Wenndt, Danilo Moreta, Will Stafstrom- as well as many of those who worked with Rebecca before my time, including Chia-Lin Chung, Jesse Poland, Tiffany Jamann, and Santiago Mideros. Many thanks go to the lab's non-student personnel, Molly Towne, AC Repka, and Nick Morales.

All field and greenhouse experiments were only possible thanks to knowledgeable and hardworking personnel, namely Chad Thomas at the Dimock Lab and Paul Stachowski at the Musgrave Research Farm. My work on machine learning was only possible thanks to the collective effort of our National Robotics Initiative grant members, including Rebecca and Mike, Gore Lab members Ethan Stewart and Nicholas Kaczmar, and Lipson Lab members Hod Lipson, Chad DeChant, and Harvey Wu. My fellow graduate students in the Synapsis Plant Breeding & Genetics Graduate Student Organization, first among them Lauren Brzozowski, have given me advice, guidance, criticism, and camaraderie, all of which shaped my professional and personal lives for the better. My family has given me love and support from afar.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AUDPC: Area under disease progress curve

BLUP: Best linear unbiased predictor

CNN: Convolutional neural network

DAMP: Damage-associated molecular pattern

DLA: Diseased leaf area

DTA: Days to anthesis

ETI: Effector-triggered immunity

GBS: Genotyping-by-sequencing

GLS: Grey leaf spot

GST: Glutathione S-transferase

GWAS: Genome-wide association study

MAF: Minor allele frequency

MAS: Marker-assisted selection

MDR: Multiple disease resistance

NAM: Nested association mapping

NIL: Near-isogenic line

NLB: Northern leaf blight

NLR: Nucleotide-binding leucine-rich repeat

PAMP: Pathogen-associated molecular pattern

PRR: Pattern-recognition receptor

PTI: PAMP-triggered immunity

QTL: Quantitative trait locus

RIL: Recombinant inbred line

SLB: Southern leaf blight

SNP: Single nucleotide polymorphism

SSR: Simple sequence repeat

CHAPTER 1

MULTIPLE DISEASE RESISTANCE IN PLANTS[1]

ABSTRACT

Many plants, both in nature and in agriculture, are resistant to multiple diseases. Although much of the plant innate immunity system provides highly specific resistance, there is emerging evidence to support the hypothesis that some components of plant defense are relatively nonspecific, providing multiple disease resistance (MDR). Understanding MDR is of fundamental and practical interest to plant biologists, pathologists, and breeders. This review takes stock of the available evidence related to the MDR hypothesis. Questions about MDR are considered primarily through the lens of forward genetics, starting at the organismal level and proceeding to the locus level and, finally, to the gene level. At the organismal level, MDR may be controlled by clusters of R genes that evolve under diversifying selection,by dispersed, pathogen-specific genes, and/or by individual genes providing MDR. Based on the few MDR loci that are well-understood, MDR is conditioned by diverse mechanisms at the locus and gene levels.

INTRODUCTION

Plants must defend themselves against a wide range of pathogens with diverse offensive strategies. This review considers the evidence regarding multiple disease resistance (MDR), with an eye to understanding its importance and mechanisms. Because we approach this evidence with crop improvement and protection in mind, we focus mostly on naturally occurring genetic variation

1 Wiesner-Hanks T, Nelson R. 2016. Multiple Disease Resistance in Plants. Annual Review of Phytopathology 54: 229-252.

affecting resistance, with a secondary interest in the potential of transgenic resistance. Although MDR is a highly desirable plant trait, the underlying genetic architecture and biological mechanisms are not as well understood as those of single-disease resistances. As such, in addition to reviewing the literature surrounding MDR, we also speculate as to what mechanisms may in the future be revealed to mediate MDR.

Defining Multiple Disease Resistance

We use “host plant resistance to two or more diseases” as the definition of MDR (98, p. 203). This definition includes all forms of heritable host plant resistance, including both qualitative resistance and quantitative resistance (see 101 for a review of relevant terminology). It does not necessarily imply shared causal loci; a plant with MDR could carry several distinct R genes effective against different pathogens.

Note that this definition excludes nonhost resistance. Because most plants are nonhosts to most diseases, MDR should be distinguished from the many nonhost resistances of any given plant. MDR is also distinct from multiple disease susceptibility (MDS), the loss of baseline levels of resistance. A loss-of-function mutation in some critical defense gene may lead to extreme susceptibility to many pathogens (i.e., MDS), whereas overexpression of the same gene may (or may not) impart MDR. Although both are biologically interesting, only the latter might be used to improve host plant resistance.

Multiple Disease Resistance: **Why and Why Not?**

There is reason to expect, a priori, that there might be a nonspecific component to disease resis- tance in plants. The plant innate immunity system, broadly speaking, has two main branches: a nonspecific form based on the recognition of pathogen-association molecular patterns (PAMPs) and a highly specific form based on the recognition of pathogen effectors, known as PAMP- triggered immunity (PTI) and effector-triggered immunity (ETI), respectively (29). Nonspecific PTI can result when plant pattern-recognition receptors (PRRs) detect highly conserved PAMPs such as bacterial flagellin or fungal chitin (9, 164). These PRRs are in turn targeted and silenced by pathogen effectors. Genetic variation that affects the sensitivity of plant PRRs, or their silencing by pathogens, would in turn affect resistance to many pathogens from similar taxa. Alternatively, if plants recognize host damage features that occur as a general consequence of pathogenesis [damage-associated molecular patterns (DAMPs)] (36, 102), they may be resistant to pathogens with similar effects on host tissue. Once signals (PAMPs, DAMPs, or effectors) have been recognized, genetic variation could affect the many downstream signaling cascades, in turn affecting the sensitivity, degree, or mechanism of the activated defense response.

Variation in constitutive defenses might also affect multiple resistances. Physical barriers and antimicrobial compounds, which have been implicated in both host and nonhost resistance (66), may hinder pathogens with similar invasion strategies or damage similar pathogen taxa. Other plant traits, like developmental timing or reduced herbivory, may allow the host to evade infection by pathogens with similar life cycles or vectors.

Considering the potential for general defense strategies, one might expect that MDR would be a common phenomenon. Before considering the available evidence, it is worthwhile to reflect on why

MDR might not be common. Three hypotheses come to mind. The first is that although many defense systems can be broadly effective, they are not all constitutively active. Thus, plants must distinguish threats from nonthreats, creating room for specificity. The second is that the evolutionary arms race involves pathogens evolving to thwart plant basal defense mechanisms, which is in turn subject to reciprocal evolution by plants: a cycle of selection described by Jones & Dangl (66). The third is that defense strategies that defeat one set of pathogens may make the plant more vulnerable to another (e.g., cell death can stop biotrophs but facilitate pathogenesis for necrotrophs).

Genetic Scales

MDR can be conditioned by genetic variation at any scale, from many genes across the genome (e.g., of a highly resistant variety) to the single gene (e.g., the *Lr34* gene in wheat). Resistance at the whole-genome scale could be conditioned by multiple unlinked loci that each provide protection against single pathogens (**Figure 1.1a**) or by chromosomal segment(s) that individually provide MDR (**Figure 1.1b,c,d**). Resistance at the level of a chromosomal segment may be conditioned by clusters of tightly linked genes (**Figure 1.1b**) or by individual genes with pleiotropic effects (**Figure 1.1c,d**). There are a number of mechanisms by which a single gene might provide resistance to multiple diseases, several of which have been empirically implicated. The bulk of this review discusses the evidence for MDR at these varying genetic scales.

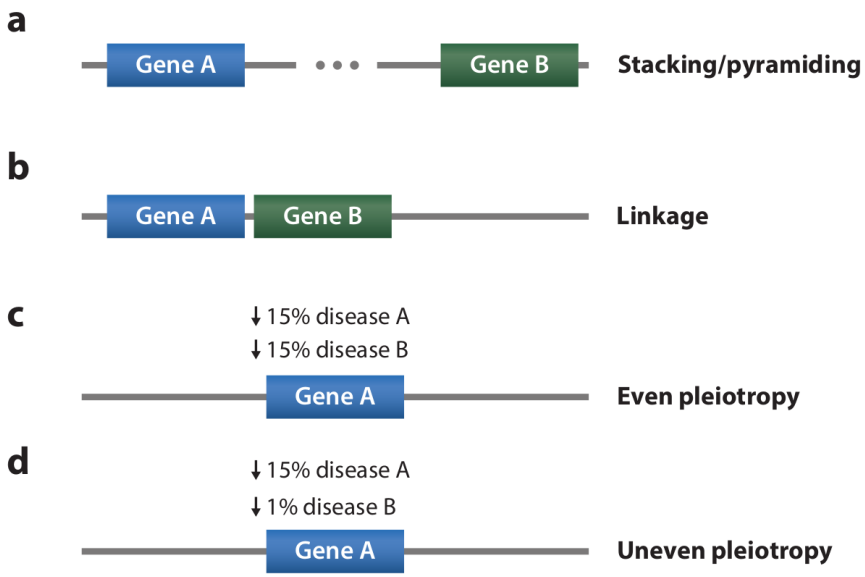


Figure 1.1. Four genetic scenarios by which a plant might be resistant to multiple diseases. (a) Loci conditioning resistance to single diseases (either R genes or quantitative trait loci effective against single pathogen species) may be stacked/pyramided within a single genotype. These loci may be on separate chromosomes or the same chromosome. (b) Loci conditioning resistance to single diseases may be in tight linkage with one another and thus typically transmitted as a unit from one generation to the next. These loci may be tightly or loosely linked, in coupling or in repulsion. (c,d) A single locus may have pleiotropic effects on multiple diseases. It may have roughly comparable effect sizes on both diseases, which we term (c) even pleiotropy, or highly divergent effect sizes, which we term (d) uneven pleiotropy. The key difference between even and uneven pleiotropy is that the former can be detected by genetic mapping methods, whereas the latter may be undetectable due to the small effect size on one of the diseases. The examples given here involve resistance to two diseases, but trade-offs are also possible; in some cases, within any of the above scenarios, resistance to one disease may be associated with susceptibility to another.

MULTIPLE DISEASE RESISTANCE AT THE WHOLE-GENOME SCALE

Given the abundance and diversity of plant pathogens, MDR is of clear evolutionary importance and agricultural interest. Potato, for example, is attacked by least 12 major diseases and pests in Europe (62). Wheat plants are often infected by multiple pathogens at a given time (49). Legumes are vulnerable to a large number of diseases of many taxa, with plants often infected by more than one pathogen (98). Resistance to multiple diseases is therefore of great importance to plant pathologists and breeders. MDR has been noted as a valued trait for more than a century; resistance to multiple diseases was documented in cowpea in 1902 (103, 146).

Multiple resistances are frequently described among the merits of plant genetic resources. To quantify the importance of MDR to modern plant breeding, we searched for relevant citations in the Journal of Plant Registrations (JPR), the official registration publication of the Crop Science Society of America since 2007. A Web of Science title search for "resistan∗ " (which captures resistance, resistant, etc.) in the JPR returned 115 results. Of these, 30 described new varieties, mapping populations, or other germplasm resources with resistance to single diseases, and 70 described germplasm resistant to multiple diseases. MDR is clearly a highly valued trait in plant genetic resources.

Germplasm Screening

Crop improvement programs routinely screen germplasm collections for resistance to multiple diseases (e.g., 11, 48, 49, 56, 107). The prevalence of lines with MDR varies highly from study to study, often even across studies in the same host-pathogen systems. In many instances, resistances to multiple diseases have been found to be correlated across the entirety of a germplasm collection (26, 49, 50, 93,

106, 148). These significant correlations can exist across panels even if individual lines with a high degree of MDR are rare. For example, in a screening of spring wheat landraces for five leaf spot diseases, resistances were correlated for 11 of the 15 pairs of pathogens tested, but only less than 1% of accessions were resistant to three or more diseases (50).

Even with similar diseases in the same species, the frequency with which MDR occurs can vary highly among studies, as in the case of resistance to multiple leaf spot diseases in wheat, which are caused by various fungal pathogens and one bacterial pathogen. In a panel of diverse lines screened for two fungal leaf spot diseases, 11% carried partial resistance to both (48). In another panel of lines screened for three fungal leaf spot diseases, only 2.4% carried partial resistance to all three pathogens (1). As noted above, less than 1% of spring wheat landraces screened for five leaf spot diseases were partially resistant to three or more diseases (50).

Multienvironment Trials

Most of the disease screening studies referenced above were done under controlled conditions at a single location. An alternative approach is to use multienvironment field trials to assess the stability of resistance to multiple pathogen species and genera. In a study on fava bean, 43 accessions were tested for reaction to two diseases (138). Eleven accessions with stable resistance to both diseases were identified and confirmed under controlled conditions. In a multilocation study of resistance to multiple *Fusarium* species, 25 winter wheat genotypes were tested at six locations across Europe with 17 strains of three *Fusarium* species (137). When the wheat genotypes were assessed for their responses to the 59 combinations of strain, year, and location, resistances to different species were highly correlated.

Resistance to Fusarium head blight caused by multiple species was therefore inferred to be species nonspecific.

Screening of Wild Relatives

MDR may be sourced from wild or cultivated crop relatives, sometimes in hopes of identifying genes with broader-spectrum resistance phenotypes than are available within the cultivated gene pool. Wheat-rye translocations have been used in several instances to bring MDR from rye to wheat (55, 162). Screening of wild crop relatives can capture highly effective forms of MDR. Jansky & Rouse (60), for example, identified an interspecific hybrid potato clone with resistance to five diverse diseases. Fetch *et al.* (37) tested accessions of the wild progenitor of barley (*Hordeum spontaneum*) for resistance to six fungal pathogens and found that resistance to most of them was present at high frequencies, with greater frequencies of resistance for populations sourced from more moist (likely disease-conducive) environments.

Accounting for Population Structure

To know whether correlations among resistances to different diseases reflect a common genetic basis, the population structure of the germplasm must be understood. Population structure refers to the patterns of genetic relatedness among populations of the same species. This can be estimated from the geographical origin of different lines or from known pedigrees but is most reliably inferred from molecular marker data. For example, population structure largely explained resis- tance correlations in two maize diversity panels, with tropical lines being generally more resistant than temperate lines (109, 148).

Such patterns may reflect historical selection for multiple resistances. If two pathogens thrive in similar environmental conditions, then plant breeders, including the farmers who selected landraces, will select for lines with resistance to both. Cox *et al.* (26) found higher levels of MDR in *Triticum tauschii* accessions from humid areas and hypothesized that the more severe disease pressure from diverse, humidity-loving pathogens led to more stringent selection for resistance. Among wheat accessions from European and Asian breeding programs, resistance to multiple leaf spot diseases was more strongly influenced by region of origin than whether the accession was a landrace or improved variety (49).

Resistance correlations may vary among subpopulations, driven by either highly resistant or susceptible lines. For example, in papaya, different genetic subgroups were found to have distinctive tendencies to provide resistance to various pathogens (139). Resistance to cassava bacterial blight was found to be correlated with resistance to cassava anthracnose disease in one panel of improved cassava varieties (31) but not in a partially overlapping panel of varieties from the same breeding program (40). In a wheat diversity panel, resistances to two fungal leaf spot diseases were correlated in spring wheat accessions, many of which were highly susceptible to both diseases, but not in winter wheat accessions, which were mostly somewhat resistant (48).

Structured Populations

Correlations in biparental families and other structured populations are more straightforward to interpret than correlations in diverse germplasm. To determine whether MDR at the whole- genome level is due to one or more loci, the typical approach is to make crosses and to evaluate patterns of segregation in the progeny. Resistance imparted by a dominant gene of substantial effect (*e.g.*, an R

gene) segregates in a 3:1 ratio in an F_2 population, whereas multiple genes of small effect give a continuous phenotypic distribution. When multiple resistances are correlated in a biparental family, they can be inferred to be controlled by similar chromosomal segments, which can be mapped through cosegregation with molecular markers. For example, resistances to two bacterial diseases were correlated in segregating populations of sweet corn (107). In a diallel analysis of alfalfa resistance to multiple pathogens, correlations were observed within several biparental populations (53). For one pair of pathogens, correlations were positive in two populations and negative for another, suggesting different architecture of causal loci in different populations. In contrast, if multiple resistances do not cosegregate in a biparental population, it implies that they are mediated by different genes, as was shown to be the case for MDR derived from an elite Australian wheat variety (68).

Correlated resistances have also been seen in recurrent selection programs. In alfalfa, recurrent selection for resistance to root rot caused by a single *Fusarium* species led to improved resistance to three Fusarium wilts (90). In *Brassica rapa*, Mitchell-Olds *et al.* (91) conducted three cycles of selection for resistance to each of three diseases and tested the responses for the one selected and two nonselected diseases. Resistance to both an oomycete and an ascomycete responded strongly to direct selection, and each of these diseases also responded significantly to selection for the other. Resistance to a third disease responded less strongly to direct selection and did not respond to selection for either of the other diseases.

MULTIPLE DISEASE RESISTANCE AT THE LOCUS LEVEL

Genetic mapping, supported by the use of molecular markers, has allowed genes influencing both quantitative and qualitative traits to be associated with particular chromosomal segments [quantitative

trait loci (QTLs); 124]. We use the term QTL mapping to refer to linkage mapping in structured populations (77, 158) and the term genome-wide association study (GWAS) to refer to the identification of trait-associated loci in diversity panels (112, 161).

Quantitative Trait Loci Colocalization

As QTLs for different resistances are mapped on plant genomes, the spatial relationships among them (in the chromosomal context) can be assessed. Colocalization of QTLs for different diseases can provide suggestive evidence for MDR loci. Most QTL mapping studies have focused on a single disease, and we could only find a few studies characterizing QTL colocalization for resistance to multiple diseases. Some examples are provided below, and a case study on maize is presented later.

For a given host and set of pathogens, different resistances might be conditioned by the same loci or by distinct loci, and it is not always easy to distinguish these two scenarios. QTLs for resistance to three *Phytophthora* species, all causing cocoa black pod, were found to greatly overlap (115). In ryegrass (*Lolium* spp.), a forage crop important in Europe and Australia, between one and seven QTLs were identified for each of four diseases in an interspecific mapping population (64). One locus, syntenous with a QTL for MDR in rice (149), conferred resistance to three of the four diseases. Clustering of resistance loci has been observed in several legume species (89, 129) and for loci conditioning rust resistance in wheat (18, 84). In other cases, genome-wide MDR is found to be the result of single-disease resistance QTLs that co-occur, as with the nonoverlapping QTLs for resistance to two *Phytophthora* diseases in pepper (12).

Genome-Wide Association Study Colocalization

Colocalizing GWAS associations can also imply a common genetic basis of multiple resistances. A multivariate GWAS of resistance to three fungal foliar diseases in a maize diversity panel identified a single marker significantly associated with all three diseases, implicating a glutathione S-transferase gene (see Oxidative Stress below) (148). Subsequent GWAS analysis in the maize nested association mapping (NAM) population found additional evidence for MDR loci (see Case Study: Maize below). In a combined GWAS for spring wheat resistance to five leaf spot diseases, 32 loci were significantly associated with the different resistances, but no loci were associated with more than one disease (50).

Meta-Analysis

After resistance QTLs have been mapped in different populations of a given plant species, the overall architecture of MDR can be further clarified by meta-analysis, the rigorous integration of QTLs from studies in multiple mapping populations. A meta-analysis of the many QTL studies on barley disease resistance published since 1992 allowed integration of 166 QTLs from 28 studies (119). From these, 20 meta-QTLs were inferred, eight of which corresponded to MDR loci. Some MDR QTLs were associated with resistance to diverse fungal pathogens with a range of lifestyles (biotrophic, hemibiotrophic, and necrotrophic), whereas others were associated with resistance only to biotrophs.

Several meta-analyses have suggested that MDR loci are relatively common in rice. When 94 disease resistance QTLs from 16 mapping studies were integrated onto the same genetic map, QTLs and known R genes were found to cluster by several measures, although the analysis was limited by the low resolution of the mapping studies (149). A subsequent meta-analysis of 572 rice disease resistance QTLs from 56 mapping studies found evidence for 116 meta-QTLs, 76 of which conferred resistance to more than one disease (71).

Quantitative Trait Loci in Repulsion

Just as tightly linked QTLs can cause positive correlations between resistances, QTLs that are linked in repulsion can cause negative correlations between resistances. For example, mapping of QTLs for resistance to seven diseases in a wheat biparental family revealed a QTL cluster on chromosome 3DL (165). Resistance QTLs for yellow leaf spot and Septoria tritici blotch, inherited from one parent, were linked in repulsion to resistance QTLs for leaf rust and stem rust, which were inherited from the other. Similarly, the wheat *Sr2* locus, which confers resistance to stem rust, powdery mildew, and leaf rust (84), was tightly linked in repulsion to the *Fhb1* locus, which confers resistance to Fusarium head blight (38). Understanding the genetic architecture of resistance to multiple diseases gives insight into the potential challenges and opportunities for improving MDR.

Intermediate Model: **Uneven Pleiotropy**

We have discussed scenarios in which a genetic locus confers resistance to either a single disease or to multiple diseases. There is evidence for an intermediate model as well, in which a locus has differential effects on different pathogens, which we term uneven pleiotropy (**Figure 1.1d**). When QTLs were mapped in pepper for resistance to two *Colletotrichum* species, the major QTLs for resistance to the two species did not colocalize, but the major QTL for each colocalized with a minor QTL for the other (79). This could also explain instances in which QTL mapping for multiple resistances finds few to no loci with pleiotropic effects, even though those resistances are correlated, as suggested by Balint-Kurti *et al.* (6).

R-Gene Clusters

Results from several systems have shown that a given MDR locus may be either simple (when a single gene underlies an MDR locus) or complex (when multiple genes underlie the locus). Finemapping studies for dozens of pathosystems have revealed that complexes of homologous R genes often underlie resistance loci (54), including MDR loci. A well-studied example is the lettuce Resistance Gene Candidate 2 (*RGC2*) cluster, which contains several dozen R-gene homologs (88). The gene complement varies across lettuce accessions (74), and most of the genes confer resistance to the oomycete *Bremia lactucae*, but one gene provides resistance to the root aphid (152). Given the size and diversity of this cluster, it is possible that other, as-yet-undiscovered specific resistances are encoded by other *RGC2* genes. Given the prevalence of R-gene clusters in plant genomes and some examples of R-gene clusters at MDR loci, we expect that more MDR loci will eventually be explained by clusters of tightly linked R genes.

R genes within a cluster can provide resistance for distantly related pathogen taxa because variations in R-gene sequences among paralogs or even homologs can result in novel specificities. For example, the allelic *Arabidopsis* genes *HRT* and *RPP8* confer resistance to Turnip crinkle virus and the oomycete *Peronospora parasitica*, respectively (25). The potato cyst nematode resistance gene *Gpa2* and Potato virus X resistance gene *Rx1* are two members of a four-gene cluster with 88% sequence similarity (136). This is likely not a unique phenomenon in potato, as most resistance gene homologs and known R genes cluster tightly on the potato genome (5). The *multiple resistance complex J* (*MRC-J*) cluster in *Arabidopsis* is fascinating both for the diversity of R-gene targets and the genetic arrangement of the R genes it contains. Genes imparting race-specific resistance to several bacteria, viruses, a fungus, and an oomycete have all been characterized within this cluster, which also contains

roughly a dozen or so uncharacterized resistance gene homologs (25, 43, 97, 127). Two R genes arranged head-to-head in the cluster, *RRS1* and *RPS4*, were found to both be necessary for three of these resistances (96).

GENE-LEVEL INSIGHTS

Relatively few disease resistance QTLs have been cloned to date. Among these, several have been shown to have conferred MDR, including *Lr34* and *Lr67* in wheat (72, 95), *GH3-2* in rice (42), and *mlo* in barley (61). Below, we review the literature on cloned MDR loci, as well as speculate on what mechanisms might underlie as-yet-undiscovered cases of single-gene MDR.

Recognition of Conserved Signals

Disruption of recognition pathways can lead to the loss of multiple resistances (51, 52, 67), and successful pathogens have evolved ways to rapidly suppress the basal resistance provided by PAMP-triggered immunity (66). Conversely, variation in plant genes can perhaps produce MDR by altering the perception of certain PAMPs by plant PRRs or by affecting the inhibition of these PRRs by pathogen effectors. Another potential mechanism for single-gene MDR is the recognition by single plant R genes of effectors from multiple pathogens. Several instances of dual-specificity R genes have been noted within single pathosystems; for example, the *Arabidopsis RPM1* gene confers resistance to strains of *Pseudomonas* carrying either of two *avr* genes (46), and the tomato *Pto* gene recognizes *Pseudomonas* strains with either of two dissimilar effectors (69).

For R genes to recognize effectors from distinct pathogens, effector motifs would have to be conserved across species. This is not unlikely; comparison of predicted effector proteins from a wide

range of fungi and oomycetes found that effector sequences were often fairly conserved within clades, even for species of differing lifestyles (73). Groups of effectors from *Pseudomonas* and *Ralstonia* induced similar patterns of necrosis on diverse panels of tomato, pepper, and lettuce lines, even when the effector sequences were fairly divergent, suggesting similar targets in the hosts (151). Effector targets can be conserved among even extremely distant taxa. A yeast two-hybrid screening of effectors from *Pseudomonas syringae* and the oomycete *Hyaloperonospora arabidopsidis* found that effectors from these two distant species targeted several common *Arabidopsis* proteins (94). Host variation in effector targets might thus allow a plant to elude multiple pathogens. We further hypothesize that MDR may result, in some instances, from multispecificity R genes.

Hormone Signaling

Once threats have been recognized, plants rely on shared signaling pathways to initiate defense responses; recent reviews have described the roles of salicylate (140), jasmonate (17, 118), ethylene (15), abscisic acid (28), and the crosstalk between these pathways (13, 70, 116) in response to biotic and abiotic stress. Loci affecting hormone pathways have been suggested to underlie known resistance QTLs. Natural variation in the potato *aos2* gene, which affects jasmonic acid production, is suspected to underlie resistance QTLs against *Phytophthora infestans* and *Erwinia carotovora* (105). The rice *GH3-2* locus, which mediates resistance to *Xanthomonas oryzae* and *Magnaporthe grisea*, was found to encode a synthetase that produces the main form of auxin in rice (42). Less-direct evidence has also connected QTLs for MDR to loci controlling hormone signaling. Mapping of predicted defense genes in maize suggested several homologs of the rice Myb transcription factor, implicated in regulation of the jasmonic acid pathway, as candidate genes underlying QTLs for MDR (145).

Sugar Signaling and Partitioning

A successful pathogen is able to draw nutrients from the host plant, whereas a successful plant is able to deny the pathogen these nutrients. Changing concentrations or ratios of sugars in plant tissue can induce plant defense genes, influence plant hormone pathways, and induce resistance to various diseases (10). Certain genes underlying sugar signaling and transport have been shown to have roles in MDR.

Feeding sucrose to rice plants through the roots was shown to induce expression of defenseresponse genes in a manner highly similar to a challenge by *Magnaporthe oryzae* (44). This relationship seems to apply to multiple diseases. Rice plants transformed with the maize *PRms* gene accumulated higher levels of sucrose in the leaves and showed increased resistance to infection by several fungi and a bacterial pathogen. Rice lines with constitutive overexpression of the cell wall invertase gene *GIF1* accumulated more apoplastic hexoses and sucrose, leading to constitutively activated defense genes and elevated resistance to several diseases (125). The resistance allele of the wheat *Lr67* gene (shown to underlie the loci *Pm46*, *Sr55*, *Yr46*, and *Ltn3*), which confers resistance to leaf rust, stripe rust, stem rust, and powdery mildew, encodes a hexose transporter that inhibits hexose uptake from the apoplast by host cells (94). This inflation of the apoplastic hexose:sucrose ratio is thought to be associated with sugar signaling of pathogen invasion. Similar sugar transporters (STP genes) in *Arabidopsis* have been shown to play a role in basal defense (80) and to be upregulated in response to pathogen infection (41).

As with any disease resistance mechanism, the benefits of altering sugar transport must be considered in light of potential trade-offs for other traits. Constitutive expression of defense response

genes may hinder plant growth, and alternative modes of sugar partitioning may alter critical plant traits such as yield, staygreen, etc. Resistance is ultimately useful to the farmer when it results in improved productivity, nutrition, stability, or quality.

Cell Death and the Hypersensitive Response

Plant cell death is an important defense mechanism against biotrophic pathogens and a gateway to infection for necrotrophic pathogens. Several genes implicated in cell death have been shown to contribute to MDR. The most famous example of this is the recessive *mlo* gene in barley, which provides resistance to several biotrophic pathogens (19). *MLO* is a negative regulator of the defense response, and genotypes that are homozygous for weak or null alleles at this locus manifest an overactive defense response. The gene was originally implicated in resistance to *Blumeria graminis* ssp. *hordei* and has been implicated in response to leaf wounding, leaf senescence, herbicide treatment, and a challenge with rice blast pathogen *M. oryzae* (108). It has also been implicated as a susceptibility factor for multiple necrotrophic pathogens. Plants homozygous for mlo show leaf-tip necrosis, reflecting the overactive induction of cell death. A similar but more exaggerated phenomenon is seen with the *accelerated-cell-death 6* (*ACD6*) gene in *Arabidopsis*, which conditions resistance to diverse pathogens as well as to herbivory (134). This gene is associated with necrosis and plant stunting, highlighting the fitness trade-offs that may be associated with defense strategies.

Lesion mimics are mutant plants that spontaneously develop lesions resembling a hypersensitive response or lesions caused by pathogens (99). These mutants offer an excellent system in which to study cell death and its effects on MDR. In rice, lesion mimic lines have been repeatedly shown to have increased resistance to *M. oryzae* and *Xanthomonas oryzae* pv. *oryzae* (92, 153, 154).

Interestingly, the mutations that impart dual resistance, or any resistance at all, are only a small subset of the characterized lesion mimic mutations in rice. Lesion mimic mutants also vary in terms of their physiological trade-offs. *Arabidopsis* lines heterozygous for two distinct lesion mimic mutations were healthier than homozygous mutants but still retained an elevated hypersensitive response when challenged with virulent *Pseudomonas* isolates (47).

Although cell death can be used to stop pathogens that require living tissue, it can facilitate infection by those that feed on dead tissue. In addition to being more resistant to powdery mildew, *mlo* barley is more susceptible to the necrotrophic *Ramularia collo-cygni* and to laboratory infection by the biotrophic *M. oryzae* (61, 87). Similarly, a barley lesion mimic mutation conferring high levels of resistance to the biotrophic fungus *Puccinia hordei* also conferred hypersensitivity to the necrotrophic oomycete *Pyrenophora teres* f. sp. *teres* (150). Variation in cell death can also work in the opposite direction, with deficiencies in hypersensitive response conferring resistance to necrotrophs but susceptibility to biotrophs. *Arabidopsis* mutants deficient in hypersensitive response, although more susceptible to *P. syringae*, were much more resistant to the necrotrophic fungi *Botrytis cinerea* and *Sclerotinia sclerotiorum* (45).

Oxidative and Chemical Stress

Upon challenge by a pathogen, plants begin to form a wide array of reactive oxygen species (ROS) in a process known as the oxidative burst (135). These ROS strengthen plant cell walls, serve as a signal to induce the disease defense response, and potentially create a hostile environment for invading pathogens. Necrotrophic pathogens, in turn, can elicit host production of ROS and secrete toxic compounds to kill host tissue (45). Mitigating this chemical and oxidative stress is critical for

maintaining plant health. Plants produce enzymes that detoxify ROS in response to all manner of abiotic and biotic stresses (3). There is evidence that genes affecting these processes of generating and mitigating toxic compounds can underlie MDR.

In addition to managing exogenous chemicals, plants must manage the endogenous ROS they produce in response to pathogen attack. Multivariate analysis of resistance to three fungal foliar diseases in a maize diversity panel implicated a glutathione S-transferase (GST) gene (148). As GSTs have been noted for their roles in mitigating oxidative stress and detoxifying xenobiotic compounds (86), this is a plausible candidate for resistance to multiple, mostly necrotrophic, fungal diseases.

As indicated above, the wheat *Lr34* gene provides resistance to several diseases of wheat (72, 76, 123). It has been effective and in wide use for more than 100 years. The *LR34*/*Yr18*/*Pm38* locus was assumed to be a complex of multiple genes, but when it was cloned in 2009, MDR was found to be conferred by a single gene encoding an ATP-binding cassette (ABC) transporter (72). Although neither the substrate of the ABC transporter nor the mechanism by which it provides resistance is known, it is suspected to have a role in transporting or sequestering xenobiotic compounds.

Oxalate oxidases, which catalyze the production of hydrogen peroxide from oxalate, have been hypothesized to underlie MDR loci in rice (114) and wheat (35). Rice has four tandemly duplicated oxalate oxidase genes that have been suggested to underlie a QTL for resistance to rice blast and bacterial blight in rice (114). However, overexpression lines of these four oxalate oxidase genes were not more resistant to either *Xanthomonas oryzae* pv. *oryzae* or *M. oryzae* (160). Still, it is possible that natural variation in oxalate oxidases or similar proteins could affect MDR, as a cluster of 12 germin-like proteins (formerly known as oxalate oxidase-like proteins) was shown to underlie a major rice blast QTL and also to contribute to sheath blight resistance (85).

Antimicrobial Peptides

Plants, animals, bacteria, and fungi produce peptides with broad antimicrobial activity (157). Antimicrobial peptides vary greatly in structure, targets, and efficacy. Plant defensins, for example, are small, cysteine-rich antimicrobial peptides that have long been recognized as an ancient, basal component of plant defense against diseases (131). These compounds, found in high concentration in plant seeds and cell walls, have been shown to inhibit a diverse array of fungi and oomycetes, as well as a more limited number of bacteria *in vitro* (22, 100).

Although the use of antimicrobial peptides as transgenic sources of resistance has been well documented, there is more limited evidence that they may underlie resistance QTLs. The pea Pi39 defensin has been implicated in QTLs conferring resistance to *Fusarium solani* f. sp. *pisi* (27) and Ascochyta blight, caused by three distinct species of fungi (133). Given the abundance, diversity, and broad-spectrum activity of antimicrobial peptides, it is likely that variation in other genes encoding antimicrobial peptides affects MDR as well.

CASE STUDY: **MAIZE**

A series of studies, beginning with forward genetics, was conducted to understand the genetic architecture of disease resistance in maize. In 2006, Wisser *et al.* (147) summarized the available evidence from 50 QTL mapping studies on maize diseases. Although QTLs and R genes for multiple diseases showed significant clustering, finding candidate MDR loci was limited by the resolution of mapping studies at the time. Subsequent mapping studies for resistance to three fungal, predominantly necrotrophic foliar diseases—southern leaf blight (SLB), northern leaf blight (NLB), and gray leaf spot

(GLS)—have described the genetic architecture of these three resistances at increasingly high resolution.

Populations and Correlations

Many maize populations have been screened for these three diseases. In diversity panels, these resistances have been highly correlated. Roughly 250 lines of the Goodman diversity panel (39), representative of the genetic diversity of maize, were screened for these three diseases (148). Resistances were correlated (r = 0.55 to 0.67) even after adjusting for population structure, kinship, and maturity effects. A subset of lines from this panel was used as parents to construct the maize NAM population. The NAM consists of 25 biparental families generated by crossing 25 diverse founders to a common parent, B73 (156). Across the entire population of ∼5,000 lines, resistances were also correlated (r = 0.42 to 0.59), in large part because of population structure rather than segregation within families (109). In both the Goodman panel and the NAM, tropical lines tended were more broadly resistant than temperate lines (109, 148). This is likely because tropical environments tend to have more disease pressure, historically necessitating stronger selection.

Structured Populations

These resistances were found to be less correlated in structured mapping populations. Within the NAM, correlations were much weaker within each of the biparental families (r = −0.07 to 0.41) than among their diverse parental lines (r = 0.62 to 0.77) (109). These three resistances are also loosely correlated in other structured families: the intermated B73 × Mo17 (IBM) biparental population (r = 0.16 to 0.42) (6), a biparental population derived from an MDR line (Ki14) and B73 (r = 0.25 to 0.62) (166), and a

set of near-isogenic lines (NILs) derived from an SLB-resistant line (NC250P) and B73 (r = 0.09 to 0.38) (8). Altogether, this suggests that MDR is mostly mediated by multiple unlinked QTLs conferring resistance to single diseases, rather than mostly by pleiotropic QTLs.

Quantitative Trait Loci Colocalization

In each of the studies mentioned above, resistance QTLs for SLB, NLB, and GLS were mapped, giving deeper insight into the genetic architecture of MDR. Resistance QTLs rarely colocalized in any population. In the IBM biparental population, in only one instance did QTLs for two diseases colocalize (6). In the Ki14 × B73 biparental population, a greater degree of MDR was detected (166). Seventeen resistance QTLs were identified, five of which conditioned resistance to two or more diseases. One locus (in the 1.06 bin of the maize genome, discussed below) conditioned resistance to all three diseases and was also associated with an effect on flowering time.

The NAM population was designed to permit both QTL mapping and GWASs. Joint linkage mapping (i.e., locating QTLs in one or more biparental population) identified 32 QTLs for SLB resistance, 29 for NLB resistance, and 16 for GLS resistance (7, 75, 110). Of these, five colocalized between NLB and SLB, six between GLS and NLB, and one between GLS and SLB resistance (although the two QTLs had opposite effects) (7). The results of QTL meta-analysis tell a similar story: tightly linked and pleiotropic resistance QTLs are either rare or difficult to find. There are several plausible explanations for this (see Missing Loci? Below).

Genome-Wide Association Study Colocalization

Significant associations identified by GWASs can also be used to dissect these correlated resistances at greater resolution. An initial multivariate GWAS for these three diseases in ~250 lines of the Goodman diversity panel found only a single polymorphism associated with all three diseases, despite high genetic correlations between the traits (148). This early study was based on a relatively small number of single-nucleotide polymorphism (SNP) markers (n = 858), likely limiting its resolution.

GWASs have also been conducted in the NAM for SLB resistance (75), NLB resistance (110), and GLS resistance (7).We compared the associations found in these studies in much the same way as the QTLs above. These studies should have more power to detect causal MDR loci, given the larger number of lines (n = 5,000) and SNPs (n = 1.6 million) utilized. Because of the limited LD breakdown in the NAM population (P. Bradbury, personal communication), we considered SNPs to colocalize if they were within 1 megabase of each other. This was supported empirically; at more stringent cutoffs for physical proximity, association results for the same disease with slightly different mapping approaches showed almost no colocalization.

GWAS co-localization has demonstrated several aspects of MDR in maize. Associations from GWASs were integrated into the same physical map and then combined if they were within 1 megabase. After this, there were 121 regions significantly associated with NLB resistance, 115 with SLB resistance, and 99 with GLS resistance. Of these, 21 chromosomal regions colocalized between NLB and SLB, 25 between NLB and GLS, 14 between GLS and SLB, and 4 between all three diseases. From another multitrait GWAS in the NAM, which included the same NLB and SLB phenotypic data as above (141), there were 44 regions associated with NLB resistance and 68 with SLB resistance, only 6 of which overlapped. That standardized effect estimates of co-localizing associations were positively correlated (r = 0.36 to r = 0.47) suggests that these colocalizing associations are truly pleiotropic or

linked loci, rather than spurious coincidences. This supports the hypothesis that pleiotropic or linked loci are rare.

Missing Loci?

Why are loci with pleiotropic disease effects hard to identify, even with similar diseases and correlated resistances? Several phenomena, or a combination thereof, could explain this. It may be that most pleiotropic loci have effects that are too small to be detected by mapping, or that loci with large effects on resistance to one disease may have weaker effects on resistance to another (uneven pleiotropy) (**Figure 1.1d**), as suggested by Balint-Kurti et al. (6). Uneven pleiotropy would be quite difficult to detect in the case of quantitative disease resistance, as most loci conditioning quantitative resistance are expected to have small effects, but loci with negligibly small effects will not pass the significance threshold during QTL mapping; thus, the minor effect of a locus with uneven pleiotropy would be difficult to detect. It may simply be that QTLs for single-disease resistances, selected in environments with high disease pressure, co-occur in lines with a high degree of MDR (gene pyramiding) (**Figure 1.1a**).

Dissection of Quantitative Trait Loci

Although most large-effect disease QTLs in maize appear to be disease-specific, genetic dissection has revealed several QTLs that provide resistance to multiple diseases. A series of studies has been conducted to dissect putative MDR loci using NILs. A study typically identifies a pair of lines differing for a locus that conditions resistance to a disease of primary interest and then analyzes the response of the lines to other diseases. Below, we review a series of such studies conducted on NILs derived from a

cross between B73 and Tx303 that collectively illustrate the complex genetics and diverse mechanisms that can underlie MDR (24, 59).

Two QTLs, located on the same chromosome but in different bins (genome sections roughly 20 cM in length), were introgressed from the broadly resistant line Tx303 into the more susceptible B73 line (24). One (in bin 1.06) conferred resistance to NLB and Stewart's wilt, whereas the other (in bin 1.02) conferred resistance to NLB, Stewart's wilt, and common rust. Both NLB and Stewart's wilt are vascular diseases, with the former caused by a fungus and the latter by a bacterium.

The QTL in bin 1.06, which was shown to hinder leaf penetration by the NLB pathogen *Setosphaeria turcica* (24), proved recalcitrant to conventional genetic analysis, as recombination rates were vanishingly low (59). Whether this QTL is truly pleiotropic (**Figure 1.1c**) or a cluster of linked loci affecting separate diseases (**Figure 1.1b**) remains to be seen. This region is considered an important adaptive region in maize because of the many other QTLs for important traits that have been found there, including an MDR QTL derived from another population (166). Although limited recombination hindered the fine-mapping of this QTL, a mutant for a leucine-rich repeat receptor-like kinase (LRR-RLK) gene, *Pan1*, in the region showed resistance to both NLB and Stewart's wilt (59). This suggests that the wild-type allele is useful in pathogenesis for both vascular pathogens. *Pan1* had previously been shown to play a role in cytoskeletal dynamics (actin organization) required for proper stomatal development (21).

The QTL in bin 1.02, which was shown to restrict entry of *S. turcica* into the vascular tissue (24), could be dissected through recombination. This showed that although the Stewart's wilt and common rust resistances may be due to pleiotropy, the NLB locus is a tightly linked, separate locus (58) (**Figure 1.1b**). Multiple cycles of recombination allowed the narrowing down to four genes, which

were further assessed using mutants, implicating a remorin gene. Members of the remorin gene family, involved in membrane rafts and plasmodesmatal function, have been implicated in inhibiting Potato virus X mobility through plasmodesmata (113) and in promoting infection by *P. infestans* (14). Thus, a nonspecific role in restricting plasmodesmatal movement by different pathogens is plausible.

TRANSGENIC MULTIPLE DISEASE RESISTANCE

With direct transfer of genes among genotypes and species, the potential scope of resistance sources expands to different species, genera, and kingdoms. The extent to which this opportunity will be realized to produce MDR depends on whether known MDR genes are effective in heterologous systems, the technical ease of transformation, and the social and political context that influences the regulatory environment. Although the GMO debate has been politically fraught, the use of resistance genes from wild species could lead to reduced reliance on pesticides and thus could have some appeal from an environmental perspective. Here, we review the existing literature on transgenic methods and their implications for MDR.

Constitutive Defense and its Drawbacks

Most defense mechanisms require the plant to recognize threats and to activate defense response pathways. Constitutive expression of defense response genes can enable plants to bypass the recognition step, creating a hostile environment for invaders. There are many routes to affecting resistance via constitutively expressed defense genes. R-gene overexpression can lead to activation of defense response pathways and subsequent MDR, as was the case with the *Pto* gene in *Arabidopsis* (128). Defense-response genes on the downstream end can also be used. Overexpression of the rice

peroxidase *OsPrx114*, which catalyzes oxidation of substrates by peroxides during the oxidative burst, imparted resistance to multiple necrotrophs in carrot (143).

The *Arabidopsis NPR1* gene, which regulates systemic acquired resistance and has homologs in other crops, has been suggested as a source MDR (32). Overexpression of *NPR1* has led to MDR in *Arabidopsis* (20, 155), tomato (83), carrot (142), and strawberry (120). Interestingly, the baseline expression of pathogenesis-related (PR) proteins in these *NPR1* overexpression lines was not different from the wild-type expression level in Arabidopsis or carrot but was higher in tomato and strawberry. Because *NPR1* activates defense response pathways in a dosage-dependent manner (32), it seems that constitutive expression of this upstream defense regulator can induce MDR in multiple ways: by constitutively activating defense pathways (an undesirable trait) or by increasing the sensitivity, intensity, or duration of the defense response (a potentially desirable trait).

There are good reasons why plants do not typically show constitutive activation of defense pathways, however. Many defense responses are expected to incur a cost to the host plant in some way, by either consuming limited resources or indirectly affecting growth (16, 122, 144). Evidence for fitness costs of transgenic MDR supports this. Overexpression of the *RPM1* R gene in *Arabidopsis* led to stunted plants with lower seed production (134). Although *Arabidopsis NPR1* overexpression lines are developmentally normal, *NPR1* overexpression led to reduced growth in strawberry (120). *Arabidopsis* mutants that overexpressed the MAP kinase kinase 7 gene, previously shown to regulate basal and systemic acquired resistance, had high levels of MDR, but this was associated with a bushy, dwarf plant morphology (159). Constitutive expression of defense-related genes will ultimately be useful from a crop improvement standpoint if the transgene increases sensitivity only to signals from relevant threats, without sacrificing plant performance in the absence of pathogens.

Antimicrobial Peptides

As with other disease defense mechanisms, plants that constitutively express antimicrobial compounds may be able to bypass the recognition of effectors. Transgenic constitutive overexpression of the potato Snakin-1 peptide, originally found to have antibacterial function in vitro, imparted resistance to the bacteria *Rhizoctonia solani* and *E. carotovora* in potato (2). Plant defensins have also been transferred from one plant species to another with great success in many cases. Both tobacco and peanut plants transformed with the mustard defensin gene *BjD* are highly resistant to multiple distinct fungal diseases (126). The radish defensin gene *Rs-AFP2* has been transformed into tobacco (130), tomato (23), pear (78), wheat (82), and rice (63), providing *in vitro* resistance to a wide range of economically important fungal pathogens. Transformation of potato with the *Nicotiana megalosiphon*–derived peptide *NmDef02* conferred resistance to Alternaria solani and P. infestans (111). Using plant defensins in transgenic plants is still a fairly conservative strategy, as the defensins are less likely to have broad phytotoxic properties.

Of course, transgene sources are not limited to the plant kingdom. Transforming plants with antimicrobial peptides from distantly related taxa arms them with chemical weapons that may be quite novel to potential pathogens. Synthetic analogs of magainin, an antimicrobial peptide from the African clawed frog, have been used to confer MDR in transgenic tobacco (22, 81) and banana (22). The *msrA1* gene, which encodes for a chimeric protein derived from antimicrobial peptides of the giant silk moth and bee venom, imparted resistance to fungal and bacterial pathogens in potato (104) and to fungal pathogens in *Brassica juncea* (117). As many of these authors have noted, the broad antimicrobial activity of these peptides comes with a caveat: Novel transgenic peptides may have

unforeseen impacts on the plant, its pollinators, the beneficial microbes of the rhizosphere, and/or its human consumers.

IMPROVING MULTIPLE DISEASE RESISTANCE

As we learn more about the underlying genetic and biological mechanisms behind MDR, how do we translate this knowledge into improving MDR in different crops? Methods for producing crop varieties with whole-genome MDR vary, with the strategy depending on the nature of inheritance of MDR. To make efficient use of a resistance source, it is necessary to have a clear understanding of its inheritance. For example, the breeding strategy required for utilizing monogenic resistance is distinct from one based on quantitative resistance. It would be relatively easy to manage a single gene that provided resistance to multiple diseases, and a much greater technical challenge to make practical use of a large number of loci of small effect.

A common strategy for producing MDR at the whole-genome level is to combine multiple major genes into a single line. The pyramiding or stacking of R genes or other major genes can be achieved by phenotypic selection and/or marker-assisted selection (MAS). For example, Singh *et al.* (121) and Zhou *et al.* (163) used MAS to pyramid R genes for rice blast and/or bacterial diseases in Basmati rice. Eibach *et al.* (33) used MAS to pyramid resistance QTLs for downy mildew and powdery mildew in grape, avoiding the long generation times needed to conduct phenotypic selection.

Efforts to combine major genes are influenced by the distribution and nature of those genes within the species. If loci for different resistances are not closely linked, they must be introgressed independently of one another. If they are linked, they are easy to introgress if linked in coupling and more difficult to introgress if they are linked in repulsion. For example, the wheat MDR locus *Sr2* was

difficult to combine with the Fusarium head blight resistance locus *Fhb1* because the two were linked in repulsion (38). Screening by molecular markers eventually identified recombination events between the two loci; the resultant recombinant lines can be used as donors of both the *Fhb1* and *Sr2* loci with limited loss of stacked resistance due to further recombination.

Although molecular markers can be useful, phenotypic selection can also be used to combine resistances. Teran *et al.* (129) describe a strategy for developing breeding lines with resistance ´ to the five most important diseases of common bean in Latin America by screening multiparent populations with multiple pathogens. They found relatively high co-incidence of resistance to three fungal diseases and a viral disease. The authors note that their success in creating MDR lines was probably enhanced by the clustering of resistance genes in the bean genome. Elgin *et al.* (34) compared a range of selection methods for developing alfalfa populations with resistance to multiple diseases. They found that sequential selection for one disease at a time was ineffective (resistance to a given disease would respond to selection but would be lost when the population was later selected for resistance to other diseases), suggesting that resistances were genetically unlinked.

The reliance on major genes can be appealing because of their potential to provide complete resistance and the relative ease with which they can be analyzed. The main downside to their exploitation is their potential lack of durability, or their long-term performance in the face of pathogen evolution (65). Polygenic, quantitative resistance is considered to be the most durable form of resistance (124). Single R genes are generally race specific and relatively rapidly overcome as pathogen populations evolve under selection pressure. Broad-spectrum resistance is logically more likely to be durable than resistance for which compatible pathogen strains are already known. MDR can be regarded as an exceptional form of broad-spectrum resistance and thus potentially particularly

difficult to overcome. Consistent with this, single genes associated with MDR, such as *mlo* and *Lr34*, are among the rare major resistance loci to have demonstrated durability.

CONCLUDING REMARKS

Only a few genes for quantitative resistance have been cloned to date. Some of these have been MDR loci, and their cloning has shed light on the types of mechanisms that can underlie resistance to multiple diseases. As more MDR loci reveal their secrets, we will gain insights of relevance to host-pathogen interactions and of importance to crop improvement and protection. The loci thus far associated with MDR have ranged from clusters of diversifying R genes to those involved in chemical warfare. The patterns of MDR also vary with regard to similarities in pathogen relatedness; some loci are associated with resistance to various obligate biotrophs, whereas others provide resistance to more diverse pathogens. Optimal exploitation of the potential of MDR will benefit from a deeper understanding of the underlying mechanisms and the potential trade-offs with other traits of interest to pathologists, breeders, and growers.

FUTURE ISSUES

1. Many plant defense strategies for broad resistance against biotrophs lead to susceptibility to necrotrophs and vice versa. What genes or mechanisms lead to resistance against both necrotrophic and biotrophic pathogens?
2. As more is learned about the genetic architecture underlying MDR in different crops, how can this knowledge translated into crop improvement?

3. As genetic engineering falls in price and difficulty, it will be easier to move genes from nonhosts of a given pathogen into hosts. How will this affect our understanding of host and nonhost resistance to multiple diseases? How will it change breeding for multiple resistances? Will the public accept this?

REFERENCES

1. Ali S, Singh PK, McMullen MP, Mergoum M, Adhikari TB. 2008. Resistance to multiple leaf spot diseases in wheat. Euphytica 159:167–79
2. Almasia NI, Bazzini AA, Hopp HE, Vazquez-Rovere C. 2008. Overexpression of *snakin-1* gene enhances resistance to *Rhizoctonia solani* and *Erwinia carotovora* in transgenic potato plants. Mol. Plant Pathol. 9:329– 38
3. Apel K, Hirt H. 2004. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. Annu. Rev. Plant Biol. 55:373–99
4. Atienza SG, Jafary H, Niks RE. 2004. Accumulation of genes for susceptibility to rust fungi for which barley is nearly a nonhost results in two barley lines with extreme multiple susceptibility. Planta 220:71–79
5. Bakker E, Borm T, Prins P, van der Vossen E, Uenk G, *et al.* 2011. A genome-wide genetic map of NB-LRR disease resistance loci in potato. Theor. Appl. Genet. 123:493–508
6. Balint-Kurti PJ, Yang J, Van Esbroeck G, Jung J, Smith ME. 2010. Use of a maize advanced intercross line for mapping of QTL for northern leaf blight resistance and multiple disease resistance. Crop Sci. 50:458–66
7. Benson JM, Poland JA, Benson BM, Stromberg EL, Nelson RJ. 2015. Resistance to gray leaf spot of maize: genetic architecture and mechanisms elucidated through nested association mapping and nearisogenic line analysis. PLOS Genet. 11:e1005045

8. Belcher AR, Zwonitzer JC, Cruz JS, Krakowsky MD, Chung CL, *et al.* 2012. Analysis of quantitative disease resistance to Southern leaf blight and of multiple disease resistance in maize, using near-isogenic lines. Theor. Appl. Genet. 124:433–45
9. Boller T, Felix G. 2009. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. Annu. Rev. Plant Biol. 60:379–406
10. Bolouri Moghaddam MR, Van den Ende W. 2012. Sugars and plant innate immunity. J. Exp. Bot. 63:3989–98
11. Bonman JM, Khush GS, Nelson RJ. 1992. Breeding rice for resistance to pests. Annu. Rev. Phytopathol. 30:507–28
12. Bonnet J, Danan S, Boudet C, Barchi L, Sage-Palloix A-M, *et al.* 2007. Are the polygenic architectures of resistance to *Phytophthora capsici* and *P. parasitica* independent in pepper? Theor. Appl. Genet. 115:253–64
13. Bostock RM. 2005. Signal crosstalk and induced resistance: straddling the line between cost and benefit. Annu. Rev. Phytopathol. 43:545–80
14. Bozkurt TO, Richardson A, Dagdas YF, Mongrand S, Kamoun S, Raffaele S. 2014. The plant membrane– associated REMORIN1.3 accumulates in discrete perihaustorial domains and enhances susceptibility to *Phytophthora infestans*. Plant Physiol. 165:1005–18
15. Broekaert WF, Delaure SL, De Bolle MFC, Cammue BPA. 2006. The role of ethylene in host-pathogen interactions. Annu. Rev. Phytopathol. 44:393–416
16. Brown JK. 2002. Yield penalties of disease resistance in crops. Curr. Opin. Plant Biol. 5:339–44
17. Browse J. 2009. Jasmonate passes muster: a receptor and targets for the defense hormone. Annu. Rev. Plant Biol. 60:183–205

18. Buerstmayr M, Matiasch L, Mascher F, Vida G, Ittu M, *et al.* 2014. Mapping of quantitative adult plant field resistance to leaf rust and stripe rust in two European winter wheat populations reveals co-location of three QTL conferring resistance to both rust pathogens. Theor. Appl. Genet. 127:2011–28
19. Buschges R, Hollricher K, Panstruga R, Simons G, Wolter M, *et al.* 1997. The barley *Mlo* gene: a novel control element of plant pathogen resistance. Cell 88:695–705
20. Cao H, Li X, Dong X. 1998. Generation of broad-spectrum disease resistance by overexpression of an essential regulatory gene in systemic acquired resistance. PNAS 95:6531–36
21. Cartwright HN, Humphries JA, Smith LG. 2009. PAN1: a receptor-like protein that promotes polarization of an asymmetric cell division in maize. Science 323:649–51
22. Chakrabarti A, Ganapathi TR, Mukherjee PK, Bapat VA. 2003. MSI-99, a magainin analogue, imparts enhanced disease resistance in transgenic tobacco and banana. Planta 216:587–96
23. Chen SC, Liu AR, Wang FH, Ahammed GJ. 2009. Combined overexpression of chitinase and defensin genes in transgenic tomato enhances resistance to *Botrytis cinerea*. Afr. J. Biotechnol. 8:5182–88
24. Chung C-L, Longfellow JM,Walsh EK, Kerdieh Z, Van Esbroeck G, *et al.* 2010. Resistance loci affecting distinct stages of fungal pathogenesis: use of introgression lines for QTL mapping and characterization in the maize-*Setosphaeria turcica* pathosystem. BMC Plant Biol. 10:103
25. Cooley MB, Pathirana S, Wu HJ, Kachroo P, Klessig DF. 2000. Members of the *Arabidopsis HRT*/*RPP8* family of resistance genes confer resistance to both viral and oomycete pathogens. Plant Cell 12:663–76

26. Cox TS, Raupp WJ, Wilson DL, Gill BS, Bockus WW, Browder LE. 1992. Resistance to foliar diseases in a collection of *Triticum tauschii* germplasm. Plant Dis. 76:1061–64
27. Coyne CJ, Pilet-Nayel M-L, McGee RJ, Porter LD, Smykal P, Grunwald NJ. 2015. Identification of QTL controlling high levels of partial resistance to *Fusarium solani* f. sp. *pisi* in pea. Plant Breed. 134:446–53
28. Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR. 2010. Abscisic acid: emergence of a core signaling network. Annu. Rev. Plant Biol. 61:651–79
29. Dangl JL, Jones JDG. 2001. Plant pathogens and integrated defence responses to infection. Nature 411:826–33
30. de Carvalho A, Gomes VM. 2009. Plant defensins- prospects for the biological functions and biotechnological properties. Peptides 30:1007–20
31. Dixon AGO, Ogbe FO, Okechukwu RU. 2010. Cassava mosaic disease in sub-Saharan Africa: a feasible solution for an unsolved problem. Outlook Agric. 39:89–94
32. Dong X. 2004. *NPR1*, all things considered. Curr. Opin. Plant Biol. 7:547–52
33. Eibach R, Zyprian E, Welter L, Topfer R. 2007. The use of molecular markers for pyramiding resistance genes in grapevine breeding. Vitis 46:120–24
34. Elgin JH, Hill RR, Zeiders KE. 1970. Comparison of four methods of multiple trait selection for five traits in alfalfa. Crop Sci. 10:190–93
35. Faris JD, Li WL, Liu DJ, Chen PD, Gill BS. 1999. Candidate gene analysis of quantitative disease resistance in wheat. Theor. Appl. Genet. 98:219–25
36. Ferrari S. 2013. Oligogalacturonides: plant damage-associated molecular patterns and regulators of growth and development. Front. Plant Sci. 4:1–9

37. Fetch TG, Steffenson BJ, Nevo E. 2003. Diversity and sources of multiple disease resistance in *Hordeum spontaneum*. Plant Dis. 87:1439–48

38. Flemmig EL. 2012. Molecular markers to deploy and characterize stem rust resistance in wheat. MS Thesis, N. C. State. Univ., Raleigh

39. Flint-Garcia SA, Thuillet A-C, Yu J, Pressoir G, Romero SM, *et al.* 2005. Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J. 44:1054–64

40. Fokunang CN, Ikotun T, Dixon AGO, Akem CN. 2000. Field reaction of cassava genotypes to anthracnose, bacterial blight, cassava mosaic disease, and their effects on yield. Afr. Crop Sci. J. 8:179–86

41. Fotopoulos V, Gilbert MJ, Pittman JK, Marvier AC, Buchanan AJ, *et al.* 2003. The monosaccharide transporter gene, *AtSTP4*, and the cell-wall invertase, *Atβfruct1*, are induced in *Arabidopsis* during infection with the fungal biotroph *Erysiphe cichoracearum*. Plant Physiol. 132:821–29

42. Fu J, Liu H, Li Y, Yu H, Li X, *et al.* 2011. Manipulating broad-spectrum disease resistance by suppressing pathogen-induced auxin accumulation in rice. Plant Physiol. 155:589–602

43. Gassmann W, Hinsch ME, Staskawicz BJ. 1999. The Arabidopsis *RPS4* bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. Plant J. 20:265–77

44. Gomez-Ariza J, Campo S, Rufat M, Estopa M, Messeguer J, *et al.* 2007. Sucrose-mediated priming of plant defense responses and broad-spectrum disease resistance by overexpression of the maize pathogenesis-related PRms protein in rice plants. Mol. Plant-Microbe Interact. 20:832–42

45. Govrin EM, Levine A. 2000. The hypersensitive response facilitates plant infection by the necrotrophic pathogen *Botrytis cinerea*. Curr. Biol. 10:751–57

46. Grant M, Godiard L, Straube E, Ashfield T, Lewald J, *et al.* 1995. Structure of the *Arabidopsis RPM1* gene enabling dual specificity disease resistance. Science 269:843–46

47. Greenberg JT, Ausubel FM. 1993. *Arabidopsis* mutants compromised for the control of cellular damage during pathogenesis and aging. Plant J. 4:327–41

48. Gurung S, Bonman JM, Ali S, Patel J, Myrfield M, *et al.* 2009. New and diverse sources of multiple disease resistance in wheat. Crop Sci. 49:1655–66

49. Gurung S, Hansen JM, Bonman JM, Gironella AIN, Adhikari TB. 2012. Multiple disease resistance to four leaf spot diseases in winter wheat accessions from the USDA national small grains collection. Crop Sci. 52:1640–50

50. Gurung S, Mamidi S, Bonman JM, Xiong M, Brown-Guedira G, Adhikari TB. 2014. Genome-wide association study reveals novel quantitative trait loci associated with resistance to multiple leaf spot diseases of spring wheat. PLOS ONE 9:e108179

51. Hann DR, Rathjen JP. 2007. Early events in the pathogenicity of *Pseudomonas syringae* on *Nicotiana benthamiana*. Plant J. 49:607–18

52. Heese A, Hann DR, Gimenez-Ibanez S, Jones AME, He K, *et al.* 2007. The receptor-like kinase *SER3*/*BAK1* is a central regulator of innate immunity in plants. PNAS 104:12217–22

53. Hill RR, Leath KT. 1975. Genotypic and phenotypic correlations for reaction to five foliar pathogens in alfalfa. Theor. Appl. Genet. 258:254–58

54. Hulbert SH, Webb CA, Smith SM, Sun Q. 2001. Resistance gene complexes: evolution and utilization. Annu. Rev. Phytopathol. 39:285–312

55. Hysing S-C, Hsam SLK, Singh RP, Huerta-Espino J, Boyd LA, *et al.* 2007. Agronomic performance and multiple disease resistance in T2BS.2RL wheat-rye translocation lines. Crop Sci. 47:254

56. Infantino A, Kharrat M, Riccioni L, Coyne CJ, McPhee KE, Grunwald NJ. 2006. Screening techniques and sources of resistance to root diseases in cool season food legumes. Euphytica 147:201–21

57. Jafary H, Szabo LJ, Niks RE. 2006. Innate nonhost immunity in barley to different heterologous rust fungi is controlled by sets of resistance genes with different and overlapping specificities. Mol. PlantMicrobe Interact. 19:1270–79

58. Jamann TM, Luo X, Morales L, Kolkman JM, Chung C-L, Nelson RJ. A remorin gene is implicated in quantitative disease resistance in maize. Theor. Appl. Genet. 129:591–602

59. Jamann TM, Poland JA, Kolkman JM, Smith LG, Nelson RJ. 2014. Unraveling genomic complexity at a quantitative disease resistance locus in maize. Genetics 198:333–44

60. Jansky SH, Rouse DI. 2003. Multiple disease resistance in interspecific hybrids of potato. Plant Dis. 87:266–72

61. Jarosch B, Kogel K-H, Schaffrath U. 1999. The ambivalence of the barley *Mlo* locus: mutations conferring resistance against powdery mildew (*Blumeria graminis* f. sp. *hordei*) enhance susceptibility to the rice blast fungus *Magnaporthe grisea*. Mol. Plant-Microbe Interact. 508:508–14

62. Jellis GJ. 1992. Multiple resistance to diseases and pests in potatoes. Euphytica 63:51–58

63. Jha S, Chattoo BB. 2010. Expression of a plant defensin in rice confers resistance to fungal phytopathogens. Transgenic Res. 19:373–84 Candidate ge

64. Jo Y-K, Barker R, Pfender W, Warnke S, Sim S-C, Jung G. 2008. Comparative analysis of multiple disease resistance in ryegrass and cereal crops. Theor. Appl. Genet. 117:531–43

65. Johnson R. 1981. Durable resistance: definition of, genetic control, and attainment in plant breeding. Phytopathology 71:567

66. Jones JDG, Dangl JL. 2006. The plant immune system. Nature 444:323–29

67. Kemmerling B, Schwedt A, Rodriguez P, Mazzotta S, Frank M, *et al.* 2007. The BRI1-associated kinase 1, BAK1, has a brassinolide-independent role in plant cell-death control. Curr. Biol. 17:1116–22

68. Khan MA, Shah MD, Saini RG. 2012. Multiple disease resistance of an Australian bread wheat cultivar Cook. Australas. Plant Pathol. 41:131–37

69. Kim YJ, Lin N-C, Martin GB. 2002. Two distinct *Pseudomonas* effector proteins interact with the Pto kinase and activate plant immunity. Cell 109:589–98

70. Koornneef A, Pieterse CMJ. 2008. Cross talk in defense signaling. Plant Physiol. 146:839–44

71. Kou Y, Wang S. 2012. Toward an understanding of the molecular basis of quantitative disease resistance in rice. J. Biotechnol. 159:283–90

72. Krattinger SG, Lagudah ES, Spielmeyer W, Singh RP, Huerta-Espino J, *et al.* 2009. A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. Science 323:1360–63

73. Krijger J-J, Thon MR, Deising HB, Wirsel SG. 2014. Compositions of fungal secretomes indicate a greater impact of phylogenetic history than lifestyle adaptation. BMC Genom. 15:722

74. Kuang H, Woo S-S, Meyers BC, Nevo E, Michelmore RW. 2004. Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. Plant Cell 16:2870–94

75. Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, *et al.* 2011. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat. Genet. 43:163–68

76. Lagudah ES, McFadden H, Singh RP, Huerta-Espino J, Bariana HS, Spielmeyer W. 2006. Molecular genetic characterization of the *Lr34*/*Yr18* slow rusting resistance gene region in wheat. Theor. Appl. Genet. 114:21–30

77. Lander ES, Botstein D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–99

78. Lebedev VG, Dolgov SV, Lavrova N, Lunin VG, Naroditski BS. 2002. Plant-defensin genes introduction for improvement of pear phytopathogen resistance. Acta. Hortic. 596:167–172

79. Lee J, Hong J-H, Do JW, Yoon JB. 2010. Identification of QTLs for resistance to anthracnose to two *Colletotrichum* species in pepper. J. Crop Sci. Biotechnol. 13:227–33

80. Lemonnier P, Gaillard C, Veillet F, Verbeke J, Lemoine R, *et al.* 2014. Expression of *Arabidopsis* sugar transport protein STP13 differentially affects glucose transport activity and basal resistance to *Botrytis cinerea*. Plant Mol. Biol. 85:473–84

81. Li Q, Lawrence CB, Xing HY, Babbitt RA, Bass WT, *et al.* 2001. Enhanced disease resistance conferred by expression of an antimicrobial magainin analog in transgenic tobacco. Planta 212:635–39

82. Li Z, Zhou M, Zhang Z, Ren L, Du L, *et al.* 2011. Expression of a radish defensin in transgenic wheat confers increased resistance to *Fusarium graminearum* and *Rhizoctonia cerealis*. Funct. Integr. Genom. 11:63–70

83. Lin W-C, Lu C-F, Wu J-W, Cheng M-L, Lin Y-M, *et al.* 2004. Transgenic tomato plants expressing the *Arabidopsis NPR1* gene display enhanced resistance to a spectrum of fungal and bacterial diseases. Transgenic Res. 13:567–81

84. Mago R, Tabe L, McIntosh RA, Pretorius Z, Kota R, *et al.* 2011. A multiple resistance locus on chromosome arm 3BS in wheat confers resistance to stem rust (*Sr2*), leaf rust (*Lr27*) and powdery mildew. Theor. Appl. Genet. 123:615–23

85. Manosalva PM, Davidson RM, Liu B, Zhu X, Hulbert SH, *et al.* 2009. A germin-like protein gene family functions as a complex quantitative trait locus conferring broad-spectrum disease resistance in rice. Plant Physiol. 149:286–96

86. Marrs KA. 1996. The functions and regulation of glutathione S-transferases in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 47:127–58

87. McGrann GRD, Stavrinides A, Russell J, Corbitt MM, Booth A, *et al.* 2014. A trade off between *mlo* resistance to powdery mildew and increased susceptibility of barley to a newly important disease, ramularia leaf spot. J. Exp. Bot. 65:1025–37

88. Meyers BC, Chin DB, Shen KA, Sivaramakrishnan S, Lavelle DO, *et al.* 1998. The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell 10:1817–32

89. Miklas PN, Delorme R, Stone V, Daly MJ, Stavely JR, *et al.* 2000. Bacterial, fungal, and viral disease resistance loci mapped in a recombinant inbred common bean population ("Dorado"/XAN 176). J. Am. Soc. Hortic. Sci. 125:476–81

90. Miller-Garvin JE, Viands DR. 1994. Selection for resistance to Fusarium root rot, and associations among resistances to six diseases in alfalfa. Crop Sci. 34:1461–65

91. Mitchell-Olds T, James RV, Palmer MJ, Williams PH. 1995. Genetics of *Brassica rapa* (syn. *campestris*). 2. Multiple disease resistance to three fungal pathogens: *Peronospora parasitica*, *Albugo candida* and *Leptosphaeria maculans*. Heredity 75:362–69

92. Mizobuchi R, Hirabayashi H, Kaji R, Nishizawa Y, Yoshimura A, *et al.* 2002. Isolation and characterization of rice lesion-mimic mutants with enhanced resistance to rice blast and bacterial blight. Plant Sci. 163:345–53

93. Monteagudo AB, Rodino AP, Lema M, De la Fuente M, Santalla M, De Ron AM. 2006. Resistance to ˜ infection by fungal, bacterial, and viral pathogens in a common bean core collection from the Iberian Peninsula. HortScience 41:319–22

94. Moore JW, Herrera-Foessel S, Lan C, Schnippenkoetter W, Ayliffe M, *et al.* 2015. A recently evolved hexose transporter variant confers resistance to multiple pathogens in wheat. Nat. Genet. 47:1494–98

95. Mukhtar MS, Carvunis A-R, Dreze M, Epple P, Steinbrenner J, *et al.* 2011. Independently evolved virulence effectors converge onto hubs in a plant immune system network. Science 333:596–601

96. Narusaka M, Shirasu K, Noutoshi Y, Kubo Y, Shiraishi T, *et al.* 2009. *RRS1* and *RPS4* provide a dual resistance gene system against fungal and bacterial pathogens. Plant J. 60:218–26

97. Narusaka Y, Narusaka M, Park P, Kubo Y, Hirayama T, *et al.* 2004. *RCH1*, a locus in *Arabidopsis* that confers resistance to the hemibiotrophic fungal pathogen *Colletotrichum higginsianum*. Mol. Plant-Microbe. Interact. 17:749–62

98. Nene YL. 1988. Multiple-disease resistance in grain legumes. Annu. Rev. Phytopathol. 26:203–17

99. Neuffer MG, Calvert OH. 1975. Dominant disease lesion mimics in maize. J. Hered. 66:265–70

100. Ng TB. 2004. Antifungal proteins and peptides of leguminous and non-leguminous origins. Peptides 25:1215–22

101. Niks RE, Qi X, Marcel TC. 2015. Quantitative resistance to biotrophic filamentous plant pathogens: concepts, misconceptions, and mechanisms. Annu. Rev. Phytopathol. 53:445–70

102. Nothnagel EA, McNeil M, Albersheim P, Dell A. 1983. Host-pathogen interactions. Plant Physiol. 71:916–26

103. Orton WA. 1902. The wilt diseases of the cowpea and its control. US Dep. Agric. Bur. Plant Ind. Bull. 17:9–20

104. Osusky M, Zhou G, Osuska L, Hancock RE, Kay WW, Misra S. 2000. Transgenic plants expressing cationic peptide chimeras exhibit broad-spectrum resistance to phytopathogens. Nat. Biotechnol. 18:1162–66

105. Pajerowska-Mukhtar KM, Mukhtar MS, Guex N, Halim VA, Rosahl S, *et al.* 2008. Natural variation of *potato allene oxide synthase 2* causes differential levels of jasmonates and pathogen resistance in *Arabidopsis*. Planta 228:293–306

106. Pan RS, More TA. 1996. Screening of melon (*Cucumis melo* L.) germplasm for multiple disease resistance. Euphytica 88:125–28

107. Pataky J, Williams MM. 2010. Reactions of sweet corn hybrids to prevalent diseases. Bull. Univ. Ill., Urbana, IL

108. Piffanelli P, Zhou F, Casais C, Orme J, Jarosch B, *et al.* 2002. The barley MLO modulator of defense and cell death is responsive to biotic and abiotic stress stimuli. Plant Physiol. 129:1076–85

109. Poland J. 2010. *The Genetic Architecture of Quantitative Resistance in Maize*. PhD Thesis, Cornell Univ., Ithaca, NY

110. Poland JA, Bradbury PJ, Buckler ES, Nelson RJ. 2011. Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. PNAS 108:6893–98

111. Portieles R, Ayra C, Gonzalez E, Gallo A, Rodriguez R, *et al.* 2010. *Nmdef02*, a novel antimicrobial gene isolated from *Nicotiana megalosiphon* confers high-level pathogen resistance under greenhouse and field conditions. Plant Biotechnol. J. 8:678–90

112. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. 2000. Association mapping in structured populations. Am. J. Hum. Genet. 67:170–81

113. Raffaele S, Bayer E, Lafarge D, Cluzet S, German Retana S, *et al.* 2009. Remorin, a Solanaceae protein resident in membrane rafts and plasmodesmata, impairs Potato virus X movement. Plant Cell Online 21:1541–55

114. Ramalingam J, Vera Cruz CM, Kukreja K, Chittoor JM, Wu J, *et al.* 2003. Candidate defense genes from rice, barley, and maize and their association with qualitative and quantitative resistance in rice. Mol. Plant-Microbe Interact. 16:14–24

115. Risterucci AM, Paulin D, Ducamp M, N'Goran JAK, Lanaud C. 2003. Identification of QTLs related to cocoa resistance to three species of *Phytophthora*. Theor. Appl. Genet. 108:168–74

116. Robert-Seilaniantz A, Grant M, Jones JDG. 2011. Hormone crosstalk in plant disease and defense: more than just jasmonate-salicylate antagonism. Annu. Rev. Phytopathol. 49:317–43

117. Rustagi A, Kumar D, Shekhar S, Yusuf MA, Misra S, Sarin NB. 2014. Transgenic *Brassica juncea* plants expressing MsrA1, a synthetic cationic antimicrobial peptide, exhibit resistance to fungal phytopathogens. Mol. Biotechnol. 56:535–45

118. Santino A, Taurino M, De Domenico S, Bonsegna S, Poltronieri P, *et al.* 2013. Jasmonate signaling in plant development and defense response to multiple (a)biotic stresses. Plant Cell Rep. 32:1085–98

119. Schweizer P, Stein N. 2011. Large-scale data integration reveals colocalization of gene functional groups with meta-QTL for multiple disease resistance in barley. Mol. Plant-Microbe Interact. 24:1492–1501

120. Silva KJP, Brunings A, Peres NA, Mou Z, Folta KM. 2015. The *Arabidopsis NPR1* gene confers broad spectrum disease resistance in strawberry. Transgenic Res. 24:693–704

121. Singh A, Singh VK, Singh SP, Pandian RTP, Ellur RK, *et al.* 2012. Molecular breeding for the development of multiple disease resistance in Basmati rice. AoB Plants 4:pls029

122. Smedegaard-Petersen V, Tolstrup K. 1985. The limiting effect of disease resistance on yield. Annu. Rev. Phytopathol. 23:475–90

123. Spielmeyer W, McIntosh RA, Kolmer J, Lagudah ES. 2005. Powdery mildew resistance and *Lr34*/*Yr18* genes for durable resistance to leaf and stripe rust cosegregate at a locus on the short arm of chromosome 7D of wheat. Theor. Appl. Genet. 111:731–35

124. St. Clair DA. 2010. Quantitative disease resistance and quantitative resistance loci in breeding. Annu. Rev. Phytopathol. 48:247–68

125. Sun L, Yang DL, Kong Y, Chen Y, Li XZ, *et al.* 2014. Sugar homeostasis mediated by cell wall invertase GRAIN INCOMPLETE FILLING 1 (GIF1) plays a role in pre-existing and induced defence in rice. Mol. Plant Pathol. 15:161–73

126. Swathi Anuradha T, Divya K, Jami SK, Kirti PB. 2008. Transgenic tobacco and peanut plants expressing a mustard defensin show resistance to fungal pathogens. Plant Cell Rep. 27:1777–86

127. Takahashi H, Miller J, Nozaki Y, Takeda M, Shar J, *et al.* 2002. RCY1, an *Arabidopsis thaliana RPP8/HRT* family resistance gene, conferring resistance to cucumber mosaic virus requires salicylic acid, ethylene and a novel signal transduction mechanism. Plant J. 32:655–67

128. Tang X, Xie M, Kim YJ, Zhou J, Klessig DF, Martin GF. 1999. Overexpression of *Pto* activates defense responses and confers broad resistance. Plant Cell 11:15–29

129. Teran H, Jara C, Mahuku G, Beebe S, Singh SP. 2013. Simultaneous selection for resistance to five ´ bacterial, fungal, and viral diseases in three Andean × Middle American inter-gene pool common bean populations. Euphytica 189:283–92

130. Terras F. 1995. Small cysteine-rich antifungal proteins from radish: their role in host defense. Plant Cell 7:573–88

131. Thomma B, Cammue B, Thevissen K. 2002. Plant defensins. Planta 216:193–202

132. Tian D, Traw M, Chen J, Kreitman M, Bergelson J. 2003. Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. Nature 423:74–77

133. Timmerman-Vaughan GM, Frew TJ, Russell AC, Khan T, Butler R, *et al.* 2002. QTL mapping of partial resistance to field epidemics of Ascochyta blight of pea. Crop Sci. 42:2100

134. Todesco M, Balasubramanian S, Hu TT, Traw MB, Horton M, *et al.* 2010. Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. Nature 465:632–36

135. Torres MA, Jones JDG, Dangl JL. 2006. Reactive oxygen species signaling in response to pathogens. Plant Physiol. 141:373–78

136. Van Der Vossen EAG, Van Der Voort JNAM, Kanyuka K, Bendahmane A, Sandbrink H, *et al.* 2000. Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: a virus and a nematode. Plant J. 23:567–76

137. van Eeuwijk FA, Mesterhazy A, Kling CI, Ruckenbauer P, Saur L, *et al.* 1995. Assessing non-specificity of resistance in wheat to head blight caused by inoculation with European strains of *Fusarium culmorum*, *F. graminearum* and *F. nivale* using a multiplicative model for interaction. Theor. Appl. Genet. 90:221–28

138. Villegas-Fernandez AM, Sillero JC, Emeran AA, Flores F, Rubiales D. 2011. Multiple-disease resistance in *Vicia faba*: multi-environment field testing for identification of combined resistance to rust and chocolate spot. Field Crop. Res. 124:59–65

139. Vivas M, Silveira SF, Pio-Viana A, Amaral-Junior AT, Ferreguetti GA, Pereira MG. 2015. Resistance to multiple foliar diseases in papaya genotypes in Brazil. Crop Prot. 71:138–43

140. Vlot AC, Dempsey DA, Klessig DF. 2009. Salicylic acid, a multifaceted hormone to combat disease. Annu. Rev. Phytopathol. 47:177–206

141. Wallace JG, Bradbury PJ, Zhang N, Gibon Y, Stitt M, Buckler ES. 2014. Association mapping across numerous traits reveals patterns of functional variation in maize. PLOS Genet. 10:e1004845

142. Wally O, Jayaraj J, Punja ZK. 2009. Broad-spectrum disease resistance to necrotrophic and biotrophic pathogens in transgenic carrots (*Daucus carota* L.) expressing an *Arabidopsis NPR1* gene. Planta 231:131–41

143. Wally O, Punja ZK. 2010. Enhanced disease resistance in transgenic carrot (*Daucus carota* L.) plants over-expressing a rice cationic peroxidase. Planta 232:1229–39

144. Walters D, Heil M. 2007. Costs and trade-offs associated with induced resistance. Physiol. Mol. Plant Pathol. 71:3–17

145. Wang G-X, Chen Y, Zhao J-R, Li L, Korban SS, *et al.* 2007. Mapping of defense response gene homologs and their association with resistance loci in maize. J. Integr. Plant Biol. 49:1580–98

146. Webber HJ, Orton WA. 1902. A cowpea resistant to root-knot (*Heterodera radicicola*). US Dep. Agric. Bur. Plant Ind. Bull. 17:23–26

147. Wisser RJ, Balint-Kurti PJ, Nelson RJ. 2006. The genetic architecture of disease resistance in maize: a synthesis of published studies. Phytopathology 96:120–29

148. Wisser RJ, Kolkman JM, Patzoldt ME, Holland JB, Yu J, *et al.* 2011. Multivariate analysis of maize disease resistances suggests a pleiotropic genetic basis and implicates a GST gene. PNAS 108:7339–44

149. Wisser RJ, Sun Q, Hulbert S, Kresovich S, Nelson RJ. 2005. Identification and characterization of regions of the rice genome associated with broad-spectrum, quantitative disease resistance. Genetics 169:2277–93

150. Wright SAL, Azarang M, Falk AB. 2013. Barley lesion mimics, supersusceptible or highly resistant to leaf rust and net blotch. Plant Pathol. 62:982–92

151. Wroblewski T, Caldwell KS, Piskurewicz U, Cavanaugh KA, Xu H, *et al.* 2009. Comparative large-scale analysis of interactions between several crop species and the effector repertoires from multiple pathovars of *Pseudomonas* and *Ralstonia*. Plant Physiol. 150:1733–49

152. Wroblewski T, Piskurewicz U, Tomczak A, Ochoa O, Michelmore RW. 2007. Silencing of the major family of NBS-LRR-encoding genes in lettuce results in the loss of multiple resistance specificities. Plant J. 51:803–18

153. Wu C, Bordeos A, Madamba MRS, Baraoidan M, Ramos M, *et al.* 2008. Rice lesion mimic mutants with enhanced resistance to diseases. Mol. Genet. Genom. 279:605–19

154. Yin Z, Chen J, Zeng L, Goh M, Leung H, *et al.* 2000. Characterizing rice lesion mimic mutants and identifying a mutant with broad-spectrum resistance to rice blast and bacterial blight. Mol. Plant-Microbe Interact. 13:869–76

155. Yu D, Chen C, Chen Z. 2001. Evidence for an important role of WRKY DNA binding proteins in the regulation of NPR1 gene expression. Plant Cell 13:1527–40

156. Yu J, Holland JB, McMullen MD, Buckler ES. 2008. Genetic design and statistical power of nested association mapping in maize. Genetics 178:539–51

157. Zasloff M. 2002. Antimicrobial peptides of multicellular organisms. Nature 415:389–95

158. Zeng ZB. 1994. Precision mapping of quantitative trait loci. Genetics 136:1457–68

159. Zhang X, Dai Y, Xiong Y, DeFraia C, Li J, *et al.* 2007. Overexpression of Arabidopsis MAP kinase kinase 7 leads to activation of plant basal and systemic acquired resistance. Plant J. 52:1066–79

160. Zhang XY, Nie ZH, Wang WJ, Leung DWM, Xu DG, *et al.* 2013. Relationship between disease resistance and rice oxalate oxidases in transgenic rice. PLOS ONE. 8:e78348

161. Zhang Z, Ersoz E, Lai C, Todhunter RJ, Tiwari HK, *et al.* 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42:355–60

162. Zhou J, Zhang H, Yang Z, Li G, Hu L, *et al.* 2012. Characterization of a new T2DS.2DL-?R translocation triticale ZH-1 with multiple resistances to diseases. Genet. Resour. Crop Evol. 59:1161–68

163. Zhou Y-L, Xu J-L, Zhou S-C, Yu J, Xie X-W, *et al.* 2009. Pyramiding *Xa23* and *Rxo1* for resistance to two bacterial diseases into an elite indica rice variety using molecular approaches. Mol. Breed. 23:279–87

164. Zipfel C. 2008. Pattern-recognition receptors in plant innate immunity. Curr. Opin. Immunol. 20:10–16

165. Zwart RS, Thompson JP, Milgate AW, Bansal UK, Williamson PM, *et al.* 2010. QTL mapping of multiple foliar disease and root-lesion nematode resistances in wheat. Mol. Breed. 26:107–24

166. Zwonitzer JC, Coles ND, Krakowsky MD, Arellano C, Holland JB, *et al.* 2010. Mapping resistance quantitative trait loci for three foliar diseases in a maize recombinant inbred line population: evidence for multiple disease resistance? Phytopathology 100:72–79

CHAPTER 2

TRANSCRIPTOMICS OF THE BIOTROPHY-NECROTROPHY TRANSITION AND R-GENE-MEDIATED RESISTANCE IN THE *SETOSPHAERIA TURCICA*-MAIZE PATHOSYSTEM[2]

ABSTRACT

Hemibiotrophic interactions are multifaceted and complex, as the mechanisms deployed by pathogens and their host(s) can change throughout the course of pathogenesis. In this study, RNA-seq was used to explore the *Setosphaeria turcica*- maize pathosystem as it transitioned from biotrophy to necrotrophy in four combinations of host and pathogen. Both pathogen and host transcriptomes shifted dramatically throughout the biotrophic-necrotrophic transition. Pathogen avirulence in the presence of a corresponding host R gene led to an apparent arrest of pathogen growth and development. Gene-poor regions of the *S. turcica* genome tended to contain more genes differentially expressed during infection, but these genes did not appear to evolve more rapidly. Overall differences of gene expression in the host between resistant and susceptible interactions were largely quantitative- of degree and timing- rather than qualitative differences in the identity of the genes and pathways involved. RNA-seq data were used to identify two plausible candidates for the maize *Ht2* R gene. Maize gene expression in response to *S. turcica* infection could be used to predict maize resistance phenotypes from non-inoculated gene expression levels.

2 Wiesner-Hanks T, Mideros S, Wu D, Haridas S, Andreopoulos W, Singan V, Daum C, Barry K, Saha S, Condon B, Grigoriev IV, Nelson RJ, and Turgeon BG. Transcriptomics of the biotrophy-necrotrophy transition and R-gene-mediated resistance in the *Setosphaeria turcica*-maize pathosystem. In preparation.

INTRODUCTION

Fungal plant pathogens are typically divided into three classes based on lifestyle and nutritional relationship with their host: biotrophs, necrotrophs, and hemibiotrophs. Biotrophs enter host cells without triggering plant defenses and thrive on living host tissue. Necrotrophs kill host tissue and are sustained by released nutrients. Hemibiotrophs begin their relationship with the host as biotrophs, then transition to a necrotrophic lifestyle (Perfect and Green 2008).

Setosphaeria turcica (*Exserohilum turcicum*) is a hemibiotrophic ascomycete and the causal agent of northern leaf blight (NLB) of maize (Leonard et al. 1989, Lim et al. 1974, Chung et al. 2010). Several races have been described (Leonard et al. 1989, Ferguson et al. 2004, Ferguson et al. 2007) based on their ability to infect maize lines carrying the resistance genes *Ht1*, *Ht2*, *Ht3*, and/or *HtN*. *S. turcica* race 1 is virulent on maize carrying *Ht1*, while race 23 is virulent on maize carrying *Ht2* and/or *Ht3*, etc. NLB is one of the most important maize foliar diseases, causing an estimated economic loss of roughly $1.9 billion in 2016 in the US (Mueller et al. 2017). Despite its economic importance, *S. turcica* has not been studied as much as other hemibiotrophic fungi such as *Colletotrichum* spp., *Zymoseptoria tritici* or *Magnaporthe oryzae*.

Previous transcriptomic studies of hemibiotrophs have shown that gene expression profiles of both pathogen and host are usually dissimilar during the biotrophic and necrotrophic phases (Brunner et al. 2013; Gan et al. 2013; Yang et al. 2013; Palma-Guerrero et al. 2016). In the dicot pathogen *Colletotrichum higginsianum*, for example, categories of genes typically associated with virulence (e.g., those encoding enzymes for secondary metabolite (SM) biosynthesis, small secreted proteins/effector proteins (SSPs), cell wall degradation (CAZys), transporters, peptidases, transcription factors) are transcribed in ‘waves’ roughly corresponding to lifestyle phase (Gan et al, 2013; O’Connell

et al. 2012). SM, SSP, and CAZy associated genes tend to be upregulated in the early biotrophic phase, while transporters and hydrolases are associated with the later necrotrophic phase (O'Connell et al. 2012). While most SM genes are upregulated in the fungus *in planta*, expression profiles are diverse: some are upregulated in appressoria *in vitro* and also *in planta*, and others are upregulated in appressoria *in planta* and in the biotrophic phase, some during both bio- and necrotrophic phases and others only in the necrotophic phase (Dallery 2017).

Another observation regarding virulence-related fungal genes is their location in the genome. In some fungi, genes mediating fungal virulence are found preferentially in gene-poor regions. SSPs in particular tend to be located in gene-poor, high A+T regions in *Leptosphaeria maculans* (Rouxel et al. 2011; Winter et al. 2018). In the well-studied rice blast fungus *M. oryzae*, effector-encoding genes are preferentially located near telomeric regions, while for the emerging wheat blast strain, many effector genes are located on gene poor mini chromosomes (Peng et al., 2018).

Previous RNA-seq studies have elucidated the temporal dimensions of hemibiotrophy in different host-pathogen interactions. In many pathosystems, gene expression profiles in both pathogen and host are typically quite different during these two phases (Brunner et al. 2013; Gan et al. 2013; Yang et al. 2013; Palma-Guerrero et al. 2016; Wang et al. 2018). Certain pathogen infection strategies, such as the production of polyketides and small peptides, have been observed in many pathosystems (Gan et al, 2013; O'Connell et al. 2012). Other offensive strategies, such as the production of hydrophobins during necrotrophy, have so far been noted in only a single pathosystem (Yang et al. 2013; Meinhardt et al. 2014).

Many studies have explored the transcriptomic aspects of R-gene-mediated (qualitative) resistance, but findings vary widely from pathosystem to pathosystem, defying generalizations.

Differential gene expression (DGE) in response to infection may be more pronounced in the resistant host than the susceptible host (Kawahara et al. 2012; Sonah et al. 2016) or just the opposite (Matić et al. 2016). The genes or pathways that are up- or downregulated by the resistant and susceptible hosts may be very similar (Sonah et al. 2016) or quite distinct (Han et al. 2015). Some plant defense strategies against hemibiotrophs, such as chitinase production, appear to be fairly conserved among various hosts (Yang et al. 2013; Meinhardt et al. 2014; Gharbi et al. 2017).

Many maize loci conditioning NLB resistance manifest their effects at varying stages of symptom development, making the maize-*S. turcica* pathosystem a useful model to study hemibiotrophic plant-pathogen interactions. For instance, the R gene *HtN* delays lesion formation (Hurni et al. 2015), while the nearby R gene *Ht2* slows the rate of lesion growth (Chung et al. 2010a). Quantitative resistance loci (QRL) can also have stage-specific effects. Two maize QRL derived from the same resistant donor line were found to impart resistance via distinct mechanisms: one reduces hyphal penetration of the leaf, hindering the pathogen during early biotrophy, while the other inhibits penetration of the host vascular tissue and delays the biotrophy-necrotrophy transition (Chung et al. 2010). These diverse host strategies provide many avenues for comparison of different resistance mechanisms.

To explore the shared and unique aspects of biotrophy and necrotrophy, and of resistant and susceptible interactions, we analyzed 78 RNA-seq libraries derived from four combinations of host (either carrying or lacking the *Ht2* resistance gene) and pathogen (either carrying or lacking the *AVRHt2* gene, which causes a resistance response in maize carrying the *Ht2* gene; Table 2.1) and the appropriate controls. Each host/pathogen combination was assayed over several time points to explore

the different aspects of biotrophy and necrotrophy and which patterns were shared or unique. The patterns of gene regulation during the biotrophic and necrotrophic phases were quite distinct, and characterization of these sets gives a clear insight into host and pathogen strategy during the two phases.

	***S. turcica* isolate**	
Maize host line	StNY001 [*AVRHt2*]	St28A [*avrHt2*]
Zm*Ht2*+	Resistant *Ht2/AVRHt2*	Susceptible *Ht2/avrHt2*
Zm*Ht2*-	Susceptible *ht2/AVRHt2*	Susceptible *ht2/avrHt2*

Table 2.1. Genotypes and expected R-gene mediated interaction of host lines and pathogen isolates used in this study.

RESULTS

Fungal genome structure

The StNY001 genome was sequenced and assembled for the first time, allowing a comparison to the previously reported St28A genome (Condon et al. 2013) and more accurate alignment of transcript

reads from StNY001. The assembled genomes were comparable by most metrics, though the St28A assembly was both slightly larger and contained fewer gaps (Table 2.2). Alignment with nucmer showed that 98.5% of the St28A genome and 87.5% of the StNY001 genome were covered by regions of at least 99% sequence similarity, the larger St28A assembly contained proportionally more unique genomic regions. Most gene models were common to both isolates: 11,012/12,028 (91.6%) of St28A gene models and 11,012/12,547 (87.8%) of StNY001 gene models had a reciprocal best hit (RBH) in the other isolate. These shared gene models were often identical: 7,024/11,012 (63.8%) of shared models had 100% sequence identity.

Table 2.2. Summary of *S. turcica* genome assemblies.

Genome characteristic	**St28A**[a]	**StNY001**[b]
Genome sequence total (Mb)	43.01	38.42
Genome scaffold count	407	489
Genome contig count	1,959	1,823
Scaffold N50	8	45
Scaffold L50 (Mb)	2.14	0.23
% genome covered by gaps	11.10%	6.00%
# gene models	12,028	12,547
# shared gene models	11,012	11,012

# unique gene models	1,016	1,535

[a] https://genome.jgi.doe.gov/Settu3/Settu3.home.html

[b] https://genome.jgi.doe.gov/Settur3/Settur3.home.html

Sequencing and alignment of reads

Sequencing of 78 RNA-seq libraries- 24 from mock-inoculated maize tissue, 24 from maize tissue inoculated with *S. turcica* isolate 28A (St28A), 24 from maize tissue inoculated with *S. turcica* isolate NY001 (StNY001), and 6 from axenic *S. turcica* tissue- yielded 3.6 billion reads in total, for an average of 47 million (M) reads per library (Table 2.3). In total, 81.4% of reads were successfully aligned to a unique position overlapping with a single annotated transcript from the appropriate genome. The percentage of reads mapped was comparable in most tissue types (roughly 80-82%), except in the St28A axenic libraries, in which only 73.0% of reads aligned.

Although reads were aligned to both fungal and maize genome FASTA files simultaneously, alignment of fungal reads to the maize genome and vice versa was rare. Of the 328 M reads from the six *S. turcica* axenic libraries, 17,083 (0.0052%) mapped to the maize genome. Of the 1,067 M reads from the 24 maize mock-inoculated libraries, only 5,062 (0.00047%) mapped to the *S. turcica* genome.

Table 2.3. Summary of sequencing and alignment of reads.

Source	**M reads**	**# libs**	**M reads/lib**	**% mapped**	**% unmapped**
All libraries	3634	78	47	81.4	18.6
Mock-inoculated maize	1067	24	44	81.4	18.6
St28A-inoculated maize	1121	24	47	82	18
StNY001-inoculated maize	1118	24	47	82	18
Axenic St28A	119	3	40	73	27
Axenic StNY001	209	3	70	80.4	19.6

Differential Expression

Host and pathogen gene counts in inoculated tissue were tested for differential expression (DE) relative to gene counts in the appropriate negative control: mock-inoculated maize tissue for host genes and axenic cultures for fungal genes. A large proportion of both host and pathogen genes were DE at some point in at least one host/pathogen combination: 49.8% (19,741/39,625) of maize gene models, 67.3% of St28A gene models, and 33.4% of StNY001 gene models were significantly DE ($p < 0.01$ after false discovery rate [FDR] correction) at least at one time point during pathogenesis.

In all host/pathogen combinations, the proportion of genes that were significantly DE ($p < 0.01$ after FDR correction) *in planta* generally increased over time in both host and pathogen (Fig. 2.1). These patterns of differential expression were more pronounced in host/pathogen combinations involving St28A than those involving StNY001. In *S. turcica*, the proportion of genes that were DE (*in planta v.* axenic) was 2-3x higher in St28A than in StNY001 across all time points, regardless of the presence or absence of *Ht2* in the host. The proportion of maize genes that were down- or up-regulated in inoculated tissue vs. mock-inoculated tissue at most time points was similar across all four host-pathogen combinations, although by 10 DPI there were more up- or down-regulated genes in maize infected with St28A (13.3-17.2%) than in maize infected with StNY001 (9.9-11.9%). The proportion of StNY001 genes up- or down- regulated *in planta* was similar in the *Ht2+* and *Ht2-* hosts.

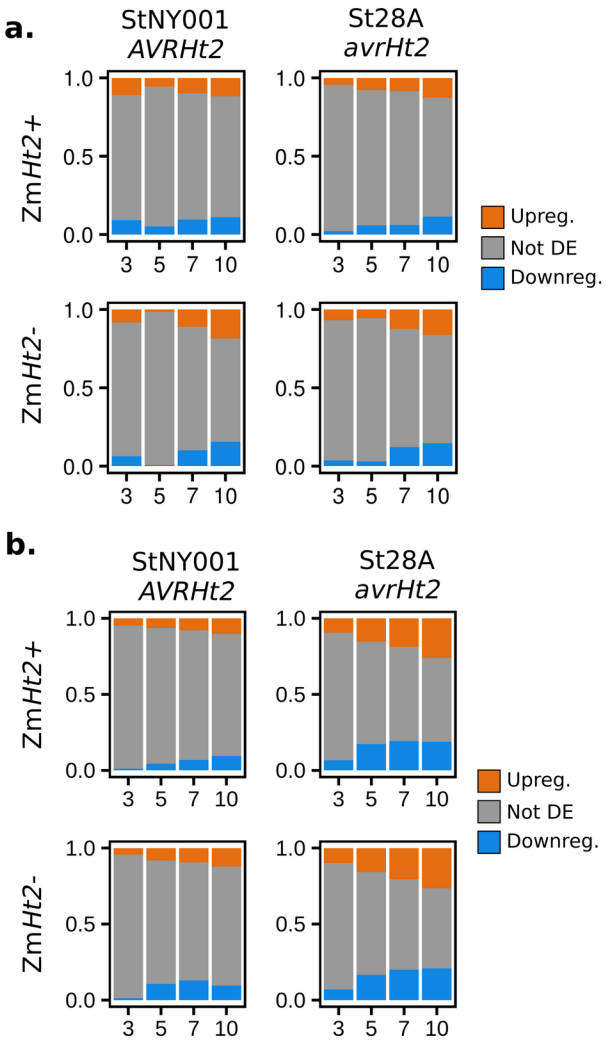


Figure 2.1: The proportion of genes that was significantly differentially expressed (DE) during infection relative to negative control tended to increase over time in both the host maize (**a**) and pathogen *S. turcica* (**b**) in all host-pathogen combinations.

Gene models common to both fungal isolates were roughly 2.5 times more likely to be differentially expressed (DE) *in planta* compared to gene models unique to one isolate (Fig. 2.2). Across all time points in all host/pathogen combinations, 25.6% of *S. turcica* gene models shared by both isolates were significantly DE between *in planta* and axenic libraries, compared to only 10.6% of isolate-specific gene models. Baseline (axenic) expression of isolate-specific genes was also much lower. The median transcript count in axenic libraries for gene models present in both isolates was 12.5, while the median count for gene models found in only a single isolate was 0.684. These findings were not surprising, as transcript reads from axenic cultures were used to augment gene calling, so genes that were highly expressed in axenic cultures were presumably more likely to be called as gene models.

Examination of the most significantly DE gene models (*in planta* expression vs. axenic) in *S. turcica* revealed several interesting trends. Of the 42 gene models that were among the 10 most significantly DE in at least one host/pathogen/DPI combination, only 12 had a predicted function or protein domain (Supplementary Table S2.1). We consider those with no predicted function or protein domain to be prime targets for functional analysis and discovery of novel virulence determinants.

Some gene models were highly DE across all hosts/timepoints, e.g. StNY001 protein ID 170288 (predicted chitin-binding domain), which suggests a general role across all phases of pathogenesis. Others seemed to serve a biotrophy-specific role, such as St28A 19500 (multicopper oxidase, type 1/2/3) and 185131 (peptidase S10), and StNY001 436826 (haem peroxidase), all among the top DE genes at 3-7 DPI in both hosts, or a necrotrophy-specific role, such as StNY001 538563 (galactosyl transferase), among the top DE genes on both hosts at 10 DPI (Supplementary Table S2.1).

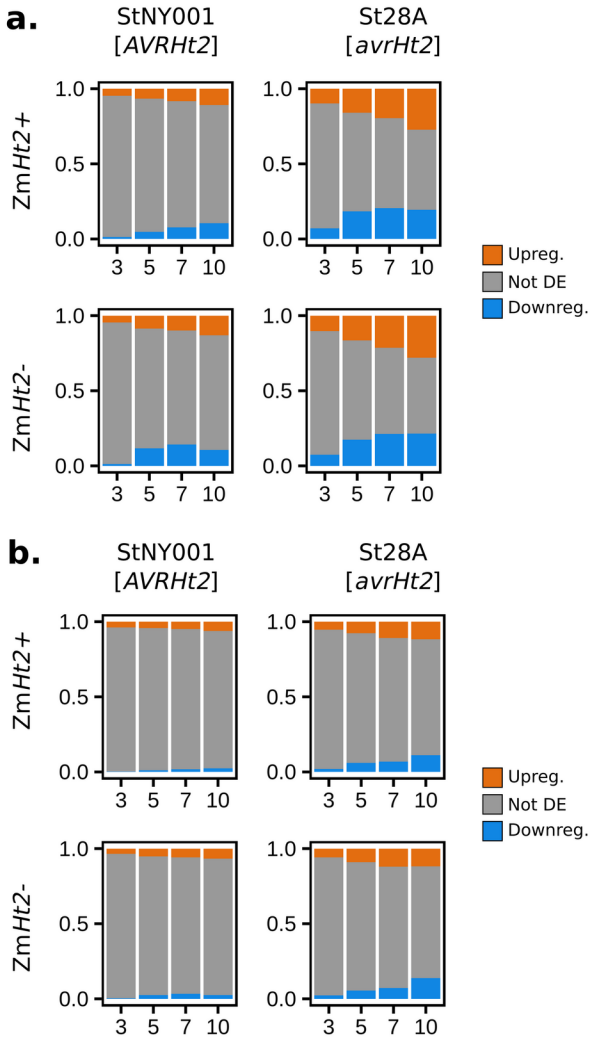


Figure 2.2: *S. turcica* gene models present in both isolates are more likely to be significantly differentially expressed (DE) *in planta* relative to axenic. **a.** Gene models found in both St28A and StNY001. **b.** Gene models found in only one isolate.

We also analyzed expression patterns of all *S. turcica* gene models with a given protein domain, in addition to considering gene models individually. Enrichment analysis identified 22 Pfam domains significantly enriched among up- or down-regulated *S. turcica* gene models (Supplementary Table S2.2, 15 enriched among upregulated gene models, 7 enriched among downregulated gene models). Three protein families were enriched among upregulated gene models in most combinations of host/isolate/timepoint (Fig. 2.3), again suggesting a general role: cytochrome P450 proteins (PF00067), major facilitator superfamily (MFS) proteins (PF07690), and zinc-binding dehydrogenases (PF00107). All three of these protein types are often associated with secondary metabolite gene clusters. PF1500324 (amino acid permease), PF00775 (dioxygenase), and PF00067 were notable in being upregulated at 5 days in the resistance interaction only. The same was true of PF00561 (alpha/beta hydrolase fold) at 10 days. PF02133 (permease) was upregulated in all three susceptible interactions at 10 days, but not in the resistant interaction.

In maize multiple gene models with the same predicted protein domain were among the most highly DE models across most of the host/isolate/timepoint combinations (Supplementary Table S2.3). This included predicted cytochrome P450 genes (GRMZM2G118809, GRMZM2G161472), ABC transporter genes (GRMZM2G413774, GRMZM2G391815, GRMZM2G415529), and terpene synthase genes (AC214360.3_FG001, GRMZM2G028306, GRMZM2G127087). Other gene models were among the most highly DE models predominantly in one host, e.g. GRMZM2G169240/GRMZM2G169261 (fatty acid desaturases, mostly DE in the Zm*Ht2+* host), or at a specific timepoint, e.g. GRMZM2G078667 (dirigent-like protein, top DE gene model at 3 DPI) or GRMZM2G090980 (zinc-binding dehydrogenase, top DE gene model at 10 DPI). The domain-level trends of maize differential expression- which classes of genes or predicted protein domains were often

induced or suppressed during pathogenesis- were strongly influenced by the phase of pathogenesis and the presence/absence of R-gene-mediated resistance, as described below.

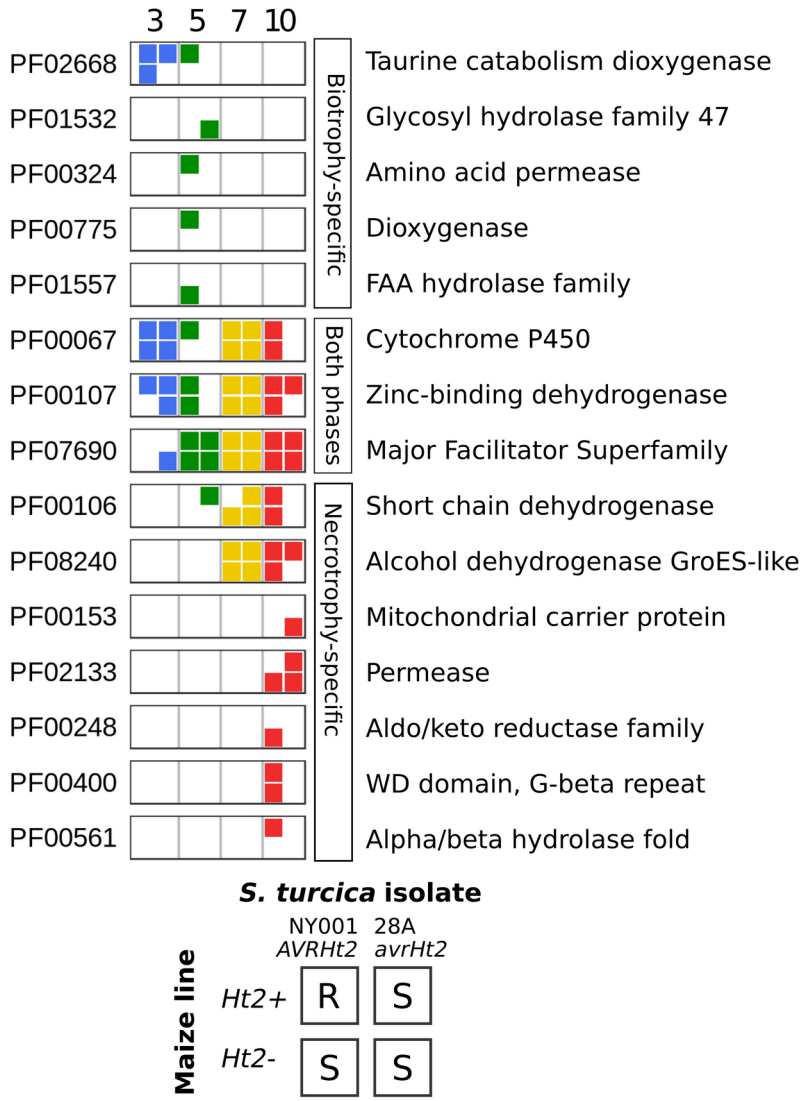


Figure 2.3: Pfam domains enriched among *S. turcica* gene models upregulated *in planta*. Each 2x2 square represents the four combinations of host and pathogen, as shown in the key at bottom.

The biotrophy-necrotrophy transition

Necrotic lesions first appeared between 7 and 10 DPI, reflecting the transition from biotrophy to necrotrophy. During this transition, fungal biomass in susceptible maize tissue, as measured by proportion of fungal reads over total reads, grew rapidly from 1.9-2.8% at 7 DPI to 4.8-14% at 10 DPI (Fig. 2.4). In contrast, fungal biomass in the resistant Zm*Ht2*+/StNY001[*AVRHt2*] reaction was far less in this same period (doubled from 0.3-0.5% to 0.6-1.0%).

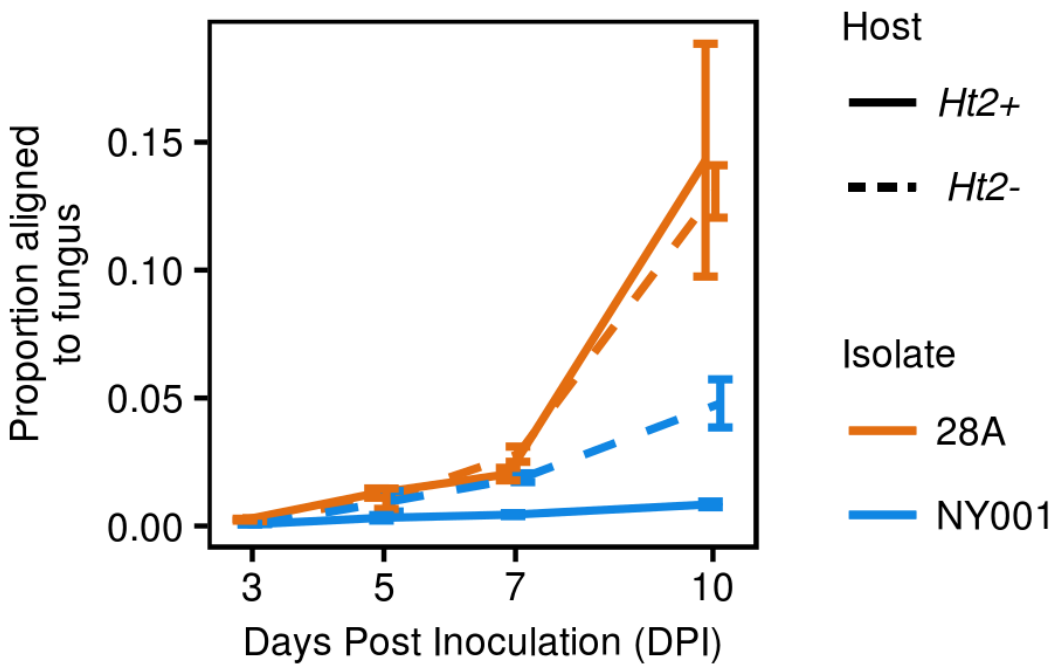


Figure 2.4: Fungal biomass increased dramatically between 7 and 10 DPI, except in the resistant interaction (Zm*Ht2*+/StNY001). Lines show mean proportion of aligned reads from inoculated maize tissue that aligned to the *S. turcica* genome, rather than to the maize genome. Bars indicate standard error.

Total biomass was significantly lower in the resistant *Ht2*+/NY001 reaction than in the susceptible *Ht2*-/NY001 reaction at 7 and 10 dpi (Fig. 2.4). The mean fungal biomass for strain NY001 in the susceptible interaction increased 2.5x, while in the resistant interaction it increased 1.9x between 7 and 10 DPI. During this same period, the increase in fungal biomass was much larger in the two susceptible interactions involving strain 28A. The presence of the *Ht2* gene did not impair the relative growth of pathogen biomass as it entered the necrotrophic phase.

The biotrophy-necrotrophy transition was characterized by sharp, global changes in gene expression patterns in both host and pathogen, as shown by principal component analysis (PCA) of the transcriptomes. In *S. turcica*, the greatest global expression differences were between the axenic libraries and the four sets of *in planta* libraries (Fig. 2.5a). *In planta* expression patterns in *S. turcica* changed very little between 3 and 7 DPI, while expression patterns between 7 and 10 DPI changed dramatically in the three susceptible combinations, but not in the single resistance interaction.

In the host transcriptomes, the differentiation between inoculated and control (mock-inoculated) tissue was less extreme (Fig. 2.5b). Expression patterns did change gradually throughout the biotrophic phase, but changes of a similar degree were observed in the mock-inoculated tissue, suggesting that these were more due to tissue maturation than infection by *S. turcica*. As in the pathogen, there was then a large shift between expression at 7 DPI and 10 DPI.

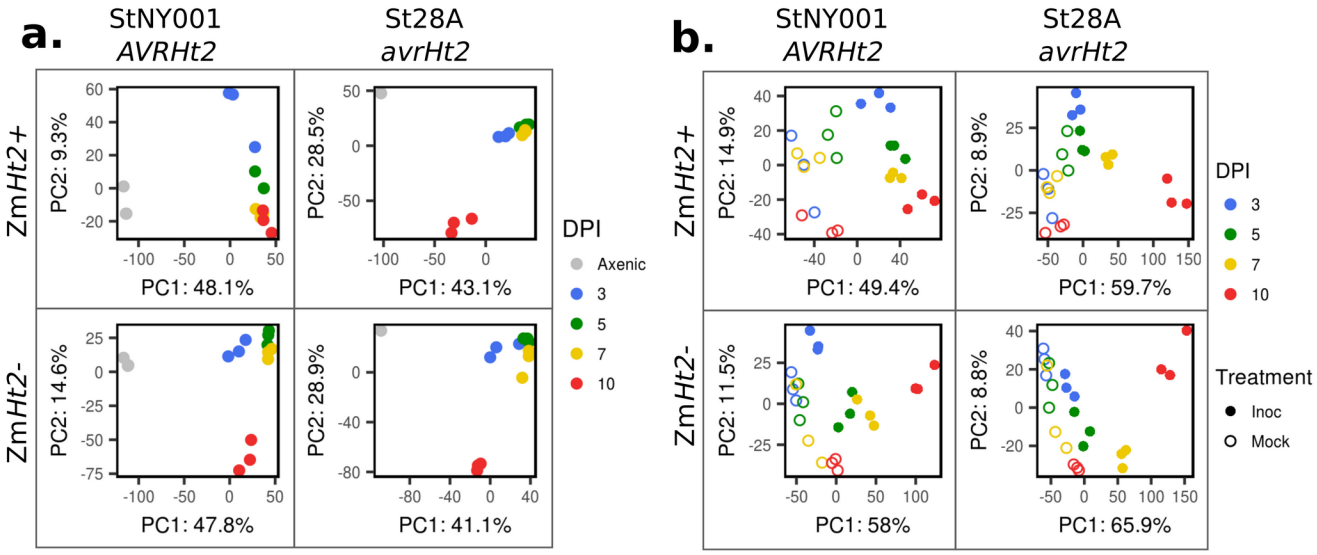


Figure 2.5: Global gene expression in both host and pathogen changes dramatically between 7 and 10 DPI in susceptible host/pathogen interactions, but not in the resistant interaction (Zm*Ht2*+/StNY001, top left of each panel). **a.** Principal component analysis (PCA) of regularized-log-transformed counts of *S. turcica* transcripts in inoculated maize tissue and axenic culture (control). **b.** PCA of regularized-log-transformed counts of maize transcripts in inoculated and mock-inoculated (control) tissue.

The types of maize genes upregulated at each time point characterize the changes happening during this shift. In maize, 100 Pfam domains were significantly enriched ($p < 0.01$ after Bonferroni correction) among upregulated genes in at least one host-pathogen-DPI combination, while only 14 domains were similarly enriched among downregulated maize genes. Pfam domains enriched among upregulated maize genes fell into three broad categories: enriched only during biotrophy (3-7 DPI), enriched only during necrotrophy (10 DPI), and enriched across both phases (Fig. 2.6). Five of the seven “biotrophy-specific” domains (PF13855 through PF00139) were enriched at 10 DPI in the R

reaction, fitting with the previous observation that the host transcriptome of the R reaction at 10 DPI most closely resembled the R host transcriptome at 7 DPI. Proteins with LRR domains are strong candidates for a role in resistance (DeYoung and Innes 2006).

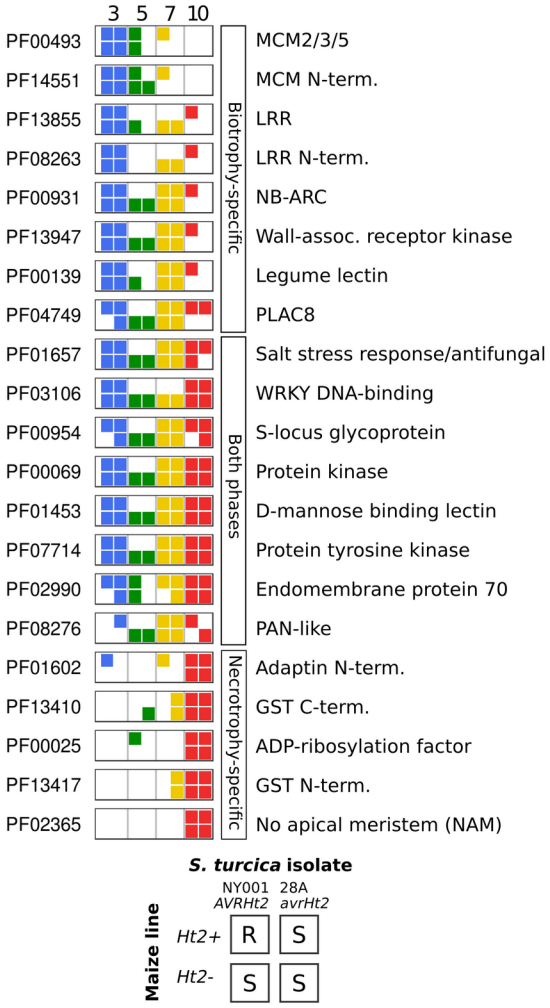


Figure 2.6: Pfam domains enriched among maize gene models upregulated in infected leaves were often enriched only during the pathogen's biotrophic phase (3-7 DPI) or necrotrophic phase (10 DPI). Each 2x2 square represents the four combinations of host and pathogen, as shown in the key at bottom.

Resistant v. susceptible interactions

The four combinations of host and pathogen provided two ways to investigate the effects of fungal *AVR* and host R genes in interactions: 1) a pathogen strain with or without an *AVR* gene encoding an effector (+/- *AVRHt2*) and 2) a host with or without an R gene encoding a protein for resistance (+/- *Ht2*). Only one combination, *AVRHt2/ Ht2+* brokered a resistance reaction; the other three (*AVRHt2/Ht2-*, *avrHt2/Ht2+*, *avrHt2/Ht2-*) led to susceptible interactions.

Differences in host gene expression in resistant and susceptible interactions were most notable at the start of infection. In response to StNY001[*AVRHt2*] at 3 DPI, thousands of genes were DE (both up- and down- regulated) by the resistant host (Zm*Ht2+*), while there were few genes DE in the susceptible host (Zm*Ht2-*)(Table 2.4). By 5 DPI, this trend reversed and more than a thousand genes were DE (both up- and down- regulated) by the susceptible host (Zm*Ht2-*), but not in the resistant host (Zm*Ht2+*). This trend quickly diminished: by 10 DPI, roughly 75% of genes up- or down-regulated in one host were similarly up- or down-regulated in the other. Those genes that were significantly DE in both hosts in response to StNY001[*AVRHt2*] at 3 DPI were 1.32 times more strongly up- or down-regulated in the R reaction. Expression of genes in the resistant host (Zm*Ht2+*) inoculated with virulent strain St28A[*avrHt2*] or avirulent strain StNY001[*AVRHt2*] tended to be similarly up- or down-regulated (Table 2.5).

Very few host genes were regulated in an opposite manner between the R and S interactions, i.e. upregulated in one host in response to StNY001 and downregulated in the other. Of the 16,057 maize genes DE in response to StNY001 at any time point, only 37 were upregulated in Zm*Ht2+* and downregulated in Zm*Ht2-* or vice versa. The host genes DE in the susceptible and resistant interaction

also tended to be of the same classes. No Pfam domains or GO terms were significantly enriched among host genes up- or downregulated in Zm*Ht2*+/StNY001 compared to those up- or downregulated in Zm*Ht2*-/StNY001 at any time point.

Table 2.4: Number of gene models up- or downregulated in maize with or without R gene (*Ht2*+/*Ht2*-), inoculated with *S. turcica* with the corresponding *AVR* gene (StNY001[*AVRHt2*]).

	Upregulated			**Downregulated**		
DPI	Zm*Ht2*+ only (R)	Both	Zm*Ht2*- only (S)	Zm*Ht2*+ only (R)	Both	Zm*Ht2*- only (S)
3	3056	1316	420	3187	454	341
5	762	1461	1611	843	1211	1053
7	1683	2228	1138	2239	1563	776
10	1330	3340	1619	1187	3156	1397

Table 2.5: Number of gene models up- or downregulated in maize with the corresponding R gene (Zm*Ht2*), inoculated with fungus with or without the functional *AVRHt2* gene.

	Upregulated			**Downregulated**		
DPI	StNY001 [*AVRHt2*] only (R)	Both	St28A [*avrHt2*] only (S)	StNY001 [*AVRHt2*] only (R)	Both	St28A [*avrHt2*] only (S)
3	1335	3037	272	1413	2228	310
5	1707	516	86	1861	193	48
7	842	3069	1313	844	2958	1049
10	531	4139	3212	525	3818	2334

Fungal effectors and secondary metabolites

We had strong *a priori* interest in three fungal gene classes: polyketide synthase (PKS) encoding genes, non-ribosomal peptide synthetase (NRPS) encoding genes, and those encoding small secreted proteins (SSP), as all classes encode proteins that may act as effectors. The number of predicted genes in these classes were similar in both isolates, and SSPs were by far the most numerous. In St28A, we predicted 357 SSP, 23 *NRPS*, and 27 *PKS* genes, and in StNY001, 434 SSP, 22 *NRPS*, and 29 *PKS* genes (Supplementary Table S2.4). Of the 56 *NRPS/PKS* genes predicted in either isolate, 49 were present in both isolates. There were two predicted hybrid *PKS/NRPS* genes in both isolates, one of which (protein

179218 in St28A, 444637 in StNY001) has been identified as *AVRHt1* (Mideros et al. 2018). Knowing the aforementioned, we investigated the other (protein 47468 in St28A, protein 444637 in StNY001) as a candidate *AVRHt2* gene. The predicted gene model was 575 aa longer in StNY001 than in St28A. However, expression of this gene *in planta* was extremely low, in contrast to the highly expressed *AVRHt1*, and it was not on a scaffold genetically associated with *AVRHt2* (Mideros et al. 2018).

Predicted fungal SSP-, PKS-, and NRPS-encoding genes followed the same general pattern of expression in both isolates (Table 2.6). Roughly 20-30% were upregulated *in planta* at any given time point, a proportion that increased slightly over time. Expression profiles of the upregulated PKS- and NRPS-encoding genes varied, but most often they were highly upregulated evenly across all timepoints or increasingly so over time (Fig. 2.7). The overall percentage of *S. turcica* genes upregulated *in planta* roughly tripled from 9.3-9.4% to 25.5-26.0% between 3 and 10 DPI (Fig. 2.1). SSP-, PKS-, NRPS-encoding genes were 3x more likely than the average gene model to be upregulated at the start of pathogenesis. The AVRHt1 gene, a hybrid PKS/NRPS (protein 179218 in St28A, 444637 in StNY001) was upregulated in all combinations of host/pathogen/timepoint.

Table 2.6: Upregulation of predicted NRPS-, PKS-, and SSP-encoding genes in *S. turcica* on the resistant Zm*Ht2*+ host. Upregulated NRPS and PKS-encoding genes are listed by JGI protein ID in parentheses for each time point.

			# upregulated by DPI			
Isolate	**Gene class**	**# genes**	**3**	**5**	**7**	**10**
St28A	All	12028	1120	1837	2259	3126
St28A	SSP	357	67	60	73	78
St28A	NRPS	23	6 [36641 54477 179218 169407 155102 173669]	5 [36641 54477 179218 155102 173669]	5 [36641 54477 179218 155102 173669]	6 [18754 54477 179218 169407 155102 173669]
St28A	PKS	27	5 [179218 161586 92491 93994 158064]	4 [179218 92491 93994 158064]	5 [179218 161586 92491 93994 158064]	8 [179218 152662 30113 158567 161586 92491 93994 22370 100517 158064]
StNY001	All	12547	1185	1917	2467	3196
StNY001	SSP	434	64	59	61	79
StNY001	NRPS	22	3 [444637 227272 550530]	4 [444637 543585 468908	5 [444637 543585 468908	5 [444637 543585 468908

				550530]	550530 607936]	550530 607936]
StNY001	PKS	29	5 [444637 443749 434732 552647 606828]	5 [444637 443749 434732 552647 606828]	5 [444637 443749 434732 552647 606828]	5 [444637 443749 434732 552647 606828]

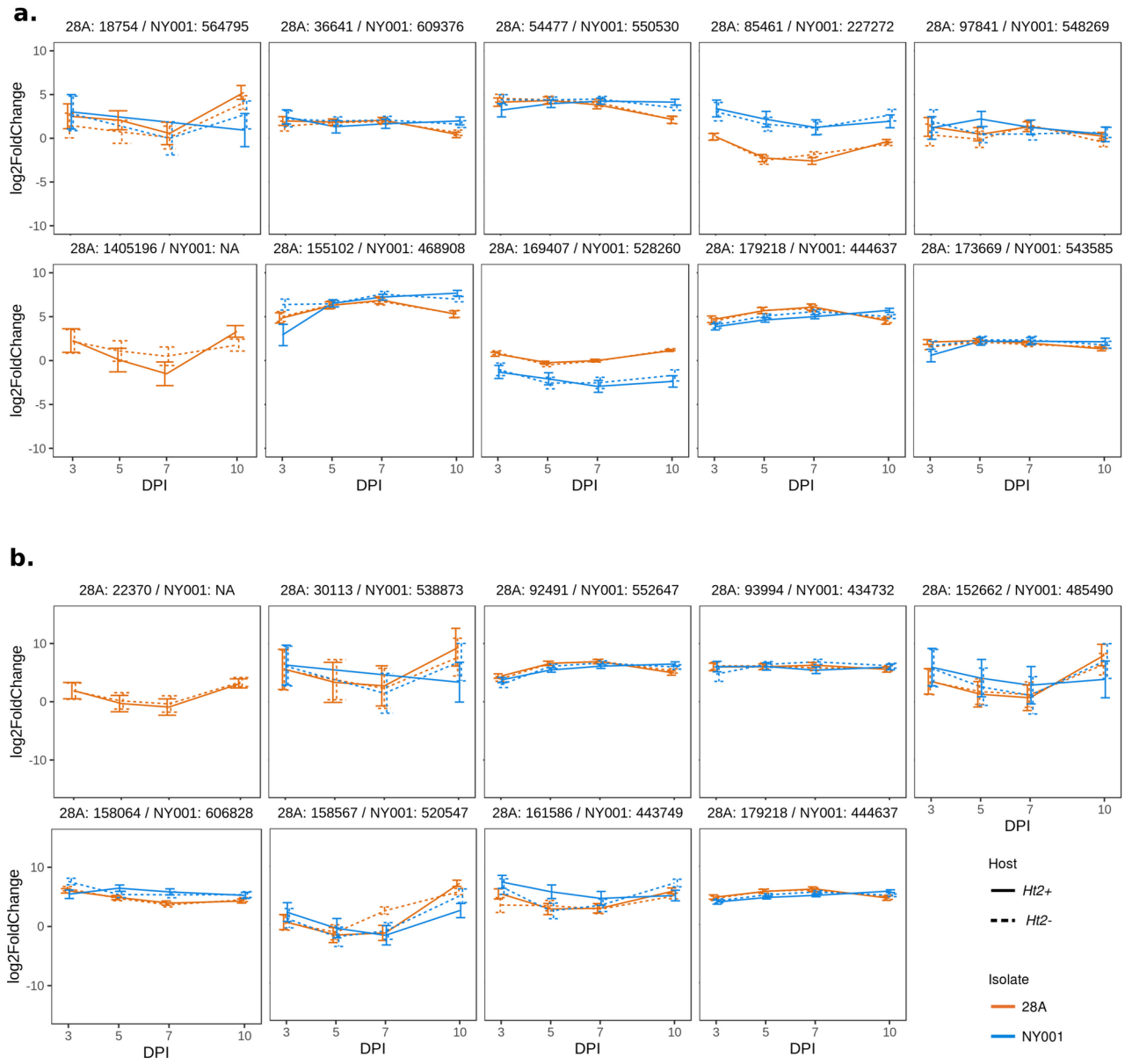


Figure 2.7: Expression profiles of *S. turcica* predicted PKS- and NRPS-encoding genes that are significantly upregulated in at least one host/pathogen/DPI combination.

Fungal gene-sparse regions

In both *S. turcica* strain 28A and NY001, genes that lay 10 kb or further from their nearest neighbor had a much higher mean ratio of *in planta*/axenic expression than genes less than 10 kb from their nearest neighbor (Fig. 2.8). In St28A, genes >10 kb from the nearest neighbor (63 genes, 0.52% of total) had a mean *in planta*/axenic expression ratio of 200, compared to the global average of 1.09. In StNY001, genes >10 kb from their nearest neighbor (25 genes, 0.20% of total) had a mean expression ratio of 9.6, compared to the global mean of 1.0.

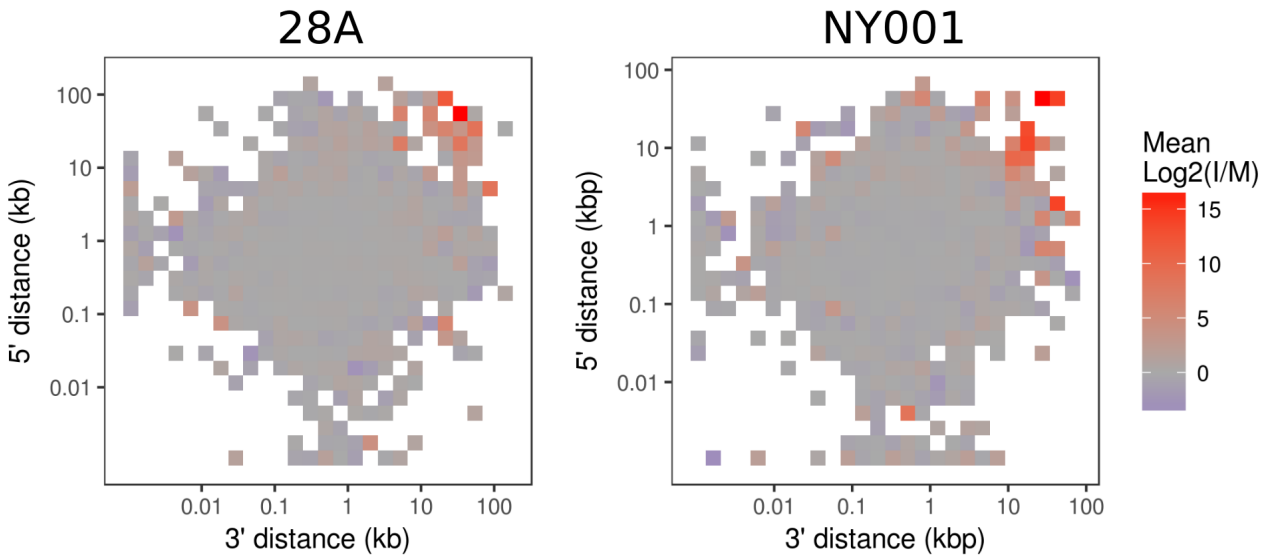


Figure 2.8: *S. turcica* gene models in gene-sparse regions tend to be highly expressed *in planta* in both St28A (left) and StNY001 (right). All gene models were binned by distance to nearest proximal gene in both the 3' and 5' directions. Color indicates mean log-transformed ratio of *in planta* to axenic expression in a given bin.

Predicted fungal SSP-, PKS-, or NRPS-encoding genes also tended to lie in more gene-sparse regions (Table 2.7). Relative to the average gene model, the average predicted SSP in either strain of *S. turcica* tended to have one fewer gene within 25 kb on either side and be 500-1100 bp farther from its closest neighboring genes, while a predicted PKS-encoding gene had 3-5 fewer nearby genes and be 1400-4000 bp farther from its closest neighbors. These differences between predicted SSP- and SM-encoding genes and non-SSP or non-SM genes were all significant as determined by Wilcoxon rank-sum test ($p < 0.01$, after Bonferroni correction). Predicted SSP-encoding genes also tended to be more likely to be very distant (>10 kb) from their closest neighbor. In isolate 28A, 6 of 357 predicted SSP genes (1.7%, protein IDs 25985, 32729, 161801, 164559, 1382176, and 1384223) were >10 kb from the closest neighboring gene, compared to 0.52% of non-SSP genes (p = 0.010 by Fisher's exact test). In isolate NY001, 5 of 434 predicted SSP genes (1.2%, protein IDs 169173, 444690, 518715, 531761, and 609418), compared to 0.20% of non-SSP genes (p = 0.0034 by Fisher's exact test). There were no predicted PKS- or NRPS-encoding genes >10 kb away from their closest neighbor.

S. turcica genes in gene-sparse regions did not have especially high ratios of non-synonymous to synonymous mutations (dN/dS) relative to predicted orthologs in other pathogenic Pleosporales (Supplementary Fig. S2.1). Thus, these isolated genes did not show signs of more extreme diversifying or directional selection than the average *S. turcica* gene.

Table 2.7: Number of nearby gene models tends to be lower and distance to nearest gene model tends to be higher for predicted *S. turcica* small secreted protein (SSP), polyketide synthase (*PKS*), and non-ribosomal peptide synthase (NRPS) encoding genes. P-values are given for Wilcoxon Rank-Sum (WRS) test comparing given class of predicted gene to all gene models not in that class.

Gene class	**St28A Gene models in 50 kb window**		**NY001 Gene models in 50 kb window**	
	Mean	p (WRS)	Mean	p (WRS)
All	15.6		16.2	
SSP	14.6	2.22E-05	14.9	4.59E-06
PKS	12.2	5.24E-04	11.3	1.03E-05
NRPS	13.4	2.10E-02	13.4	3.10E-02

Gene class	**St28A Distance (kb) to nearest gene model**		**StNY001 Distance (kb) to nearest gene model**	
	Mean	p (WRS)	Mean	p (WRS)
All	1.82		1.24	
SSP	2.9	3.10E-12	1.9	7.00E-15
PKS	5.9	8.88E-05	2.67	5.46E-06
NRPS	2.1	1.50E-02	1.5	1.40E-01

Effects of the *Ht2* introgression

The maize lines used in this study were the most isogenic *Ht2+/Ht2-* pair available, but the exact size of the introgression was not known. After testing gene models for DE between non-inoculated Zm*Ht2+* and Zm*Ht2-* tissue, we found clear breakpoints at 148.8 Mb and 160.8 Mb on chromosome 8 of the AGPv3 reference genome. This region contained 368 gene models, including the R gene *HtN* (GRMZM2G164612), known to be very close to *Ht2* from genetic mapping (Chung et al. 2010b; Jamann 2013).

In non-inoculated maize tissue, 109/368 (29.6%) of the gene models in the *Ht2* introgression were DE between Zm*Ht2+* and Zm*Ht2-* at at least one time point. The introgression also affected the baseline expression of genes *in trans*: of the gene models outside the introgression, 1,149/38,955 (2.95%) were DE between the two non-inoculated hosts at at least one time point. However, there was no statistically significant enrichment of any Pfam domain or GO term among the genes DE between non-inoculated Zm*Ht2+* and Zm*Ht2-* tissue. Thus, though we observed many differences in expression, the biological significance of these differences, if any, was unclear.

Maize benzoxazinoid pathway

Benzoxazinoids, a class of hydroxamic acids, have been studied mostly for their role in plant defense against insect herbivores (Niemeyer 2009), but also have a role in defense against fungal pathogens, including *S. turcica* in maize (Long et al. 1975; Long et al. 1978; Ahmad et al. 2011). We divided maize benzoxazinoid pathway genes into three phases: biosynthesis of DIMBOA-Glc, differentiation of DIMBOA-Glc into various related compounds, and activation of these compounds by deglucosidation (Fig. 2.9).

Each set of genes had distinctive expression patterns. The genes that collectively synthesize DIMBOA glucoside (*Bx1*, *Bx2*, *Bx3*, *Bx4*, *Bx5*, and *Bx6*) were all downregulated in the later stages of virulence (7-10 DPI), possibly due to inhibition feedback caused by DIMBOA accumulation (Ahmad et al., 2011). The pathogen isolate influenced certain genes; *Bx3* and *Bx5* were both significantly more downregulated at 10 DPI in response to St28A than in response to StNY001. The presence or absence of *Ht2* did not influence expression of any of the genes involved in DIMBOA glucoside production.

We observed two trends among the genes that convert DIMBOA-Glc into different benzoxazinoid compounds (termed "differentiation" phase in Fig. 2.9; the products are still glucosinated and therefore inactive). First, *Bx10*, *Bx11*, and *Bx12*, all of which encode O-methyltransferases of similar or identical function (Meihls et al. 2013), were all strongly upregulated at 10 DPI (245x - 26,800x in susceptible host/pathogen combinations), suggesting a necrotrophy-specific role to HDMBOA. We observed different responses in the different host/pathogen combinations; *Bx10* and *Bx11* were significantly upregulated by 7 DPI in the three susceptible host/pathogen interactions, but not in the resistant interaction. Second, *Bx14*, an O-methyltransferase gene responsible for the conversion of DIM2BOA to HDM2BOA (Handrick et al., 2016), was upregulated during both biotrophy (43x - 477x, 3-7 DPI) and necrotrophy (300x - 9360x, 10 DPI). This gene was notable as the only Bx pathway gene significantly upregulated at 3 DPI. Bx13 was not significantly DE at any time point in any host/pathogen combination.

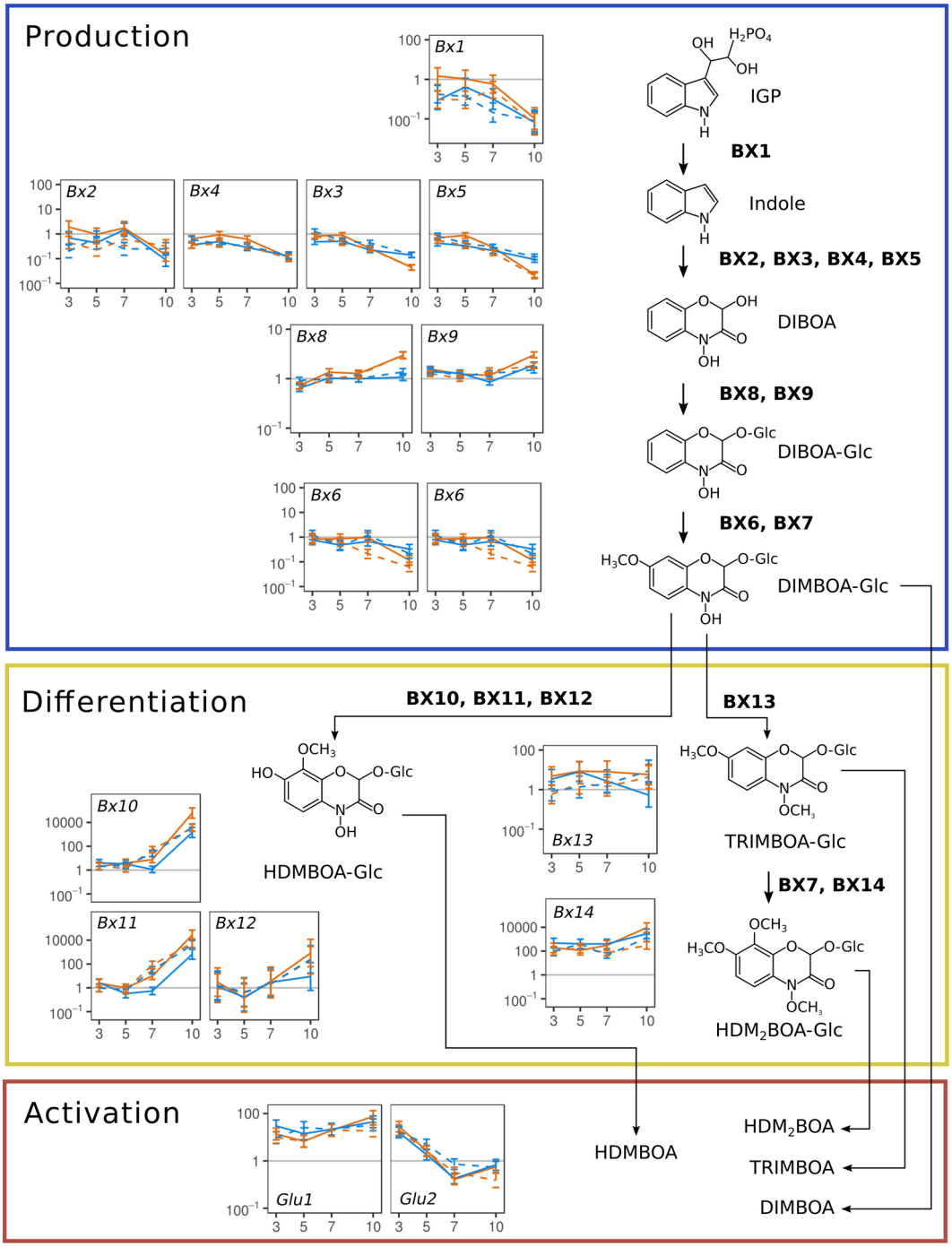

Production
Bx1
Bx2
Bx4
Bx3
Bx5
Bx8
Bx9
Bx6
Bx6
IGP
BX1
Indole
BX2, BX3, BX4, BX5
DIBOA
BX8, BX9
DIBOA-Glc
BX6, BX7
DIMBOA-Glc
Differentiation
BX10, BX11, BX12
BX13
HDMBOA-Glc
Bx10
Bx11
Bx12
Bx13
Bx14
TRIMBOA-Glc
BX7, BX14
HDM2BOA-Glc
Activation
Glu1
Glu2
HDMBOA
HDM2BOA
TRIMBOA
DIMBOA

Figure 2.9: Expression of maize benzoxazinoid pathway genes. Pathway genes have been divided into three categories based on their role: production, differentiation, and activation. Y-axis of each expression graph represents ratio of gene expression in inoculated tissue vs. mock-inoculated tissue in log scale. Orange lines represent expression in St28A-inoculated lines, blue lines represent expression in StNY001-inoculated lines. Solid lines indicate expression in Zm*Ht2*+, while dashed lines indicate expression in Zm*Ht2*-.

The maize beta-glucosidase genes *Glu1* and *Glu2*, which have similar sequences but different substrate specificity (Czjzek et al. 2000) and expression dynamics (Gomez-Anduro et al. 2011), were expressed quite differently in response to infection. *Glu1* was highly upregulated across all time points. In contrast, *Glu2* was upregulated at 3 DPI in all hosts (16x - 28x) but returned to baseline (non-inoculated) levels by 5 DPI and was not significantly DE thereafter. Expression of these genes was not significantly different ($p < 0.01$) between the four host/pathogen combinations at any time point.

Our observations were mostly consistent with HPLC profiling of maize benzoxazinoid compounds by Ahmad et al. (2011), with the key exception that they observed accumulation of HDMBOA by 5 DPI, much earlier than we saw upregulation of *Bx10/Bx11/Bx12*. However, this was observed in plants inoculated at 8 days old, much younger than those used in this study.

Maize diversity and resistance

We used several methods to relate the patterns of gene expression observed in this study to chromosomal regions associated with NLB by linkage mapping or GWAS. We found no relationship

between whether a gene was DE in response to *S. turcica* and whether it was in or near an NLB-associated QTL or SNP. We were, however, able to predict NLB phenotypes from non-inoculated expression of NLB-induced genes with fairly high accuracy.

NLB-responsive genes were not any more likely than non-NLB-responsive genes to be located within QTL for NLB resistance. Roughly one fifth (20.45%, or 8,044/39,323) of AGPv3 gene models were located inside joint linkage QTL for NLB resistance in the maize NAM population calculated by Poland et al. (2011). Under a Fisher's exact test, there was no significant difference between the proportion of genes that were DE in genes lying within NLB resistance QTL and those lying outside these QTL. A variety of comparisons were made (e.g. proportion in/not in QTL that were downregulated during necrotrophy in at least one sample, proportion that were upregulated during biotrophy only in the R interaction, etc.), but in no case was a significant difference in the proportion of pathogen-induced genes observed, even before correcting for multiple tests.

Maize genes that were significantly DE in at least one inoculated tissue (49.8% of genes) did tend to be closer to GWAS hits for NLB resistance than genes that were not DE in response to *S. turcica*. DE genes were roughly 460 kb closer to the nearest association (15.10 Mb vs. 15.56 Mb), a significant ($p = 0.0059$) difference by t-test after log-transformation of distances. However, closer examination showed that this was mostly a function of gene density, as maize genes in gene-rich regions were slightly more likely to be DE in response to NLB infection. The distance to nearest NLB association was strongly affected by gene density (Supplementary Fig. S2.2), and after accounting for gene density, there was no significant difference between the proximity to NLB associations between DE and non-DE genes ($p = 0.785$).

Though NLB-inducible genes showed no relationship to NLB resistance loci, expression of these genes was a fairly good predictor of NLB phenotype in diverse maize lines. We used ridge regression to predict NLB phenotype of lines in the Goodman maize diversity panel from gene expression in non-inoculated leaf tissue (Kremling et al. 2017), measuring prediction accuracy (r) by 10-fold cross-validation (Fig. 2.10). Those maize genes that were significantly upregulated in response to StNY001 at 3 DPI in Zm*Ht2*+, but not in Zm*Ht2*- (i.e. those genes characteristic of a resistance response), were organized in increasing order of p-value for the test of significant difference between inoculated and mock-inoculated expression levels. From the expression of the top 100 NLB-induced genes, we could predict NLB phenotype with a median accuracy of $r = 0.29$, compared with a median r of 0.16 when sampling 100 random highly-expressed genes. This difference in accuracy diminished as more genes were considered: the expression of 3000 NLB-induced genes could predict NLB phenotypes with median r of 0.55, compared to a median r of 0.46 when using 3,000 randomly sampled genes. This was unsurprising, as larger random samples will tend to have proportionally more overlap with the NLB-induced set.

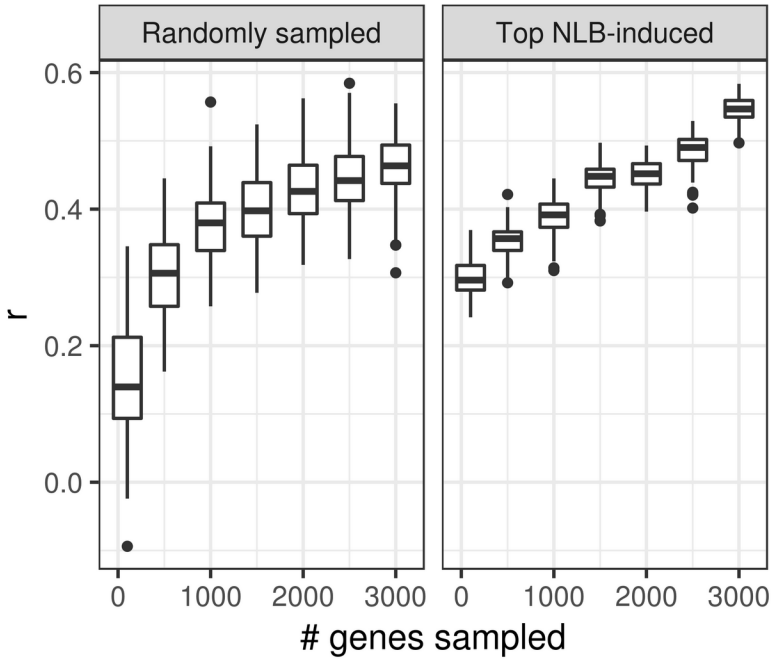


Figure 2.10: Non-induced expression of maize genes that are induced in response to NLB infection is more predictive of NLB phenotype than non-induced expression of maize genes in general. Boxplots show accuracies (*r*) when predicting NLB phenotype from expression of randomly sampled genes (left) or of NLB-responsive genes (right) in non-inoculated leaf tissue of 282 diverse maize lines. Boxplots indicate quartiles of *r* from 100 bootstraps. Points indicate *r* values further than 1.5 times the inter-quartile distance from the median *r*.

Candidate *Ht2* genes

We used several criteria to identify the most plausible *Ht2* candidate among gene models in the B73 (AGPv3.31) reference genome. First, the gene model should be within 500 kb of the center of the *Ht2* fine-mapping region. Second, *de novo* assembly of the aligned reads should show polymorphism between Zm*Ht2*+ and Zm*Ht2*-. Third, the gene model should have a predicted domain/function indicating a potential R gene role based on the existing literature. Lastly, the gene model should be expressed *in planta*. Two gene models met all four of these criteria.

The first, GRMZM2G144028, is a predicted protein kinase roughly 300 kb from the center of the *Ht2* mapping region. It was expressed in inoculated leaf tissue, with a 1.60 – 4.19x higher baseline expression level in Zm*Ht2*- than Zm*Ht2*+. The assembled transcript sequences were polymorphic between Zm*Ht2*- and Zm*Ht2*+, with the Zm*Ht2*- allele having 100% sequence identity to the B73 reference allele and homology to OsWAK10d, a rice wall-associated kinase-like cytoplasmic kinase (Zhang et al. 2005). Upon infection, it was significantly upregulated in both hosts to a relatively low degree (1.06 – 3.52x higher than mock-inoculated, Fig. 2.11a).

The second, GRMZM2G316907, is a predicted LRR-protein kinase gene, also roughly 300 kb from the center of the *Ht2* fine-mapping region. The assembled transcripts were also polymorphic between Zm*Ht2*+ and Zm*Ht2*-, with 5 non-synonymous mutations. Baseline (mock-inoculated) expression was significantly higher in Zm*Ht2*+ across all timepoints (4.5 – 25x higher than Zm*Ht2*-). Upon inoculation, the gene was significantly induced at host/pathogen/timepoints except for 5 DPI in Zm*Ht2*+ (Fig. 2.11b), with more extreme upregulation in Zm*Ht2*- bringing the post-inoculation expression levels to a similar point in both hosts.

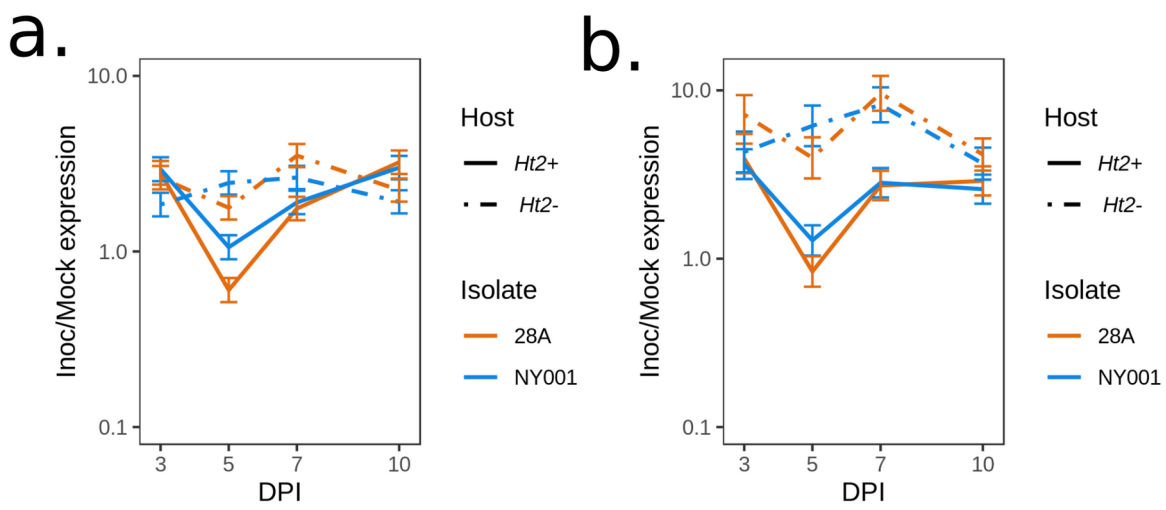


Figure 2.11: *In planta* expression patterns (fold change of expression in inoculated tissue relative to mock-inoculated expression) of GRMZM2G144028 (a) and GRMZM2G316907 (b) across different host/isolate/timepoint combinations.

Mining the unmapped reads, which could potentially contain sequence from *Ht2* that did not align to the B73 (AGPv3.31) reference genome, did not produce any usable candidate sequences. *De novo* assembly of unmapped reads from the non-inoculated maize libraries yielded 87,118 unique sequences from Zm*Ht2*+ and 154,942 from Zm*Ht2*-. Of the unique Zm*Ht2*+ sequences, 10,171 were unique to Zm*Ht2*+. After translating these sequences in all potential reading frames and comparing the amino acid sequence to the Pfam hidden Markov models for NBS and LRR domains, we identified 44 de novo assembled, Zm*Ht2*+ specific sequences containing a putative NBS domain and 51 containing a putative LRR domain. No sequences contained both a predicted NBS and predicted LRR domain. BLASTing these putative NBS/LRR sequences to maize GBS tags that had been genetically mapped

yielded none that matched a tag that had been mapped in or near the *Ht2* fine-mapping region. BLASTing the sequences of the GBS tags used to map *Ht2* in the DK888/S11 family to these *de novo*-assembled sequences also yielded no hits. Thus, these sequences could potentially contain sequence of the *Ht2* transcript, but their position in the maize genome remains unknown.

AVRHt2 candidate genes

We similarly attempted to identify one or more *AVRHt2* candidate genes using the available data. For each of the 241 gene models on an StNY001 scaffold associated with *AVRHt2* (Mideros et al. 2018), we noted its expression *in planta*, presence/absence in St28A, polymorphism between the two isolates, and whether or not it was a predicted NRPS/PKS/SSP encoding gene (Supplementary Table S2.5).

Of the twelve predicted StNY001 SSP encoding genes on the *AVRHt2*-associated scaffolds, seven were identical to the St28A gene model and four were predicted only in StNY001 but had 100% sequence identity to genomic sequence in St28A, likely attributable to technical issues associated with automated gene calling. Only one of the predicted NRPS/PKS/SSP encoding genes was deemed a plausible candidate: a predicted SSP encoding gene unique to StNY001 (protein ID 517350). StNY001 protein ID 517350 on scaffold 116 was unique to StNY001, but not significantly DE *in planta*.

All other gene models were either not expressed *in planta*, not polymorphic between the two isolates, or not predicted to have a domain or function associated with virulence. A predicted NRPS/PKS-like protein encoding gene (StNY001 protein ID 515003, St28A protein ID 174337) was not polymorphic between the two isolates. A predicted NRPS encoding gene on StNY001 scaffold 88 (StNY001 protein ID 607382, St28A protein ID 99181) was polymorphic between the two isolates, but

not expressed at any timepoint, including axenic tissue. Thirty-two gene models on an StNY001-associated scaffold were upregulated *in planta* and either specific to StNY001 or polymorphic between the two isolates, but none of these were predicted PKS/NRPS/SSP genes or contained a predicted protein domain otherwise associated with a known virulence factor. These genes of unknown function are prime targets for investigation as potential virulence determinants.

Ambiguous mapping and genome structure complicated the search for *AVRHt2*. While *AVRHt1* mapped neatly to a single marker in the middle of a scaffold, *AVRHt2* mapped to seven different small StNY001 scaffolds (Mideros et al. 2018). Many of these scaffolds were syntenic to one another, further complicating the matter. For example, parts of StNY001 scaffold 219 were syntenic to three StNY001 scaffolds- 232, 245, and 281. Several genetic markers statistically associated with the *AVRHt2*-mediated resistance mapped to these syntenic regions, and thus it was unclear which scaffolds were truly genetically linked to *AVRHt2* and which were falsely associated by multiple alignments of the marker.

DISCUSSION

Biotrophy-necrotrophy transition

Gene expression patterns in both host and pathogen were relatively static throughout the biotrophic phase of the interaction. The subsequent biotrophy-necrotrophy transition was marked by sudden and dramatic changes in gene expression. The major gene classes observed during host response, as defined by significant enrichment of Pfam domains among genes upregulated during pathogenesis, could be divided into three groups: biotrophy-specific, necrotrophy-specific, and observed throughout pathogenesis.

In the pathogen, we did not see clear patterns of different gene classes being strongly upregulated at different times, despite the fact that global gene expression changed drastically between 7 and 10 DPI. Pfam domains that were enriched among upregulated pathogen genes tended to be enriched either only in a single combination of host/pathogen/timepoint combination or across all combinations, rather than only during the biotrophic or necrotrophic phase as observed for many Pfam domains in the host. This was not driven by large differences in the richness of annotation of the two genomes: 62.7% of *S. turcica* gene models had at least one predicted Pfam domain, with an average of 1.38 predicted domains per gene model, compared to 74.5% of maize gene models with at least one predicted domain, with an average of 1.45 predicted domains per gene model.

In the plant, biotrophy-specific gene classes were mostly related to pathogen detection. Several domains characteristic of NLR genes were strongly enriched among upregulated host genes exclusively between 3 and 7 DPI. This observation is consistent with the emergent understanding that plants mount a wave of multiple recognition processes in response to infection, with PAMP detection inducing expression of NLR proteins (Lai and Eulgem 2018). The upregulation of maize mini-chromosome maintenance (MCM) genes during early biotrophy was unexpected, as this gene class is not normally associated with the plant defense response. MCM proteins form part of the initiation complex for DNA replication (Tye 1999). Though such genes play a role in diseases for which DNA replication is critical to pathogenesis, e.g. influenza in humans (Forsburg 2004) and cyst nematode infection in *Arabidopsis* (Huang et al. 2003), there has only been a single noted instance of a plant upregulating an MCM gene in response to infection by an oomycete (Wong et al. 2014) and none that we are aware of in response to a fungal pathogen. Plant MCM genes have been associated, through mostly unknown mechanisms, with salinity tolerance and low nitrogen stress response (Tuteja et al. 2011; Yang et al. 2013), so this

upregulation might simply be part of a general stress response. A more adventurous hypothesis is that MCM genes play a role in DNA replication for the purpose of extracellular DNA traps. These have so far been observed only in root tip exudates (Hawes et al. 2011), but deletion of a single extracellular DNAse gene from the maize foliar pathogen *Cochliobolus heterostrophus* drastically reduced its virulence (Park et al. 2019), suggesting that DNA-mediated resistance mechanisms are not limited solely to roots.

The genes involved in host response to necrotrophy mostly pertained to chemical warfare. Glutathione S-transferases (GSTs) detoxify xenobiotic compounds and have a known role in disease resistance in multiple pathosystems (Wisser et al. 2011). Three domains enriched among upregulated host genes during necrotrophy suggested a key role for vesicle trafficking: adaptins, a major coat protein domain (Boehm and Bonifacino 2001); Sec23/Sec24, the core components of COPII coat structures (Fromme et al. 2008); and Arf proteins, which play multiple roles in vesicle formation and membrane trafficking (D'Souza-Schorey and Chavrier 2006).

Many Pfam domains were enriched among upregulated host genes, but relatively few among downregulated host genes. This suggested that while numerous distinct pathways were induced upon infection, the genes and pathways that were downregulated were more general. Only four were enriched among at least four host-pathogen-DPI combinations, all at 7 or 10 DPI: chlorophyll A-B binding protein (PF00504), photosystem II subunit b (PF01789), CCT motif (PF06203), and CAAD domains of cyanobacterial aminoacyl-tRNA synthetase (PF14159), PsbP (PF01789).

Resistant and susceptible interactions

The differences between the resistant reaction and susceptible reactions are only interpretable in the context of the pathogen's lifestyle phases, with the starkest differences between resistance and susceptibility observed early on. At 3 DPI, the genes upregulated by the S hosts were mostly a subset of the much larger set of genes upregulated by the R host. Furthermore, those genes upregulated by both R and S hosts at 3 DPI were more strongly upregulated in the R host than the S host. Very few maize genes responded to infection in an opposite manner in R and S interactions (upregulated in Zm*Ht2*+ and downregulated in Zm*Ht2*-, or vice versa). By the time that the pathogen had begun its necrotrophic phase in the S host/pathogen interactions, it was still arrested in biotrophy in the R interaction. The resistant host transcriptome at 10 DPI more closely resembled the S transcriptomes at 7 DPI (biotrophy) than the S transcriptomes at 10 DPI (necrotrophy). The R reaction did show some hallmarks of necrotrophy, such as the formation of necrotic lesions, a twofold increase in fungal biomass, and the enrichment of necrotrophy-associated genes- GSTs, adaptins, etc.

Fungal genome structure

The two *S. turcica* isolates had similar genomes, leaving open the question of why StNY001 is less aggressive than St28A. The difference in aggressiveness was clear: even in the absence of any R gene, StNY001 accumulated only about half as much biomass by 10 DPI as St28A did. No QTL mediating aggressiveness were discovered in a mapping population derived from StNY001 and St52B, another race 23N isolate, but the radial growth QTL colocalized with the *AVRHt1* gene (Mideros et al. 2018). These results suggest that the absence of a fully functional *AVRHt1* gene present on race 1 strains such as StNY001 results in slower grown in-vitro and diminished aggressiveness in planta.

Our results supported the hypothesis of Dong et al. (2015) that gene-sparse regions of fungal pathogen genomes are “hotspots” for pathogenesis-related genes. Gene-sparse regions in both *S. turcica* genomes tended to contain genes that were highly upregulated *in planta*, as well as more predicted SSP, NRPS, and PKS genes. However, the dN/dS rate for genes in these regions was not any higher than the rate in gene-rich regions. This contradicted a core component of the above hypothesis, that pathogenesis-related genes can evolve faster in gene-sparse regions due to a greater rate of polymorphisms there.

Benzoxazinoids

Different maize benzoxazinoid compounds have distinct roles in defense against insect pests, with some defending more against phloem-feeding insects and others more against leaf-chewing insects (Zhou et al. 2018). This invites interesting comparison between maize defense against “parasitic” threats– biotrophic fungi and phloem feeding- and defense against “destructive” threats- necrotrophic fungi and leaf chewing. We found strong parallels between maize benzoxazinoid-related genes expressed during the two pathogen lifestyle phases and their known roles against different types of insect pests.

The only *Bx* gene significantly upregulated at 3 DPI was *Bx14*, which encodes an O-methyltransferase that converts DIM2BOA to HDM2BOA. Though the differential effects of HDM2BOA and DIM2BOA are not yet known, both of these compounds negatively affect phloem-feeding insect herbivores, but not chewing herbivores (Handrick et al. 2016). The *Bx10/11/12* genes were strongly upregulated as pathogenesis went on, suggesting a potential necrotrophy-specific role HDMBOA. Many of the genes in the core benzoxazinoid biosynthesis pathway were downregulated

over time, though whether this was due to a negative feedback loop as proposed by Ahmad et al. (2011) or due to inhibition of these genes by the pathogen cannot be determined from these data.

Candidate *R/AVR* genes

Because *Ht2* was mapped to a region of the maize genome containing many protein kinase genes of various types, we sought to narrow down the list of candidates using expression data and the RNA sequences themselves. Two maize gene models were deemed strong candidate genes for *Ht2* via multiple criteria. The first, GRMZM2G144028, is a protein kinase roughly 300 kb from the center of the *Ht2* mapping region, was expressed in leaf tissue and induced by NLB infection in both Zm*Ht2*+ and Zm*Ht2*-. It is highly polymorphic between Zm*Ht2*- and Zm*Ht2*+, with the ZmHt2- allele having 100% sequence identity to the B73 reference allele and homology to OsWAK10d, a rice wall-associated kinase-like cytoplasmic kinase (Zhang et al. 2005). The second, GRMZM2G316907, is an LRR-protein kinase, also roughly 300 kb from the center of the *Ht2* fine-mapping region. The assembled Zm*Ht2*+ and Zm*Ht2*- transcript sequences were polymorphic, baseline expression was higher in Zm*Ht2*+, and the gene was induced by *S. turcica* infection.

We did not find a promising candidate for *AVRHt2*. The best candidate was StNY001 protein ID 517350, a predicted SSP unique to StNY001, but this was expressed at a fairly low level, not significantly DE *in planta* (in stark contrast to *AVRHt1*), and located on a scaffold only loosely associated to *AVRHt2*. One major complication was that *AVRHt2* mapped to numerous small scaffolds (Mideros et al. 2018), most of which were syntenic to one another. Given this, we believe that *AVRHt2* is located in a repeat-rich, gene-poor region of the StNY001 genome that will be difficult to assemble correctly.

MATERIALS AND METHODS

Host/pathogen lines

Two maize near-isogenic lines, one carrying the *Ht2* resistance gene (Zm*Ht2*+) and one lacking *Ht2* (Zm*Ht2*-), were each inoculated with two fungal strains, which either induce a resistance reaction on maize carrying the *Ht2* resistance gene (*S. turcica* strain NY001 carrying *AVRHt2*) or do not induce a resistance reaction on such lines (*S. turcica* strain 28A carrying *avrHt2*). A summary of expected reactions is given in Table 2.1.

S. turcica strains St28A (race 23N, *MAT1-1*, Caldwell Co., KY, from M. Carson via P. Balint-Kurti), and StNY001 (race 1, *MAT1-1*, originally collected in New York by G. Bergstrom) were used. Genome sequence of strain 28A (https://genome.jgi.doe.gov/Settu3/Settu3.home.html) was reported earlier (Ohm et al., 2012; Condon et al., 2013). Genome sequence of strain NY001 (https://genome.jgi.doe.gov/Settur3/Settur3.home.html) was generated as part of the current study. Unless mentioned otherwise, all strains were grown on lactose casein agar (LCA) under 12 hr fluorescent light/12 dark at approximately 22 °C (Xue et al., 2013).

The two maize lines used were near-isogenic lines described by Chung *et al*. (2010). The two lines were isogenic at all locations except in the region around chromosome 8, bin 6 (bin 8.06), at which one line (08CF2008_004s) carried an introgression from maize line DK888 containing the *Ht2* resistance gene, while the other (08CF2008_027s) carried an introgression from maize line S11 lacking the *Ht2* resistance gene. These lines were referred to as Zm*Ht2*+ and Zm*Ht2*-.

Inoculation

Three-week-old maize lines (Zm*Ht2*+ and Zm*Ht2*-) were each inoculated with *S. turcica* strains NY001 (*AVRHt2*) and 28A (*avrHt2*). Prior to inoculation, cultures of *S. turcica* were grown on LCA for two weeks. Spores were dislodged by gently scraping the colony surface with a rubber policeman after application of 2 mL sterile deionized water containing 0.02% Tween 20. An aliquot was counted with a hemocytometer and spore solutions adjusted to 40,000 spores/mL. A modified airbrush sprayer (Preval) was used to spray two ml/plant. Mock-inoculated plants were sprayed with a 0.02% Tween 20 solution in sterile deionized water. Plants were then placed in a mist chamber overnight. At 3, 5, 7, and 10 days post-inoculation (DPI), samples were excised from the leaf into sterilized tubes, frozen in liquid nitrogen, and transferred to a -80C freezer.

RNA sequencing

Total RNA was extracted from frozen samples using TRIzol (Life Technologies) and purified with Qiagen RNeasy kit, as directed by the manufacturers. Quantity and quality were verified using a Nanodrop and Ribogreen. Samples were delivered to JGI for mRNA-Seq library construction and sequencing.

RNA sample prep was performed on the PerkinElmer Sciclone NGS robotic liquid handling system using Illumina TruSeq Stranded mRNA HT sample prep kit. The latter utilized poly-A selection of mRNA following the protocol outlined by Illumina (https://support.illumina.com/sequencing/sequencing_kits/truseq-stranded-mrna.html), and with the following conditions: total RNA starting material was 1 ug per sample and 10 PCR cycles were used for library amplification.

The prepared libraries were quantified using KAPA Biosystem's next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. The quantified libraries were then multiplexed with other libraries, and the pool of libraries was then prepared for sequencing on the Illumina HiSeq sequencing platform utilizing a TruSeq paired-end cluster kit, v4, and Illumina's cBot instrument to generate a clustered flow cell for sequencing. Sequencing of the flow cell was performed on the Illumina HiSeq 2500 sequencer using HiSeq TruSeq SBS sequencing kits, v4, following a 2x100 bp and 2x150 bp indexed recipes for the 300 bp and 4 kbp run, respectively.

Raw fastq file reads were filtered and trimmed using the JGI QC pipeline. In brief, using BBDuk, raw reads were evaluated for artifact sequences by kmer matching (kmer=25), allowing 1 mismatch and detected artifacts were trimmed from the 3' end of the reads. RNA spike-in reads, PhiX reads, and reads containing any missing calls ("N" calls) were removed. Quality trimming was

performed using the phred trimming method set at Q6. Finally, following trimming, reads under the length threshold were removed (minimum length 25 bases or 1/3 of the original read length - whichever is longer).

Filtered reads from each library were aligned to the combined reference transcriptomes using a tool developed at JGI utilizing BWA as the aligner with only unique mapping allowed. If a read mapped to more than one location, it was ignored. In-house tool was used to generate the raw gene counts. Raw gene counts were split one per reference. Raw gene counts were used to evaluate the level of correlation between biological replicates using Pearson's correlation and determine which replicates would be used in the DGE analysis.

NY001 DNA preparation

1000 mls of liquid LC medium were inoculated with 10^6-10^7 spores and shaken at 150-200 rpm at room temperature for two days. Mycelium was harvested on Whatman #4 filter paper using a Buchner funnel connected to a vacuum. The mycelium pad was peeled off the filter, placed in a 50 ml Falcon tube, and stored at -20ºC or directly lyophilized for 25 hrs (cap loose) or until mycelium is dry and brittle. Lyophilized mycelia were pulverized with a pre-chilled pestle in a pre-chilled mortar to a fine powder. Mycelial powder (~0.2-0.5g) was suspended in 20ml isolation buffer [150 mM EDTA pH 8.0, 50 mM Tris, pH 8.0, 1 % (w/v) sarkosyl (n-lauroyl sarcosine), 300 mg/L Protease XI (Proteinase K)] in disposable 30ml polypropylene tubes (Sarstedt) and vortexed vigorously for 30-60 sec. Cell debris was pelleted by centrifugation in an SS-34 rotor (5 min, 5000 rpm, 4ºC) and supernatant transferred to a clean tube. 20ml isolation buffer was added and centrifugation repeated. The recovered supernatant was gently mixed with an equal volume of Tris-saturated phenol and tubes centrifuged (10

min, 5000 rpm, 4ºC) to separate phases. The upper (aqueous) phase was transferred to a clean tube and mixed with an equal volume of 25:24:1 phenol:chloroform:isoamyl alcohol. Centrifugation and recovery of the top aqueous layer was repeated twice. Finally the aqueous layers were mixed with an equal volume of 24:1 chloroform:isoamyl alcohol and centrifuged as before. DNA was extracted by ethanol precipitation (1/10 vol. of 3M NaOAc plus, 2 vol. of cold absolute ethanol) for 10 min at -20C and centrifuged as above. The pellet was transferred to a 1.5 ml tube, washed with 70% ethanol twice, centrifuged (5 min, 5K rpm, 4ºC) and air dried. The DNA was dissolved in 2x 200ul TE (65ºC 10-20 min to dissolve, check frequently, not longer than 30 min), 2 ul RNase A (10ug/ul, Sigma R-6513) was added to each tube and incubated at 37ºC for 1 hr. A l.5ul aliquot was run on 0.7% gel to check if the RNA is still present.

NY001 genome sequencing and assembly

Two libraries were made, one with an insert size of 300 bp and the other 4 kbp. For the 300 bp library, 100 ng of DNA was sheared to 300 bp using the Covaris LE220 and size selected using SPRI beads (Beckman Coulter). The fragments were treated with end-repair, A-tailing, and ligation of Illumina compatible adapters (IDT, Inc) using the KAPA-Illumina library creation kit (KAPA Biosystems). The prepared libraries were quantified, multiplexed, and sequenced on HiSeq 2500 as described in the previous section but using 2x150bp indexed recipe.

For the 4 kb library, 5 µg of DNA was sheared using the Covaris g-TUBE™ (Covaris) and gel size selected for 4 kb. The sheared DNA was treated with end repair and ligated with biotinylated adapters containing loxP. The adapter ligated DNA fragments were circularized via recombination by a Cre excision reaction (NEB). The circularized DNA templates were then randomly sheared using the

Covaris LE220 (Covaris). The sheared fragments were treated with end repair and A-tailing using the KAPA-Illumina library creation kit (KAPA biosystems) followed by immobilization of mate pair fragments on strepavidin beads (Invitrogen). Illumina compatible adapters (IDT, Inc) were ligated to the mate pair fragments and 8 cycles of PCR was used to enrich for the final library (KAPA Biosystems).

Each fastq file was QC filtered for artifact/process contamination and subsequently assembled together with AllPathsLG version R49403, (Gnerre et al., 2010), to produce a 93.8 X coverage main assembly. Mitochondrial sequence was assembled separately with AllPathsLG version R49403 to produce a 90.3X coverage mitochondrial assembly with 2 scaffolds.

The genome was annotated using the JGI Annotation pipeline (Grigoriev et al, NAR 2014). Genome assembly and annotation have been deposited in DDBJ/ENA/GenBank under the accession number (TO BE PROVIDED UPON PUBLICATION). The NY001 genome is available on the JGI-DOE MycoCosm portal (Grigoriev et al., 2014; https://mycocosm.jgi.doe.gov/Settur3/Settur3.home.html).

Alignment of reads

Reads were aligned to concatenated maize and *S. turcica* FASTA/GFF files and counted with STAR (Dobin 2013). The maize AGPv3.31 B73 reference genome (Schnable et al. 2009) was used, as functional annotation of the maize AGPv4 genome (Jiao et al. 2017) was still ongoing. For *S. turcica*, the St28Av2.0 and StNY001v2.0 genomes were sequenced, assembled, and annotated by JGI as described above. Transcript counts were then tested for differential expression between inoculated tissue and the relevant control library (mock-inoculated maize or axenic *S. turcica*) using DESeq2

(Love et al. 2014). Differential expression was considered significant if $p < 0.01$ after multiple test correction.

GO term enrichment among significantly differentially expressed genes was calculated using TopGO using the weight01 algorithm, a combination of the weight and elim algorithms described by Alexa *et al.* (2006). Enriched terms were calculated separately for each timepoint for each host/pathogen combination. Enrichment of Pfam domains among up- or downregulated genes was tested with a Fisher's exact test and p-values were adjusted via the Benjamini-Hochberg False Discovery Rate (FDR) correction.

Unmapped reads from mock-inoculated maize libraries were *de novo* assembled using Trinity 2.2.0 (Grabherr et al. 2011). Each Trinity-assembled transcript was translated into all six reading frames using transeq from EMBOSS (Rice et al. 2000). Potential PFAM matches for these six translations were obtained using HMMer (Finn et al. 2011). Coverage over the *Ht2* fine-mapping region in the Ht2+/- lines was calculated using samtools (Li et al. 2009).

Gene-sparse regions

For each *S. turcica* gene, we counted the number of genes within 25 kb from the beginning and end of the transcript. Genes that were less than 25 kb from the end of a scaffold were included. If a gene was the last gene on a scaffold, the distance to the end of the scaffold was substituted for the distance to the nearest gene in that direction. If a given gene overlapped with another, the distance was taken as 0, rather than a negative value. We then examined the distribution of relative expression between mock-inoculated and inoculated tissue (the log-2 fold change) for groups with a given number of neighboring genes in a 50 kb window.

As the distribution of log-transformed distance to neighboring gene had high kurtosis, and the distribution of the number of neighboring genes had strong negative skew, a Wilcoxon rank-sum test was used to test the hypothesis that these values were similarly distributed in predicted SSP or SM genes and the rest of the genome. Mean expression changes are given as arithmetic means of the log2-fold changes, i.e. geometric means of the fold changes themselves.

S. turcica gene models

Gene models were considered common to both *S. turcica* isolates if the gene models were reciprocal best hits (RBHs). Predicted protein models were BLASTed all-vs-all with blastp default settings, and a custom Python script was used to find RBHs for which the BLAST E-value in both directions was less than 1e-10.

SSP effector genes were predicted using EffectorP (Sperschneider et al. 2016). All gene models with a secretion signal predicted by SignalP (Petersen et al. 2011) were evaluated with EffectorP. As a control, these predicted SSPs were compared to SSPs predicted by the criteria of Ohm et al. (2012), under which a gene was classified if its predicted protein product was fewer than 200 amino acids in length, contained a secretion signal as determined by SignalP, and had no transmembrane domain as determined by TMHMM v2.0 (http://www.cbs.dtu.dk/services/TMHMM/), unless it appeared in the first 40 amino acids. PKS-encoding genes were predicted using antiSMASH web platform, fungal version (Weber et al. 2015). NRPS-encoding genes were predicted using the method described in Bushley and Turgeon (2010).

Estimating dN/dS

Rates of dN/dS were estimated between all St28A and StNY001 gene models and the gene models from five other pathogenic fungi in the order Pleosporales- *Cochliobolus carbonum*, *Cochliobolus heterostrophus*, *Didymella zeae-maydis*, *Leptosphaeria maculans*, and *Stagonospora nodorum*. Predicted protein sequences were searched all-vs-all using blastp in BLAST+ v2.6.0 on default settings (Camacho et al. 2009). Orthogroups were predicted with OrthoMCL (Li et al. 2003) using an inflation cutoff of 4, percent-value cutoff of 1e-5 and, and pi cutoff of 80. Sequences from orthogroups for which there were no putative paralogs (i.e. at most one gene model from each species) were extracted with a custom Python script and aligned with MAFFT v7.312 (Katoh and Standley 2013). Multiple codon alignments were made with Pal2Nal v14 (Suyama et al. 2006), and rates of dN and dS were estimated with the method of Yang and Nielsen (2000) using PAML v4.9 (Yang 2007).

Diversity of maize gene expression (282 TWAS)

The Goodman-Buckler maize diversity panel, consisting of 282 lines from various subpopulations, has been evaluated for NLB resistance (Wisser et al. 2011), and gene expression has been quantified in multiple tissue types for most lines by 3’ RNA-seq (Kremling et al. 2017). Using ridge regression and 10-fold cross-validation, we predicted NLB phenotypes from the expression of NLB-responsive genes, as well as from the expression of randomly sampled sets of genes as a baseline.

Transformed transcript counts from non-inoculated tissue of the middle of leaf 3 (L3mid) and mature adult leaves at daytime (LMAD) were taken from Kremling et al. (2018). Best linear unbiased predictions (BLUPs) for NLB resistance in the field were obtained from Judith Kolkman (personal communication). We then tested for significant associations between gene expression in L3mid and

NLB phenotype. To determine an appropriate $p < 0.05$ cutoff, we performed 1000 bootstrap iterations, in which we randomized phenotypes and found the most significantly associated gene by t-test. The empirical p-value cutoff was afterwards set at $p = 1.4e-6$ for L3mid.

LITERATURE CITED

Ahmad, S., Veyrat, N., Gordon-Weeks, R., Zhang, Y. Martin, J., Smart, L., Glauser, G., Erb, M., Flors, V., Frey, M., Ton, J. 2011. Benzoxazinoid metabolites regulate innate immunity against aphids and fungi in maize. Plant Physiol. 157:317-327.

Boehm, M. and Bonifacino, J. S. 2001. Adaptins: the final recount. Mol. Biol. Cell 12:780-794.

Brunner, C., Torriani, S. F. F., Croll, D., Stukenbrock, E. H., and McDonald, B. A. 2013. Coevolution and life cycle specialization of plant cell wall degrading enzymes in a hemibiotrophic pathogen. Mol. Biol. Evol. 30:1337-1347.

Bushley, K.E. and Turgeon, B.G. 2010. Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. BMC Evol. Biol. 10:26.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. 2009. BLAST+: architecture and applications. BMC Bioinf. 10:421.

Chung, C.L., Jamann, T., Longfellow, J. and Nelson, R. 2010. Characterization and fine-mapping of a resistance locus for northern leaf blight in maize bin 8.06. Theor. Appl. Genet. 121:205-227.

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

Condon, B.J., Leng, Y., Wu, D., Bushley, K.E., Ohm, R.A., Otillar, R., Martin, J., Schackwitz, W., Grimwood, J., MohdZainudin, N., Xue, C., Wang, R., Manning, V.A., Dhillon, B., Tu, Z.J., Steffenson, B.J., Salamov, A., Sun, H., Lowry, S., LaButti, K., Han, J., Copeland, A., Lindquist, E., Barry, K., Schmutz, J., Baker, S.E., Ciuffetti., L.M., Grigoriev., I.V., Zhong., S., and

Turgeon, B.G. 2013. Comparative genome structure, secondary metabolite, and effector coding capacity across *Cochliobolus* pathogens. PLoS Genet. 9: p.e1003233.

Czjzek, M., Cicek, M., Zamboni, V., Bevan, D.R., Henrissat, B. and Esen, A. 2000. The mechanism of substrate (aglycone) specificity in β-glucosidases is revealed by crystal structures of mutant maize β-glucosidase-DIMBOA,-DIMBOAGlc, and-dhurrin complexes. Proc. Nat. Acad. Sci. 97:13555-13560.

Dallery, J.F., Lapalu, N., Zampounis, A., Pigné, S., Luyten, I., Amselem, J., Wittenberg, A.H., Zhou, S., De Queiroz, M.V., Robin, G.P. and Auger, A. 2017. Gapless genome assembly of *Colletotrichum higginsianum* reveals chromosome structure and association of transposable elements with secondary metabolite gene clusters. BMC Genom. 18:667.

DeYoung, B.J. and Innes, R.W. 2006. Plant NBS-LRR proteins in pathogen sensing and host defense. Nat. Immunol. 7:1243-1249.

Dong, S., Raffaele, S., and Kamous, S. 2015. The two-speed genomes of filamentous pathogens: waltz with plants. Curr. Opin. Genet. Devel. 35:57-65.

D'Souza-Schorey, C. and Chavrier, P. 2006. ARF proteins: roles in membrane traffic and beyond. Nat. Rev. Mol. Cell Biol. 7:347-358.

Finn, R.D., Clements, J. and Eddy, S.R. 2011. HMMER web server: interactive sequence similarity searching. Nucl. Acids Res. 39:W29-W37.

Forsburg, S.L. 2004. Eukaryotic MCM proteins: beyond replication initiation. Microbiol. Molec. Biology Rev. 68:109-131.

Fromme, J.C., Orci, L. and Schekman, R. 2008. Coordination of COPII vesicle trafficking by Sec23. Trends Cell Biol. 18:330-336.

Gan, P., Ikeda, K., Irieda, H., Narusaka, M., O'Connell, R. J., Narusaka, Y., Takano, Y., Kubo, Y., and Shirasu, K. 2013. Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. New Phytol. 197:1236-1249.

Gharbi, Y., Barkallah, M., Bouazizi, E., Gdoura, R. and Triki, M.A. 2017. Differential biochemical and physiological responses of two olive cultivars differing by their susceptibility to the hemibiotrophic pathogen *Verticillium dahliae*. Physiol. Molec. Plant Pathol. 97:30-39.

Gnerre, S., Maccallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., Sharpe, T., Hall, G., Shea, T.P., Sykes, S., Berlin, A.M., Aird, D., Costello, M., Daza, R., Williams, L., Nicol, R., Gnirke, A., Nusbaum, C., Lander, E.S., and Jaffe, D.B. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc. Natl. Acad. Sci. USA 108(4):1513-1518.

Gómez-Anduro, G., Ceniceros-Ojeda, E.A., Casados-Vázquez, L.E., Bencivenni, C., Sierra-Beltrán, A., Murillo-Amador, B. and Tiessen, A. 2011. Genome-wide analysis of the beta-glucosidase gene family in maize (*Zea mays* L. var B73). Plant Mol. Biol. 77:159-183.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q. and Chen, Z. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature Biotech. 29:644-652.

Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., and Shabalov, I. 2014. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. 42:D699-704.

Handrick, V., Robert, C.A., Ahern, K.R., Zhou, S., Machado, R.A., Maag, D., Glauser, G., Fernandez-Penny, F.E., Chandran, J.N., Rodgers-Melnick, E. and Schneider, B. 2016. Biosynthesis of 8-O-methylated benzoxazinoid defense compounds in maize. Plant Cell: tpc-00065.

Han, Y., Zhang, K., Yang, J., Zhang, N., Fang, A., Zhang, Y., Liu, Y., Chen, Z., Hsiang, T. and Sun, W. 2015. Differential expression profiling of the early response to *Ustilaginoidea virens* between false smut resistant and susceptible rice varieties. BMC Genom. 16:955.

Hawes, M.C., Curlango-Rivera, G., Wen, F., White, G.J., VanEtten, H.D. and Xiong, Z. 2011. Extracellular DNA: the tip of root defenses? Plant Sci. 180:741-745.

Huang, X., Springer, P.S. and Kaloshian, I. 2003. Expression of the Arabidopsis MCM gene PROLIFERA during root-knot and cyst nematode infection. Phytopathol. 93:35-41.

Jamann, T. 2013. Insights into disease resistance: Genetic architecture, genes, and pleiotropy in maize. Ph.D. dissertation, Cornell University.

Katoh, K. and Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molec. Biol. Evol. 30:772-780.

Kawahara, Y., Oono, Y., Kanamori, H., Matsumoto, T., Itoh, T. and Minami, E. 2012. Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. PloS One 7:p.e49423.

Kremling, K.A., Chen, S.Y., Su, M.H., Lepak, N.K., Romay, M.C., Swarts, K.L., Lu, F., Lorant, A., Bradbury, P.J. and Buckler, E.S. 2018. Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. Nature 555:520.

Hurni, S., Scheuermann, D., Krattinger, S.G., Kessel, B., Wicker, T., Herren, G., Fitze, M.N., Breen, J., Presterl, T., Ouzunova, M. and Keller, B. 2015. The maize disease resistance gene *Htn1* against

northern corn leaf blight encodes a wall-associated receptor-like kinase. Proc. Nat. Acad. Sci.:p.201502522.

Lai, Y. and Eulgem, T. 2018. Transcript level expression control of plant NLR genes. Mol. Plant Pathyol. 19:1267-1281.

Li, L., Stoeckert, C.J. and Roos, D.S. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res.:13:2178-2189.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079.

Long, B.J., Dunn, G.M. and Routley, D.G. 1975. Relationship of hydroxamic acid content in maize and resistance to northern corn leaf blight 1. Crop Sci. 15:333-335.

Long, B.J., Dunn, G.M. and Routley, D.G. 1978. Relationship of hydroxamate concentration in maize and field reaction to *Helminthosporium turcicum* 1. Crop Sci. 18:573-575.

Love, M.I., Huber, W. and Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:550.

Matić, S., Bagnaresi, P., Biselli, C., Carneiro, G.A., Siciliano, I., Valé, G., Gullino, M.L. and Spadaro, D. 2016. Comparative transcriptome profiling of resistant and susceptible rice genotypes in response to the seedborne pathogen *Fusarium fujikuroi*. BMC Genom. 17:608.

Meihls, L.N., Handrick, V., Glauser, G., Barbier, H., Kaur, H., Haribal, M.M., Lipka, A.E., Gershenzon, J., Buckler, E.S., Erb, M. and Köllner, T.G. 2013. Natural variation in maize aphid resistance is associated with 2, 4-dihydroxy-7-methoxy-1, 4-benzoxazin-3-one glucoside methyltransferase activity. The Plant Cell 25:2341-2355.

Meinhardt, L. W., Costa, G. G. L., Thomazella, D. P. T., Teixeira, P. J. P. L., Carazzolle, M. F., Schuster, S. C., Carlson, J. E., Guiltinan, M. J., Mieczkowski, P., Farmer, A., Ramaraj, T., Crozier, J., Davis, R. E., Shao, J., Melnick, R. L., Pereira, G. A. G., and Bailey, B. A. 2014. Genome and secretome analysis of the hemibiotrophic fungal pathogen, *Moniliophthora roreri*, which causes frosty pod rot disease of cacao: mechanisms of the biotrophic and necrotrophic phases. BMC Genom. 15:164.

Mideros, S. X., Chung, C. L., Wiesner-Hanks, T., Poland, J. A., Wu, D., Fialko, A. A., Turgeon, B. G., and Nelson, R. J. 2018. Determinants of virulence and in vitro development colocalize on a genetic map of *Setosphaeria turcica*. Phytopathology 108:254-263.

Mueller, D. and Wise, K. 2014. Corn disease loss estimates from the United States and Ontario, Canada —2013. Purdue Extension publication BP-96-13-W.

Niemeyer, H.M. 2009. Hydroxamic acids derived from 2-hydroxy-2 H-1, 4-benzoxazin-3 (4 H)-one: key defense chemicals of cereals. Journal Agric. Food Chem. 57:1677-1696.

O'Connell, R.J., Thon, M.R., Hacquard, S., Amyotte, S.G., Kleemann, J., Torres, M.F., Damm, U., Buiate, E.A., Epstein, L., Alkan, N. and Altmüller, J. 2012. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. Nat. Genet. 44:1060-1067.

Ohm, R.A., Feau, N., Henrissat, B., Schoch, C.L., Horwitz, B.A., Barry, K.W., Condon, B.J., Copeland, A.C., Dhillon, B., Glaser, F., Hesse, C.N., Kosti, I., LaButti, K., Lindquist, E.A., Lucas, S., Salamov, A.A., Bradshaw, R.E., Ciuffetti, L., Hamelin, R.C., Kema, G.H.J., Lawrence, C., Scott, J.A., Spatafora, J.W., Turgeon, B.G., deWit, P.J.G.M., Zhong, S., Goodwin, S.B., and

Grigoriev, I.V. 2012. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. PLoS Path. 8:p.e1003037.

Palma-Guerrero, J., Torriani, S. F. F., Zala, M., Carter, D., Courbot, M., Rudd, J. J., McDonald, B. A., and Croll, D. 2016. Comparative transcriptomic analyses of *Zymoseptoria tritici* strains show compWinterlex lifestyle transitions and intraspecific variability in transcription profiles. Mol. Plant Pathol. 17:845-859.

Park, H.J., Wang, W., Curlango-Rivera, G., Xiong, Z., Lin, Z., Huskey, D.A., Hawes, M.C., VanEtten, H.D., and Turgeon, B.G. 2019. A DNase from a fungal phytopathogen is a virulence factor likely deployed as counter defense against host-secreted extracellular DNA. mBio 10(2): e02805-18.

Peng, Z., Oliveira-Garcia, E., Lin, G., Hu, Y., Dalby, M., Migeon, P., Tang, H., Farman, M., Cook, D., White, F.F. and Valent, B. 2018. Effector gene reshuffling involves dispensable mini-chromosomes in the wheat blast fungus. BioRxiv:359455.

Perfect, S.E., and Green, J.R. 2008. Infection structures of biotrophic and hemibiotrophic fungal plant pathogens. Mol. Plant Pathol. 2(2):101-108.

Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8:785-786.

Poland, J.A., Bradbury, P.J., Buckler, E.S. and Nelson, R.J. 2011. Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proc. Nat. Acad. Sci.:p.201010894.

Rice, P., Longden, I. and Bleasby, A. 2000. EMBOSS: the European molecular biology open software suite. Trends in Genet. 16:276-277.

Rouxel, T., Grandaubert, J., Hane, J.K., Hoede, C., Van de Wouw, A.P., Couloux, A., Dominguez, V., Anthouard, V., Bally, P., Bourras, S., Cozijnsen, A.J., Ciuffetti, L.M., DeGrave, A., Dilmaghani, A., Duret, L., Fudal, I., Goodwin, S.B., Gout, L., Glaser, N., Linglin, J., Kema, G.H.J., Lapalu, N., Lawrence, C.B., May, K., Meyer, M., Olliver, B., Poulain, J., Schoch, C.L., Simon, A., Spatafora, J.W., Stachowiak, A., Turgeon, B.G., Tyler, B.M., Vincent, D., Weissenbach, J., Amselem, J., Quesneville., H., Oliver, R.P., Wincker, P., Balesdent, M.H., and Howlett, B.J. 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. Nat. Commun. 2:202.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A. and Minx, P. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112-1115.

Sonah, H., Zhang, X., Deshmukh, R.K., Borhan, M.H., Fernando, W.G. and Bélanger, R.R. 2016. Comparative transcriptomic analysis of virulence factors in *Leptosphaeria maculans* during compatible and incompatible interactions with canola. Front. Plant Sci. 7:1784.

Sperschneider, J., Gardiner, D.M., Dodds, P.N., Tini, F., Covarelli, L., Singh, K.B., Manners, J.M. and Taylor, J.M. 2016. EffectorP: predicting fungal effector proteins from secretomes using machine learning. New Phytol. 210:743-761.

Suyama, M., Torrents, D. and Bork, P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucl. Acids Res. 34:W609-W612.

Tuteja, N., Tran, N.Q., Dang, H.Q. and Tuteja, R. 2011. Plant MCM proteins: role in DNA replication and beyond. Plant Molec. Biol. 77:537-545.

Tye, B.K. 1999. MCM proteins in DNA replication. Ann. Rev. Biochem. 68:649-686.

Wang, L., Li, Q., Liu, Z., Surendra, A., Pan, Y., Li, Y., Zaharia, L.I., Ouellet, T. and Fobert, P.R. 2018. Integrated transcriptome and hormone profiling highlight the role of multiple phytohormone pathways in wheat resistance against fusarium head blight. PloS One 13:p.e0207036.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R., Lee, S.Y., Fischbach, M.A., Müller, R., Wohlleben, W. and Breitling, R. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucl. Acids. Res. 43:W237-W243.

Winter, D.J., Ganley, A., Young, C., Liachko, I., Schardl, C., Dupont, P.Y., Berry, D., Ram, A., Scott, B. and Cox, M. 2018. Repeat elements organize 3D genome structure and mediate transcription in the filamentous fungus *Epichloë festucae*. BioRxiv:339010.

Wisser, R.J., Kolkman, J.M., Patzoldt, M.E., Holland, J.B., Yu, J., Krakowsky, M., Nelson, R.J. and Balint-Kurti, P.J. 2011. Multivariate analysis of maize disease resistances suggests a pleiotropic genetic basis and implicates a GST gene. Proc. Nat. Acad.Sci. 108:7339-7344.

Wong, J., Gao, L., Yang, Y., Zhai, J., Arikit, S., Yu, Y., Duan, S., Chan, V., Xiong, Q., Yan, J. and Li, S. 2014. Roles of small RNA s in soybean defense against *Phytophthora sojae* infection. The Plant J. 79:928-940.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molec. Biol. Evol. 24:1586-1591.

Yang, Z. and Nielsen, R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Molec. Biol. Evol. 17:32-43.

Yang, F., Li, W., and Jørgensen, H. J. L. 2013. Transcriptional reprogramming of wheat and the hemibiotrophic pathogen *Septoria tritici* during two phases of the compatible interaction. PLoS One 8:e81606.

Zhou, S., Richter, A., and Jander, G. 2018. Beyond defense: Multiple functions of benzoxazinoids in maize metabolism. Plant Cell Physiol. 59(8):1528-1537.

Supplementary Material

Supplementary Table S2.1: Ten most significantly differentially expressed *S. turcica* genes at each combination of host (R: *Ht2+*, S: *Ht2-*), pathogen isolate (NY001: *AVRHt2*, 28A: *avrHt2*), and timepoint (DPI: days post-inoculation), organized by predicted Pfam protein domain(s).

Plant	Isolate	DPI	Protein ID	Base mean expression	log2 Fold Change	p	Description
R	NY001	3	243127	1240	13.58	2.04E-166	Aspartic peptidase
S	NY001	3	243127	1240	13.35	1.58E-162	Aspartic peptidase
R	NY001	10	506629	46	7.74	4.07E-260	Aspartic peptidase
S	NY001	7	506629	46	7.33	7.86E-246	Aspartic peptidase
S	NY001	10	506629	46	7.03	4.33E-245	Aspartic peptidase
R	NY001	3	170288	324	7.57	1.15E-169	Chitin-binding, domain 3
R	NY001	5	170288	324	8.08	3.00E-301	Chitin-binding, domain 3
R	NY001	7	170288	324	8.16	0.00E+00	Chitin-binding, domain 3
R	NY001	10	170288	324	7.93	0.00E+00	Chitin-binding, domain 3
S	NY001	3	170288	324	7.83	9.31E-197	Chitin-binding, domain 3
S	NY001	5	170288	324	8.62	0.00E+00	Chitin-binding, domain 3
S	NY001	7	170288	324	8.41	0.00E+00	Chitin-binding, domain 3
S	NY001	10	170288	324	7.08	2.44E-272	Chitin-binding, domain 3
R	NY001	10	538563	52	6.39	6.04E-242	Galactosyl transferase
S	NY001	10	538563	52	5.48	5.42E-215	Galactosyl transferase
R	NY001	3	436826	3079	13.33	2.18E-218	Haem peroxidase
R	NY001	5	436826	3079	13.91	9.54E-244	Haem peroxidase
R	NY001	7	436826	3079	13.78	1.03E-239	Haem peroxidase
S	NY001	3	436826	3079	13.77	4.00E-236	Haem peroxidase
S	NY001	5	436826	3079	14.36	1.19E-260	Haem peroxidase
R	28A	7	172060	245	-2.81	1.78E-150	LysM domain, Glycine zipper 2TM domain
R	28A	10	172060	245	-3.43	0.00E+00	LysM domain, Glycine zipper 2TM domain
S	28A	7	172060	245	-2.90	7.12E-176	LysM domain, Glycine zipper 2TM domain
S	28A	10	172060	245	-3.46	0.00E+00	LysM domain, Glycine zipper 2TM domain
R	NY001	7	523400	225	8.67	5.91E-258	Methyltransferase type 12
R	NY001	10	523400	225	9.38	0.00E+00	Methyltransferase type 12
S	NY001	5	523400	225	8.99	3.18E-283	Methyltransferase type 12
S	NY001	7	523400	225	9.31	0.00E+00	Methyltransferase type 12
S	NY001	10	523400	225	8.54	2.98E-268	Methyltransferase type 12
R	28A	3	19500	329	8.88	2.77E-213	Multicopper oxidase, type 1/2/3
R	28A	5	19500	329	8.80	6.00E-234	Multicopper oxidase, type 1/2/3
R	28A	7	19500	329	7.92	2.39E-188	Multicopper oxidase, type 1/2/3
S	28A	3	19500	329	8.84	9.06E-213	Multicopper oxidase, type 1/2/3
S	28A	5	19500	329	8.59	2.42E-217	Multicopper oxidase, type 1/2/3

S	28A	7	19500	329	8.09	4.25E-199	Multicopper oxidase, type 1/2/3
R	NY001	7	448170	138	8.99	1.57E-250	Multicopper oxidase, type 1/2/3
R	NY001	10	448170	138	8.70	1.26E-241	Multicopper oxidase, type 1/2/3
S	NY001	3	448170	138	9.47	5.50E-195	Multicopper oxidase, type 1/2/3
S	NY001	5	448170	138	8.93	1.67E-249	Multicopper oxidase, type 1/2/3
S	NY001	7	448170	138	8.87	9.01E-261	Multicopper oxidase, type 1/2/3
R	28A	3	185131	3366	7.82	7.11E-177	Peptidase S10, serine carboxypeptidase
R	28A	5	185131	3366	8.05	3.77E-190	Peptidase S10, serine carboxypeptidase
R	28A	7	185131	3366	7.64	2.37E-171	Peptidase S10, serine carboxypeptidase
S	28A	3	185131	3366	7.53	1.63E-163	Peptidase S10, serine carboxypeptidase
S	28A	5	185131	3366	7.93	4.61E-184	Peptidase S10, serine carboxypeptidase
S	28A	7	185131	3366	7.80	6.96E-179	Peptidase S10, serine carboxypeptidase
S	NY001	10	539128	42	7.95	4.11E-215	Peptidase S28
R	NY001	3	599170	1794	12.60	1.96E-278	Protein of unknown function DUF3328
R	NY001	5	599170	1794	13.55	0.00E+00	Protein of unknown function DUF3328
R	NY001	7	599170	1794	13.36	0.00E+00	Protein of unknown function DUF3328
R	NY001	10	599170	1794	13.78	0.00E+00	Protein of unknown function DUF3328
S	NY001	3	599170	1794	12.30	3.00E-268	Protein of unknown function DUF3328
S	NY001	5	599170	1794	14.14	0.00E+00	Protein of unknown function DUF3328
S	NY001	7	599170	1794	13.91	0.00E+00	Protein of unknown function DUF3328
S	NY001	10	599170	1794	13.16	0.00E+00	Protein of unknown function DUF3328
S	NY001	7	389997	70	8.09	1.36E-237	Protein of unknown function DUF3455
S	28A	10	1409328	101	-7.76	9.64E-239	Ubiquitin 3 binding protein But2, C-terminal
R	28A	10	87828	76	3.69	0.00E+00	
S	28A	10	87828	76	3.69	0.00E+00	
R	28A	3	87999	545	6.75	9.51E-185	
R	28A	5	87999	545	6.50	4.96E-190	
R	28A	7	87999	545	6.14	7.86E-171	
S	28A	3	87999	545	6.58	1.14E-174	
S	28A	5	87999	545	6.45	5.01E-183	
S	28A	7	87999	545	6.20	7.12E-176	
R	28A	10	90147	686	-11.36	0.00E+00	
S	28A	10	90147	686	-11.68	0.00E+00	
R	28A	7	91593	126	-4.59	8.29E-157	
R	28A	10	91593	126	-4.12	0.00E+00	
S	28A	7	91593	126	-4.66	6.86E-188	
S	28A	10	91593	126	-3.87	1.31E-292	
R	28A	3	93458	1564	9.86	2.09E-166	
R	28A	5	93458	1564	11.05	1.40E-220	
R	28A	7	93458	1564	11.37	6.98E-234	
S	28A	3	93458	1564	9.52	1.25E-153	
S	28A	5	93458	1564	11.26	2.09E-228	
S	28A	7	93458	1564	11.78	7.70E-252	
R	28A	10	94576	115	-6.20	0.00E+00	
S	28A	10	94576	115	-5.88	0.00E+00	
R	28A	10	99782	133	-5.95	3.02E-268	
S	28A	10	99782	133	-5.84	4.45E-267	

R	28A	3	105298	6928	15.15	5.43E-168	
R	28A	5	105298	6928	14.64	3.67E-157	
S	28A	3	105298	6928	14.73	1.07E-158	
S	28A	5	105298	6928	15.07	2.01E-166	
R	28A	3	135655	973	7.53	8.80E-179	
R	28A	5	135655	973	7.71	2.65E-198	
R	28A	7	135655	973	7.22	4.93E-174	
S	28A	3	135655	973	7.24	6.96E-164	
S	28A	5	135655	973	7.58	7.52E-190	
S	28A	7	135655	973	7.26	1.13E-176	
R	28A	3	136414	8995	11.73	0.00E+00	
R	28A	5	136414	8995	11.55	0.00E+00	
R	28A	7	136414	8995	10.85	0.00E+00	
S	28A	3	136414	8995	11.40	0.00E+00	
S	28A	5	136414	8995	11.62	0.00E+00	
S	28A	7	136414	8995	11.01	0.00E+00	
R	28A	10	156179	74	-4.66	0.00E+00	
S	28A	10	156179	74	-5.04	0.00E+00	
R	28A	3	158705	9750	16.86	1.02E-171	
R	28A	5	158705	9750	16.50	1.17E-164	
S	28A	3	158705	9750	16.47	9.77E-164	
S	28A	5	158705	9750	16.62	4.23E-167	
R	28A	10	163718	164	-4.40	0.00E+00	
S	28A	10	163718	164	-4.24	0.00E+00	
R	28A	5	164809	359	-4.38	1.76E-165	
R	28A	7	164809	359	-4.46	3.05E-192	
R	28A	10	164809	359	-4.39	1.18E-247	
S	28A	7	164809	359	-4.21	2.34E-191	
S	28A	10	166088	146	-4.91	1.32E-266	
R	28A	3	168886	6963	12.25	4.03E-234	
R	28A	5	168886	6963	12.20	2.35E-233	
R	28A	7	168886	6963	11.81	1.53E-218	
S	28A	3	168886	6963	11.90	6.87E-221	
S	28A	5	168886	6963	12.11	2.21E-229	
S	28A	7	168886	6963	11.78	1.50E-217	
R	28A	10	178296	148	-9.71	8.02E-251	
S	NY001	5	425953	103	6.75	6.79E-248	
S	NY001	7	425953	103	6.31	1.76E-236	
R	NY001	3	439242	2397	14.63	3.02E-210	
R	NY001	5	439242	2397	14.83	2.86E-220	
R	NY001	7	439242	2397	14.58	2.18E-213	
S	NY001	3	439242	2397	13.97	1.47E-191	
R	NY001	3	444690	1920	15.16	2.41E-231	
R	NY001	5	444690	1920	15.58	1.07E-251	
R	NY001	7	444690	1920	15.39	2.46E-246	
R	NY001	10	444690	1920	15.40	4.18E-247	
S	NY001	3	444690	1920	14.52	3.73E-212	

S	NY001	5	444690	1920	16.00	9.55E-267	
S	NY001	7	444690	1920	15.44	2.39E-248	
S	NY001	10	444690	1920	14.18	1.53E-209	
R	NY001	3	492699	5692	16.53	4.78E-279	
R	NY001	5	492699	5692	17.24	0.00E+00	
R	NY001	7	492699	5692	17.12	1.92E-304	
R	NY001	10	492699	5692	17.28	0.00E+00	
S	NY001	3	492699	5692	16.21	2.87E-269	
S	NY001	5	492699	5692	17.66	0.00E+00	
S	NY001	7	492699	5692	17.16	0.00E+00	
S	NY001	10	492699	5692	16.01	2.61E-266	
S	NY001	10	498810	42	6.77	6.89E-257	
R	NY001	3	521286	518	8.25	7.83E-197	
R	NY001	5	521286	518	8.55	7.80E-264	
R	NY001	7	521286	518	8.47	9.90E-271	
R	NY001	10	521286	518	8.27	5.87E-261	
S	NY001	3	521286	518	8.41	2.82E-217	
S	NY001	5	521286	518	8.79	7.40E-294	
S	NY001	7	521286	518	8.52	1.67E-280	
R	NY001	3	522672	1743	10.87	5.44E-183	
R	NY001	5	522672	1743	12.14	5.23E-248	
R	NY001	7	522672	1743	11.92	1.21E-240	
R	NY001	10	522672	1743	12.50	3.01E-265	
S	NY001	5	522672	1743	12.35	8.36E-259	
S	NY001	10	522672	1743	11.35	1.09E-218	
R	NY001	3	523229	2481	14.28	4.80E-200	
R	NY001	5	523229	2481	14.82	4.57E-220	
S	NY001	3	523229	2481	13.88	1.25E-189	
R	NY001	5	593519	179	9.31	8.07E-224	
S	28A	5	1224365	2844	12.66	6.85E-165	
R	28A	3	1302593	738	12.19	3.57E-150	
S	28A	3	1302593	738	12.06	5.18E-147	

Supplementary Table S2.2: Predicted Pfam domains that were significantly enriched ($p < 0.01$ by Fisher's exact test, after Bonferroni correction) among *S. turcica* gene models upregulated or downregulated *in planta*.

Enriched among	Pfam ID	Isolate	Plant	DPI	# proteins	# DE (expected)	# DE (observed)	p	Description
Upreg	PF00067	28A	R	3	134	13.5	29	3.17E-03	Cytochrome P450
Upreg	PF00067	28A	R	7	134	26.6	44	8.34E-03	Cytochrome P450
Upreg	PF00067	28A	S	3	134	14.0	30	3.03E-03	Cytochrome P450
Upreg	PF00067	28A	S	7	134	28.0	51	8.30E-04	Cytochrome P450
Upreg	PF00067	NY001	R	3	132	6.8	23	2.75E-05	Cytochrome P450
Upreg	PF00067	NY001	R	5	132	10.0	26	4.20E-05	Cytochrome P450
Upreg	PF00067	NY001	R	7	132	12.1	29	5.47E-05	Cytochrome P450
Upreg	PF00067	NY001	R	10	132	15.3	33	1.19E-04	Cytochrome P450
Upreg	PF00067	NY001	S	3	132	6.9	20	4.22E-04	Cytochrome P450
Upreg	PF00067	NY001	S	7	132	13.9	27	6.61E-03	Cytochrome P450
Upreg	PF00067	NY001	S	10	132	17.8	34	1.34E-03	Cytochrome P450
Upreg	PF00106	28A	R	5	126	20.7	39	5.89E-03	short chain dehydrogenase
Upreg	PF00106	28A	R	7	126	25.0	46	1.81E-03	short chain dehydrogenase
Upreg	PF00106	28A	S	7	126	26.3	49	1.22E-03	short chain dehydrogenase
Upreg	PF00106	NY001	R	10	128	14.9	31	4.85E-03	short chain dehydrogenase
Upreg	PF00106	NY001	S	7	128	13.5	29	7.71E-03	short chain dehydrogenase
Upreg	PF00106	NY001	S	10	128	17.3	43	4.27E-06	short chain dehydrogenase
Upreg	PF00107	28A	R	3	59	5.9	18	4.21E-03	Zinc-binding dehydrogenase
Upreg	PF00107	28A	R	7	59	11.7	31	5.14E-05	Zinc-binding dehydrogenase
Upreg	PF00107	28A	R	10	59	15.8	36	7.85E-05	Zinc-binding dehydrogenase
Upreg	PF00107	28A	S	3	59	6.2	18	6.03E-03	Zinc-binding dehydrogenase
Upreg	PF00107	28A	S	7	59	12.3	30	3.33E-04	Zinc-binding dehydrogenase
Upreg	PF00107	NY001	R	3	59	3.0	12	5.62E-03	Zinc-binding dehydrogenase
Upreg	PF00107	NY001	R	5	59	4.5	16	7.27E-04	Zinc-binding dehydrogenase
Upreg	PF00107	NY001	R	7	59	5.4	22	4.27E-06	Zinc-binding dehydrogenase
Upreg	PF00107	NY001	R	10	59	6.9	24	7.13E-06	Zinc-binding dehydrogenase
Upreg	PF00107	NY001	S	5	59	5.6	16	7.81E-03	Zinc-binding dehydrogenase
Upreg	PF00107	NY001	S	7	59	6.2	21	7.39E-05	Zinc-binding dehydrogenase
Upreg	PF00107	NY001	S	10	59	8.0	25	2.75E-05	Zinc-binding dehydrogenase
Upreg	PF00153	28A	S	10	39	10.7	23	9.98E-03	Mitochondrial carrier protein
Upreg	PF00248	NY001	S	10	27	3.7	12	2.95E-03	Aldo/keto reductase family
Upreg	PF00324	NY001	R	5	13	1.0	7	3.89E-03	Amino acid permease
Upreg	PF00400	NY001	R	10	104	12.1	0	1.66E-03	WD domain, G-beta repeat
Upreg	PF00400	NY001	S	10	104	14.1	0	2.25E-04	WD domain, G-beta repeat
Upreg	PF00561	NY001	R	10	7	0.8	6	3.83E-03	alpha/beta hydrolase fold
Upreg	PF00775	NY001	R	5	4	0.3	4	6.61E-03	Dioxygenase

Upreg	PF01532	28A	S	5	10	1.7	8	7.31E-03	Glycosyl hydrolase family 47
Upreg	PF01557	NY001	S	5	9	0.9	6	3.96E-03	Fumarylacetoacetate (FAA) hydrolase family
Upreg	PF02133	28A	R	10	9	2.4	9	3.03E-03	Permease for cytosine/purines
Upreg	PF02133	28A	S	10	9	2.5	9	3.29E-03	Permease for cytosine/purines
Upreg	PF02133	NY001	S	10	9	1.2	7	5.02E-03	Permease for cytosine/purines
Upreg	PF02668	28A	R	3	16	1.6	8	5.89E-03	Taurine catabolism dioxygenase TauD, TfdA family
Upreg	PF02668	NY001	R	3	15	0.8	7	2.25E-04	Taurine catabolism dioxygenase TauD, TfdA family
Upreg	PF02668	NY001	R	5	15	1.1	7	4.35E-04	Taurine catabolism dioxygenase TauD, TfdA family
Upreg	PF02668	NY001	S	3	15	0.8	7	1.22E-04	Taurine catabolism dioxygenase TauD, TfdA family
Upreg	PF07690	28A	R	5	203	33.3	57	1.59E-03	Major Facilitator Superfamily
Upreg	PF07690	28A	R	7	203	40.2	77	1.04E-06	Major Facilitator Superfamily
Upreg	PF07690	28A	R	10	203	54.4	88	1.53E-04	Major Facilitator Superfamily
Upreg	PF07690	28A	S	3	203	21.2	45	3.33E-04	Major Facilitator Superfamily
Upreg	PF07690	28A	S	5	203	34.8	58	2.07E-03	Major Facilitator Superfamily
Upreg	PF07690	28A	S	7	203	42.4	84	1.70E-07	Major Facilitator Superfamily
Upreg	PF07690	28A	S	10	203	55.6	84	2.86E-03	Major Facilitator Superfamily
Upreg	PF07690	NY001	R	5	200	15.2	35	8.00E-05	Major Facilitator Superfamily
Upreg	PF07690	NY001	R	7	200	18.3	38	4.30E-04	Major Facilitator Superfamily
Upreg	PF07690	NY001	R	10	200	23.2	48	2.75E-05	Major Facilitator Superfamily
Upreg	PF07690	NY001	S	5	200	19.0	40	2.25E-04	Major Facilitator Superfamily
Upreg	PF07690	NY001	S	7	200	21.1	43	2.57E-04	Major Facilitator Superfamily
Upreg	PF07690	NY001	S	10	200	27.0	56	7.13E-06	Major Facilitator Superfamily
Upreg	PF08240	28A	R	7	62	12.3	29	1.18E-03	Alcohol dehydrogenase GroES-like domain
Upreg	PF08240	28A	R	10	62	16.6	32	9.93E-03	Alcohol dehydrogenase GroES-like domain
Upreg	PF08240	28A	S	7	62	13.0	29	2.57E-03	Alcohol dehydrogenase GroES-like domain
Upreg	PF08240	NY001	R	7	63	5.8	20	5.47E-05	Alcohol dehydrogenase GroES-like domain
Upreg	PF08240	NY001	R	10	63	7.3	25	7.13E-06	Alcohol dehydrogenase GroES-like domain
Upreg	PF08240	NY001	S	7	63	6.6	19	1.92E-03	Alcohol dehydrogenase GroES-like domain
Upreg	PF08240	NY001	S	10	63	8.5	25	3.70E-05	Alcohol dehydrogenase GroES-like domain
Down	PF00072	28A	R	5	25	4.4	17	6.05E-05	Response regulator receiver domain
Downreg	PF00072	28A	R	7	25	4.9	19	1.43E-05	Response regulator receiver domain
Downreg	PF00072	28A	S	5	25	4.2	17	6.05E-05	Response regulator receiver

									domain
Downreg	PF00072	28A	S	7	25	5.1	18	6.05E-05	Response regulator receiver domain
Downreg	PF00072	NY001	S	5	25	2.8	12	6.14E-03	Response regulator receiver domain
Downreg	PF00227	NY001	R	7	14	1.0	10	2.66E-05	Proteasome subunit
Downreg	PF00227	NY001	S	7	14	2.0	10	3.02E-03	Proteasome subunit
Downreg	PF00394	NY001	R	3	8	0.1	4	3.02E-03	Multicopper oxidase
Downreg	PF00512	28A	R	5	21	3.7	15	1.18E-04	His Kinase A (phospho-acceptor) domain
Downreg	PF00512	28A	R	7	21	4.1	15	3.34E-04	His Kinase A (phospho-acceptor) domain
Downreg	PF00512	28A	S	5	21	3.5	14	3.76E-04	His Kinase A (phospho-acceptor) domain
Downreg	PF00512	28A	S	7	21	4.3	14	3.33E-03	His Kinase A (phospho-acceptor) domain
Downreg	PF02518	28A	R	5	25	4.4	17	6.05E-05	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase
Downreg	PF02518	28A	R	7	25	4.9	17	1.90E-04	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase
Downreg	PF02518	28A	S	5	25	4.2	16	1.87E-04	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase
Downreg	PF02518	28A	S	7	25	5.1	16	1.70E-03	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase
Downreg	PF07731	NY001	R	3	9	0.1	4	3.02E-03	Multicopper oxidase
Downreg	PF07732	NY001	R	3	9	0.1	4	3.02E-03	Multicopper oxidase

Supplementary Table S2.3: Ten most significantly differentially expressed maize genes at each combination of host (R: *Ht2+*, S: *Ht2-*), pathogen isolate (NY001: *AVRHt2*, 28A: *avrHt2*), and timepoint (DPI: days post-inoculation), organized by predicted Pfam protein domain(s).

Plant	Isolate	DPI	Gene model	Base mean expression	log2 Fold Change	p	Description
R	28A	10	GRMZM5G899851	12430	6.53	1.6E-124	2Fe-2S iron-sulfur cluster binding domain
S	28A	7	GRMZM5G899851	12430	4.36	1.6E-54	2Fe-2S iron-sulfur cluster binding domain
R	28A	3	GRMZM2G099467	9066	4.18	5.61E-28	2OG-Fe(II) oxygenase superfamily
R	28A	7	GRMZM2G413774	1473	5.94	9.11E-46	ABC transporter
R	NY001	10	GRMZM2G413774	1473	5.69	3.03E-44	ABC transporter
S	28A	7	GRMZM2G413774	1473	6.65	8.85E-54	ABC transporter
S	NY001	7	GRMZM2G413774	1473	6.19	3.34E-46	ABC transporter
S	NY001	10	GRMZM2G413774	1473	6.98	2.25E-65	ABC transporter
R	28A	10	GRMZM2G391815	4848	8.04	3.64E-122	ABC transporter
S	28A	10	GRMZM2G391815	4848	7.66	7.16E-108	ABC transporter
S	NY001	10	GRMZM2G391815	4848	7.46	3.67E-102	ABC transporter
R	28A	10	GRMZM2G391815	4848	8.04	3.64E-122	ABC transporter
S	28A	10	GRMZM2G391815	4848	7.66	7.16E-108	ABC transporter
S	NY001	10	GRMZM2G391815	4848	7.46	3.67E-102	ABC transporter
R	28A	3	GRMZM2G415529	3401	5.41	5.15E-51	ABC transporter
R	28A	7	GRMZM2G415529	3401	5.6	1.23E-53	ABC transporter
R	28A	10	GRMZM2G415529	3401	8.02	7.46E-117	ABC transporter
R	NY001	3	GRMZM2G415529	3401	5.88	9.03E-61	ABC transporter
R	NY001	7	GRMZM2G415529	3401	6.1	5.43E-64	ABC transporter
R	NY001	10	GRMZM2G415529	3401	6.67	9.47E-80	ABC transporter
S	28A	3	GRMZM2G415529	3401	6.67	2.04E-72	ABC transporter
S	28A	5	GRMZM2G415529	3401	5.97	6.88E-57	ABC transporter
S	28A	7	GRMZM2G415529	3401	6.28	1.34E-69	ABC transporter
S	28A	10	GRMZM2G415529	3401	6.55	9.09E-81	ABC transporter
S	NY001	3	GRMZM2G415529	3401	5.87	1.16E-55	ABC transporter
S	NY001	5	GRMZM2G415529	3401	6.09	2.94E-59	ABC transporter
S	NY001	7	GRMZM2G415529	3401	5.64	1.39E-55	ABC transporter
S	NY001	10	GRMZM2G415529	3401	6.7	5.95E-84	ABC transporter
R	28A	7	GRMZM2G020631	13838	6.95	9.11E-46	Aldo/keto reductase family
S	28A	7	GRMZM2G020631	13838	6.68	3.67E-43	Aldo/keto reductase family
S	28A	10	GRMZM2G020631	13838	8.86	1.76E-77	Aldo/keto reductase family
R	NY001	3	GRMZM2G155058	443	3.78	5.26E-35	Calcineurin-like phosphoesterase
S	28A	3	GRMZM2G456997	1138	5.18	4.12E-33	Cysteine-rich secretory protein family
R	28A	10	GRMZM2G170016	1389	7.99	5.21E-112	Cytochrome b5-like
S	28A	10	GRMZM2G170016	1389	6.96	4.17E-84	Cytochrome b5-like

S	NY001	10	GRMZM2G170016	1389	6.44	1.35E-71	Cytochrome b5-like
R	28A	3	GRMZM2G118809	1241	5.99	5.78E-60	Cytochrome P450
R	28A	5	GRMZM2G118809	1241	5.58	9.13E-48	Cytochrome P450
R	28A	7	GRMZM2G118809	1241	5.5	4.95E-57	Cytochrome P450
R	NY001	3	GRMZM2G118809	1241	6.21	1.11E-64	Cytochrome P450
R	NY001	5	GRMZM2G118809	1241	6.44	1.65E-64	Cytochrome P450
R	NY001	7	GRMZM2G118809	1241	5.89	6.52E-66	Cytochrome P450
R	NY001	10	GRMZM2G118809	1241	5.77	2.2E-72	Cytochrome P450
S	28A	3	GRMZM2G118809	1241	8.09	3.28E-69	Cytochrome P450
S	28A	5	GRMZM2G118809	1241	6.39	2.34E-62	Cytochrome P450
S	28A	7	GRMZM2G118809	1241	5.44	1.76E-58	Cytochrome P450
S	NY001	3	GRMZM2G118809	1241	7.67	1.45E-61	Cytochrome P450
S	NY001	5	GRMZM2G118809	1241	6.64	1.55E-67	Cytochrome P450
S	NY001	7	GRMZM2G118809	1241	5.29	2.33E-55	Cytochrome P450
R	28A	10	GRMZM2G161472	3523	7.87	2.96E-113	Cytochrome P450
R	NY001	10	GRMZM2G161472	3523	4.78	1.11E-40	Cytochrome P450
S	NY001	10	GRMZM2G161472	3523	6.63	1.52E-80	Cytochrome P450
R	NY001	3	GRMZM2G078667	1739	3.95	3.42E-35	Dirigent-like protein
S	28A	3	GRMZM2G078667	1739	3.96	1.15E-34	Dirigent-like protein
S	NY001	3	GRMZM2G078667	1739	3.58	6.88E-28	Dirigent-like protein
R	28A	3	GRMZM2G181227	3506	5.47	2.3E-42	Enoyl-CoA hydratase/isomerase
R	28A	5	GRMZM2G181227	3506	4.82	9.68E-31	Enoyl-CoA hydratase/isomerase
R	NY001	3	GRMZM2G181227	3506	5.95	1.09E-50	Enoyl-CoA hydratase/isomerase
R	NY001	5	GRMZM2G181227	3506	6.09	9.86E-50	Enoyl-CoA hydratase/isomerase
R	NY001	7	GRMZM2G181227	3506	5.75	6.52E-48	Enoyl-CoA hydratase/isomerase
R	NY001	10	GRMZM2G181227	3506	5.78	4.38E-49	Enoyl-CoA hydratase/isomerase
S	28A	3	GRMZM2G181227	3506	5.89	1.46E-49	Enoyl-CoA hydratase/isomerase
S	28A	5	GRMZM2G181227	3506	5.62	1.3E-44	Enoyl-CoA hydratase/isomerase
S	28A	7	GRMZM2G181227	3506	5.64	4.23E-46	Enoyl-CoA hydratase/isomerase
S	NY001	3	GRMZM2G181227	3506	4.81	3.5E-32	Enoyl-CoA hydratase/isomerase
S	NY001	5	GRMZM2G181227	3506	6	4.13E-51	Enoyl-CoA hydratase/isomerase
S	NY001	7	GRMZM2G181227	3506	5.08	4.38E-37	Enoyl-CoA hydratase/isomerase
R	28A	5	GRMZM2G169240	588	7.56	2.42E-25	Fatty acid desaturase, DUF3474
R	NY001	5	GRMZM2G169240	588	8.59	2.77E-33	Fatty acid desaturase, DUF3474
R	NY001	7	GRMZM2G169240	588	8.26	2.42E-33	Fatty acid desaturase, DUF3474
S	NY001	3	GRMZM2G169240	588	8.5	3.92E-26	Fatty acid desaturase, DUF3474
R	28A	5	GRMZM2G169261	856	6.95	3.82E-30	Fatty acid desaturase, DUF3474
R	NY001	5	GRMZM2G169261	856	7.96	7.98E-40	Fatty acid desaturase, DUF3474
R	NY001	7	GRMZM2G169261	856	9.63	4.49E-36	Fatty acid desaturase, DUF3474
S	NY001	7	GRMZM2G169261	856	9.31	4.38E-37	Fatty acid desaturase, DUF3474
R	NY001	3	GRMZM2G462243	1759	5.82	3.73E-35	Glutathione S-transferase, C-terminal domain
R	NY001	10	GRMZM2G462243	1759	5.96	8.3E-41	Glutathione S-transferase, C-terminal domain
R	28A	7	GRMZM2G176307	5777	6.39	1.04E-44	Glyceraldehyde 3-phosphate dehydrogenase
R	NY001	7	GRMZM2G374827	682	3.87	1.64E-39	Glyceraldehyde 3-phosphate dehydrogenase

R	NY001	10	GRMZM2G374827	682	3.86	7.45E-41	Glyceraldehyde 3-phosphate dehydrogenase
S	28A	3	GRMZM2G374827	682	3.77	6.74E-37	Glyceraldehyde 3-phosphate dehydrogenase
S	NY001	3	GRMZM2G374827	682	3.2	4.35E-26	Glyceraldehyde 3-phosphate dehydrogenase
S	NY001	5	GRMZM2G374827	682	3.47	1.28E-30	Glyceraldehyde 3-phosphate dehydrogenase
S	NY001	5	GRMZM2G175480	531	7.96	4.71E-29	Helix-loop-helix DNA-binding domain
S	28A	5	GRMZM2G084779	4892	5.66	5.52E-26	K+ potassium transporter
S	NY001	5	GRMZM2G084779	4892	5.95	9.35E-29	K+ potassium transporter
S	28A	10	GRMZM2G041959	1859	6.44	1.76E-82	Kelch motif
S	NY001	10	GRMZM2G091672	223	5.63	3.76E-64	Leucine Rich Repeat, NB-ARC domain
S	28A	5	GRMZM2G132995	443	5.58	2.49E-28	MatE
S	NY001	7	GRMZM2G132995	443	5.36	1.3E-34	MatE
S	NY001	10	GRMZM2G420988	1796	9.09	4.94E-64	Mitochondrial carrier protein
R	28A	10	GRMZM2G124175	9321	6.37	4.11E-133	Aldehyde dehydrogenase, Molybdopterin-binding domain
S	28A	7	GRMZM2G124175	9321	3.96	1.36E-50	Aldehyde dehydrogenase, Molybdopterin-binding domain
S	28A	10	GRMZM2G124175	9321	5.24	1.2E-89	Aldehyde dehydrogenase, Molybdopterin-binding domain
S	28A	10	GRMZM2G179981	8589	3.47	4.53E-82	NAD dependent epimerase/dehydratase family
R	28A	3	GRMZM2G162098	455	4.15	4.31E-32	NB-ARC domain
R	NY001	3	GRMZM2G162098	455	4.93	6.56E-46	NB-ARC domain
R	NY001	7	GRMZM2G162098	455	4.09	4.76E-38	NB-ARC domain
R	NY001	10	GRMZM2G162098	455	4.62	3.13E-53	NB-ARC domain
S	28A	3	GRMZM2G162098	455	5.06	4.27E-36	NB-ARC domain
S	28A	7	GRMZM2G162098	455	4.55	1.04E-46	NB-ARC domain
S	28A	5	GRMZM2G169584	59	4.91	1.81E-27	NB-ARC domain
S	NY001	5	GRMZM2G169584	59	5.17	1.72E-30	NB-ARC domain
R	28A	5	GRMZM2G475014	207	4.93	2.03E-31	No apical meristem (NAM) protein
R	NY001	5	GRMZM2G475014	207	4.95	6.37E-32	No apical meristem (NAM) protein
R	NY001	7	GRMZM2G475014	207	3.71	3.77E-35	No apical meristem (NAM) protein
S	28A	5	GRMZM2G475014	207	5.81	4.23E-41	No apical meristem (NAM) protein
S	28A	7	GRMZM2G475014	207	4.59	1.23E-52	No apical meristem (NAM) protein
S	NY001	5	GRMZM2G475014	207	5.92	1.19E-42	No apical meristem (NAM) protein
S	NY001	7	GRMZM2G475014	207	4.02	1.01E-39	No apical meristem (NAM) protein
S	28A	3	GRMZM2G467893	788	4.45	2.22E-33	Nodulin-like
S	28A	5	GRMZM2G467893	788	4.41	2.83E-32	Nodulin-like
S	NY001	3	GRMZM2G467893	788	4.18	8.84E-29	Nodulin-like
S	NY001	5	GRMZM2G467893	788	4.28	1.97E-30	Nodulin-like
S	28A	3	GRMZM2G107228	128	5.69	1.54E-30	Peroxidase
S	NY001	3	GRMZM2G107228	128	5.4	7.67E-27	Peroxidase
R	28A	5	GRMZM2G322819	55	19.88	1.46E-31	Prolyl oligopeptidase, N-term.
R	NY001	5	GRMZM2G322819	55	19.93	5.64E-32	Prolyl oligopeptidase, N-term.
R	NY001	7	GRMZM2G322819	55	22.37	7.32E-41	Prolyl oligopeptidase, N-term.

S	28A	3	GRMZM2G322819	55	20.31	3.37E-33	Prolyl oligopeptidase, N-term.
S	NY001	3	GRMZM2G322819	55	19.51	2.92E-29	Prolyl oligopeptidase, N-term.
R	28A	3	GRMZM2G432796	1564	1.6	3.4E-28	Protein kinase domain
S	NY001	5	GRMZM2G303909	328	3.7	1.79E-27	Protein kinase domain, D-mannose binding lectin
R	28A	3	GRMZM2G443829	326	4.01	1.1E-26	Protein kinase domain, Legume lectin domain
S	28A	10	GRMZM2G087233	6376	3.77	3.79E-83	Ribosomal protein L16p/L10e
R	28A	3	GRMZM2G452896	1838	2.88	4.68E-28	Serine carboxypeptidase S28
R	NY001	3	GRMZM2G452896	1838	3.19	1.39E-34	Serine carboxypeptidase S28
R	NY001	7	AC214360.3_FG001	3018	5.48	1.98E-35	Terpene synthase, N-term/metal binding domain
R	NY001	5	GRMZM2G028306	3943	10.46	4.83E-39	Terpene synthase, N-term/metal binding domain
R	NY001	10	GRMZM2G028306	3943	8.32	8.3E-41	Terpene synthase, N-term/metal binding domain
S	NY001	10	GRMZM2G028306	3943	11.46	1.69E-69	Terpene synthase, N-term/metal binding domain
R	28A	3	GRMZM2G127087	8039	11.64	2.93E-28	Terpene synthase, N-term/metal binding domain
R	28A	5	GRMZM2G127087	8039	11.23	5.82E-25	Terpene synthase, N-term/metal binding domain
R	NY001	5	GRMZM2G127087	8039	12.92	1.96E-33	Terpene synthase, N-term/metal binding domain
S	NY001	3	GRMZM2G127087	8039	12.03	3.92E-26	Terpene synthase, N-term/metal binding domain
S	NY001	10	GRMZM2G127087	8039	12.31	1.44E-63	Terpene synthase, N-term/metal binding domain
R	28A	3	GRMZM2G315726	1654	2.69	8.87E-28	TLC domain
R	NY001	3	GRMZM2G315726	1654	2.98	2.26E-34	TLC domain
R	28A	5	AC233910.1_FG003	122	6.38	2.46E-25	Ubiquitin family, Ribosomal L40e family
R	28A	7	AC233910.1_FG003	122	5.32	1.28E-39	Ubiquitin family, Ribosomal L40e family
R	28A	7	GRMZM2G334336	975	5.35	8.66E-41	UDP-glucoronosyl and UDP-glucosyl transferase
R	28A	10	GRMZM2G334336	975	9.17	7.66E-115	UDP-glucoronosyl and UDP-glucosyl transferase
S	28A	5	GRMZM2G334336	975	5.68	1.72E-31	UDP-glucoronosyl and UDP-glucosyl transferase
S	NY001	7	GRMZM2G334336	975	5.77	7.62E-44	UDP-glucoronosyl and UDP-glucosyl transferase
R	28A	5	GRMZM2G100898	240	36.92	1.38E-32	XS domain, XH domain, XS zinc finger domain
R	28A	7	GRMZM2G100898	240	-40.86	1.93E-45	XS domain, XH domain, XS zinc finger domain
R	NY001	5	GRMZM2G100898	240	38.29	3.24E-37	XS domain, XH domain, XS zinc finger domain
R	28A	10	GRMZM2G090980	11170	5.77	3.27E-119	Zinc-binding/Alcohol dehydrogenase GroES-like domain

S	28A	10	GRMZM2G090980	11170	4.89	4.78E-85	Zinc-binding/Alcohol dehydrogenase GroES-like domain
S	NY001	10	GRMZM2G090980	11170	4.31	7.56E-66	Zinc-binding/Alcohol dehydrogenase GroES-like domain
R	28A	10	AC214635.3_FG006	2388	8.19	5.01E-117	
S	28A	10	AC214635.3_FG006	2388	6.92	9.89E-85	
S	28A	10	GRMZM2G041959	1859	6.44	1.76E-82	
R	28A	5	GRMZM2G069932	24	40.3	1.09E-65	
R	28A	10	GRMZM2G069932	24	50.05	4.31E-106	
R	NY001	3	GRMZM2G069932	24	43.21	8.51E-76	
R	NY001	5	GRMZM2G069932	24	41.63	3.7E-71	
R	NY001	10	GRMZM2G069932	24	47.39	1.1E-93	
R	28A	7	GRMZM2G119322	322	5.19	6.85E-45	
S	28A	7	GRMZM2G119322	322	5.19	1.56E-46	
S	NY001	7	GRMZM2G119322	322	4.51	1.21E-34	
R	28A	7	GRMZM2G468111	697	7.07	3.33E-44	
S	28A	5	GRMZM2G468111	697	6	1.31E-29	
S	NY001	7	GRMZM2G468111	697	6.78	3.99E-42	

Supplementary Table S2.4: Predicted PKS/NRPS encoding genes in *S. turcica* isolates.

		28A				NY001			
Type	**Isolates**	**Protein ID**	**Scaff.**	**Start**	**End**	**Protein ID**	**Scaff.**	**Start**	**End**
Hybrid PKS-NRPS	Both	179218	2	3545435	3557663	444637	254	4684	16290
Hybrid PKS-NRPS	Both	47468	20	512645	525329	470037	177	27754	38780
NRPS	28A only	1256853	10	1948473	1948775	NA	NA	NA	NA
NRPS	28A only	51661	10	1948761	1949171	NA	NA	NA	NA
NRPS	Both	29755	1	62373	76298	520000	167	11986	25806
NRPS	Both	85461	1	1419528	1425397	227272	45	153043	159310
NRPS	Both	178272	1	2520964	2524379	540897	92	49474	52889
NRPS	Both	18754	1	3193173	3196632	564795	20	22677	26349
NRPS	Both	163244	2	2696855	2699504	420336	8	415105	417341
NRPS	Both	179218	2	3545435	3557663	444637	254	4684	16290
NRPS	Both	179280	3	415602	438043	543322	1	1574976	1597178
NRPS	Both	155102	4	334954	338493	468908	144	36354	40064
NRPS	Both	1405196	4	2520629	2522879	518907	141	11943	15271
NRPS	Both	141443	6	2077032	2093337	223375	43	191738	208293
NRPS	Both	173669	7	481091	484939	543585	2	738191	742039
NRPS	Both	97841	7	1919348	1924485	548269	39	32405	38877
NRPS	Both	99043	8	1007319	1010673	441236	140	56532	59886
NRPS	Both	99181	8	1650732	1669532	607382	88	23386	42378
NRPS	Both	166785	10	1041597	1045485	503841	17	156653	160606
NRPS	Both	36641	10	1928291	1934680	609376	211	26676	33594
NRPS	Both	65284	10	1949289	1953251	442625	169	2095	4983
NRPS	Both	158798	11	720196	723991	210853	4	543150	545239
NRPS	Both	54477	14	603410	612100	550530	77	399	9098
NRPS	Both	47468	20	512645	525329	470037	177	27754	38780
NRPS	Both	169407	22	160579	163406	528260	32	173144	175215
NRPS	NY001 only	NA	NA	NA	NA	516685	108	34202	36374
NRPS	NY001 only	NA	NA	NA	NA	441768	150	23694	25761
NRPS	NY001 only	NA	NA	NA	NA	383654	169	190	1974
PKS/NRPS-like protein	Both	174337	8	1634382	1640616	515003	88	52266	58792
Type I Iterative PKS	Both	152662	1	1458156	1464486	485490	45	114497	119909
Type I Iterative PKS	Both	30113	1	1465118	1472667	538873	45	106316	113865
Type I Iterative PKS	Both	171983	3	1687808	1696182	600137	1	325270	333669
Type I Iterative PKS	Both	154973	3	3100977	3109360	544590	6	549753	558136
Type I Iterative PKS	Both	21011	3	3452451	3459694	552175	122	78362	85605
Type I Iterative PKS	Both	92491	4	326299	334342	552647	144	27722	35765
Type I Iterative PKS	Both	33341	4	2377943	2385329	552297	127	64975	72361
Type I Iterative PKS	Both	93994	5	1540110	1548361	434732	69	27969	36220
Type I Iterative PKS	Both	22370	6	70512	78088	552764	149	44436	52012
Type I Iterative PKS	Both	181366	7	1402178	1410045	547291	27	203067	210854
Type I Iterative PKS	Both	100517	9	162363	169917	455599	23	232189	239789
Type I Iterative PKS	Both	158064	9	1708392	1717224	606828	74	35754	44831
Type I Iterative PKS	Both	174734	9	1825519	1834360	574724	51	24790	33631

Type I Iterative PKS	Both	158567	10	1718469	1726126	520547	181	34964	41359
Type I Iterative PKS	Both	103204	11	397371	404288	443168	185	24254	31171
Type I Iterative PKS	Both	175651	12	567066	574223	440951	136	37201	43724
Type I Iterative PKS	Both	38126	14	428942	434480	493842	133	61201	66911
Type I Iterative PKS	Both	184365	18	254127	262657	542253	161	13725	21985
Type I Iterative PKS	Both	161582	22	425520	432714	553208	178	5738	12932
Type I Iterative PKS	Both	161586	22	472674	480118	443749	206	704	8148
Type I Iterative PKS	Both	161587	22	481619	489481	495967	206	18409	25726
Type I Iterative PKS	Both	1323930	29	9229	17122	553774	240	4691	12584
Type I Iterative PKS	Both	40268	29	77574	84157	541777	130	29331	35995
Type I Iterative PKS	NY001 only	NA	NA	NA	NA	553509	205	3671	11546
Type I Iterative PKS	NY001 only	NA	NA	NA	NA	553510	205	12367	20065
Type I Modular PKS	Both	39554	20	415948	424697	562383	168	10985	19734

Supplementary Table S2.5: All StNY001 gene models located on scaffolds associated with *AVRHt2*, their homolog in St28A (if one exists), expression summary, and SSP/NRPS/PKS prediction status.

StNY001		St28A					
Protein ID	**Scaffold**	**Protein ID**	**Scaffold**	**Polymorphic**	**Expression**	**Effector**	**Notes**
514223	81	NA	NA	NA	Upreg. in StNY001		
550711	81	35476	8	True	Upreg. in both		
436153	81	24034	8	False	Upreg. in both		
436251	81	99303	8	False	Downreg. in both		
436228	81	181802	8	False	Downreg. in both		
514227	81	166034	8	True	Downreg. in both		
550717	81	24029	8	True	Downreg. in both		
514233	81	NA	NA	NA	Downreg. In StNY001	SSP	100% ident.
531053	81	NA	NA	NA	Not DE		100% ident.
436269	81	98296	8	False	Not DE		
312248	81	NA	NA	NA	Not DE		100% ident.
312247	81	NA	NA	NA	Not DE	SSP	100% ident.
607114	81	181799	8	False	Upreg. in both		
531056	81	44505	8	False	Upreg. in both		
514244	81	166030	8	True	Upreg. in both		
531058	81	35466	8	False	Downreg. in both		
436182	81	99416	8	False	Downreg. in both		
597655	81	1407017	8	True	Downreg. in both		
514246	81	NA	NA	NA	Not DE		100% ident.
514247	81	143325	8	False	Not DE		
464712	81	174305	8	True			
514253	81	NA	NA	NA	Not DE		100% ident., polymorphisms upstream
531066	81	1386030	8	True	Downreg. in both		
514257	81	NA	NA	NA	Downreg. in both		
514258	81	99171	8	True	Downreg. in both		
514259	81	NA	NA	NA	Not DE		100% ident.
550731	81	44495	8	False	Downreg. in both		
312556	81	166021	8	False	Downreg. in both		
514266	81	157529	8	True	Upreg. in both		
312884	81	24011	8	False	Upreg. in both		
514271	81	143304	8	False	Upreg. in both		
540521	81	174297	8	False			
531074	81	143299	8	True	Not DE		
436166	81	NA	NA	NA	Not DE		100% ident.
514274	81	143296	8	True	Upreg. in both		
490154	81	123527	8	True			
312936	81	NA	NA	NA	Not DE		100% ident.

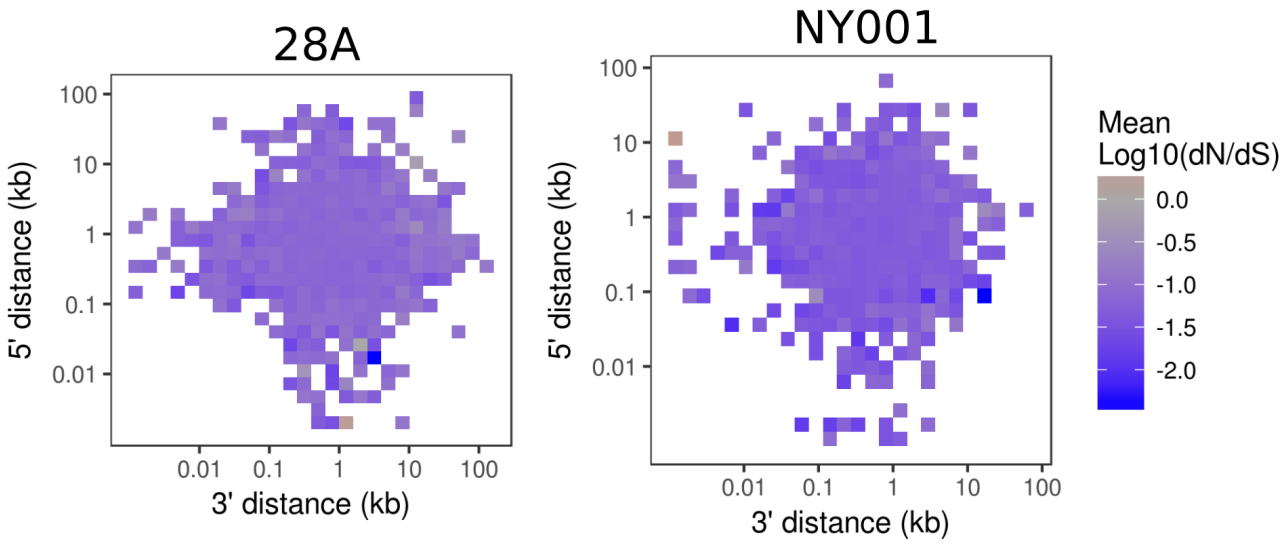
514277	81	174292	8	True	Not DE		
514278	81	44486	8	False	Downreg. in both		
514279	81	166016	8	False	Not DE		
531080	81	NA	NA	NA	Upreg. in StNY001		100% ident.
313034	81	174289	8	False	Upreg. in both		
514286	81	166015	8	True	Downreg. in both		
313216	81	NA	NA	NA	Upreg. in StNY001		100% ident.
540529	81	98593	8	True	Downreg. in both		
436181	81	98988	8	False	Upreg. in both		
579857	81	98756	8	False	Upreg. in both		
436255	81	166012	8	False	Downreg. in both		
514297	81	44481	8	False	Downreg. in both		
531086	81	35438	8	False	Upreg. in both	SSP	
313638	81	166011	8	True	Upreg. in both		
313644	81	1212335	8	False	Upreg. in both		
436210	81	166010	8	True	Upreg. in both		
436223	81	143261	8	False	Upreg. in both		
313669	81	1415037	8	True	Upreg. in both		
436252	81	166008	8	False	Downreg. in both		
490194	81	23991	8	False	Downreg. in both		
514305	81	143255	8	False	Not DE		
436157	81	166007	8	False	Downreg. in both		
436235	81	NA	NA	NA	Not DE		100% ident.
579909	81	157501	8	True	Not DE		
437032	88	98395	8	False	Upreg. in both		
551020	88	123651	8	True			
514997	88	98484	8	False	Downreg. in both		
607382	88	99181	8	True	Not DE	NRPS	
321030	88	166061	8	False	Downreg. in both		
515001	88	174338	8	True	Not DE		
515003	88	174337	8	False	Not DE	PKS/ NRPS-like protein	
515008	88	1373915	8	True	Upreg. in both		
437053	88	1431171	8	False	Downreg. in both		
551028	88	174334	8	False	Downreg. in both		
540766	88	24058	8	False	Downreg. in both		
321170	88	166054	8	False			
551031	88	166053	8	False	Downreg. in both		
531338	88	NA	NA	NA	Not DE		100% ident.
540769	88	98544	8	True	Downreg. in both		
515021	88	44521	8	False	Downreg. in both		
437067	88	143379	8	False	Upreg. in both		
321419	88	NA	NA	NA	Not DE		
580658	88	174325	8	False	Downreg. in both		
321478	88	NA	NA	NA	Not DE		100% ident.
437002	88	166048	8	True	Downreg. in both		

321579	88	99634	8	False	Upreg. in both		
515034	88	1213516	8	False	Upreg. in both		
515036	88	NA	NA	NA	Downreg. In StNY001		100% ident.
437056	88	1373890	8	True	Upreg. in both		
531348	88	35486	8	False	Downreg. in both		
321591	88	98745	8	False	Upreg. in both		
515039	88	166039	8	True	Upreg. in both		
515041	88	64978	8	True	Upreg. in both		
436999	88	181806	8	False	Upreg. in both		
436989	88	98996	8	False	Downreg. in both		
531354	88	NA	NA	NA	Not DE		
436988	88	1213210	8	True	Not DE		
322751	88	24038	8	False	Upreg. in both		
551049	88	24037	8	False			
438703	106	166067	8	False	Upreg. in both		
438701	106	98284	8	False	Upreg. in both		
551669	106	157589	8	True	Upreg. in both		
598429	106	24076	8	False	Upreg. in both		
31693	106	174347	8	False	Not DE		
492242	106	99312	8	False	Not DE		
541259	106	157592	8	True	Not DE		
531925	106	98919	8	False			
531926	106	NA	NA	NA	Not DE		100% ident.
31742	106	1407072	8	True	Upreg. in both		
516517	106	1214254	8	True	Not DE		
516519	106	1214257	8	False	Not DE		
551675	106	98552	8	True	Downreg. in both		
31786	106	NA	NA	NA	Not DE		100% ident.
531930	106	98394	8	False	Downreg. in both		
516527	106	143439	8	True	Upreg. in both		
516528	106	44545	8	False	Not DE		
551679	106	157599	8	False	Downreg. in both		
438719	106	98879	8	False	Upreg. in both		
531935	106	24087	8	False	Upreg. in both		
589504	106	123686	8	False			
598444	106	1386096	8	True	Downreg. in both		
438717	106	157605	8	True	Downreg. in both		
516547	106	185136	101	False	Not DE		
516554	106	NA	NA	NA	Not DE		
32102	106	99074	8	False	Not DE		
551685	106	98510	8	False			
551686	106	166083	8	False	Upreg. in both		
516567	106	NA	NA	NA	Downreg. In StNY001		100% ident.
551687	106	24094	8	False	Upreg. in both		
516570	106	166085	8	True	Upreg. in both		

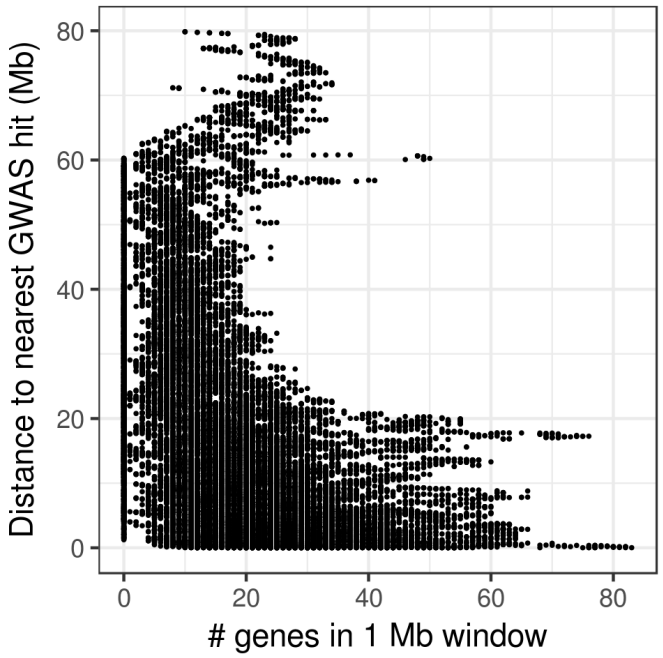
551689	106	44555	8	False	Downreg. in both		
438746	106	166088	8	False	Downreg. in both		
32338	106	1214577	8	False	Not DE		
531947	106	99152	8	False	Downreg. in both		
32417	106	1373971	8	False	Downreg. in both		
46272	116	98807	8	True	Not DE		
439732	116	166116	8	False	Upreg. in both		
551990	116	143521	8	False	Downreg. in both		
517331	116	166115	8	False	Downreg. in both		
589729	116	166114	8	False	Downreg. in both		
532193	116	44572	8	False	Downreg. in both		
46569	116	166112	8	False	Downreg. in both	SSP	
439679	116	157637	8	False	Downreg. in both		
439688	116	166110	8	True	Downreg. in both		
551997	116	174378	8	False	Downreg. in both		
541508	116	166106	8	True	Upreg. in both		
492995	116	123734	8	False			
439682	116	157628	8	False	Downreg. in both		
552002	116	98753	8	True	Downreg. in both		
439694	116	99403	8	False	Upreg. in both		
46787	116	35553	8	False	Upreg. in both	SSP	
517350	116	NA	NA	NA	Not DE	SSP	
439710	116	24117	8	False		SSP	
46809	116	35551	8	True	Not DE		
552007	116	166100	8	False	Upreg. in both		
517354	116	24114	8	False	Upreg. in both		
598695	116	99548	8	True	Not DE		
46856	116	98875	8	False	Upreg. in both		
439741	116	174368	8	False	Downreg. in both		
439701	116	157617	8	True	Not DE		
517357	116	166097	8	True	Not DE		
517358	116	NA	NA	NA	Not DE		
493030	116	98685	8	False	Upreg. in both		
532212	116	35543	8	True	Not DE		
532213	116	35542	8	False	Not DE		
409242	116	35541	8	True	Not DE		
608217	116	24104	8	False	Upreg. in both		
517375	116	35540	8	True	Not DE		
517376	116	35539	8	True	Not DE		
532216	116	24103	8	False	Not DE		
599080	137	NA	NA	NA	Upreg. in StNY001		100% ident.
518706	137	33410	5	True	Upreg. in both		
71115	137	94846	5	True	Upreg. in both		100% ident.
590081	137	1405207	5	False	Upreg. in both		
518715	137	164559	5	False	Upreg. in both	SSP	
541881	137	169072	19	True	Not DE		
441004	137	169073	19	False	Downreg. in both		

71389	137	28307	19	False	Upreg. in both		
71408	137	NA	NA	NA	Not DE		100% ident.
441030	137	37204	12	True	Not DE		
518774	137	NA	NA	NA	Not DE	SSP	100% ident.
71541	137	NA	NA	NA	Not DE		100% ident.
560104	137	1441393	12	True	Not DE		
71558	137	NA	NA	NA	Downreg. in both		
560110	137	182963	12	True	Not DE		
519568	156	24141	8	False	Upreg. in both		
552897	156	123763	8	False	Downreg. in both		
552898	156	99372	8	False	Downreg. in both		
519571	156	35573	8	True	Not DE		
519577	156	166120	8	True	Upreg. in both		
519579	156	NA	NA	NA	Not DE		100% ident.
442098	156	99451	8	False			
519583	156	123774	8	False	Downreg. in both		
542189	156	35577	8	False	Not DE		
469476	156	166124	8	True	Downreg. in both		
87908	156	98578	8	False	Downreg. in both		
519594	156	98254	8	False	Downreg. in both		
519596	156	166125	8	True	Upreg. in both		
519595	156	NA	NA	NA	Not DE		100% ident.
552907	156	123785	8	False	Downreg. in both		
87927	156	143554	8	False	Upreg. in both		
442058	156	24154	8	False	Upreg. in both		
553455	198	166132	8	False	Upreg. in both		
118706	198	99395	8	True	Not DE		
414282	198	98256	8	True	Upreg. in both		
533444	198	24167	8	False	Upreg. in both		
443588	198	174401	8	False	Upreg. in both		
521065	198	166133	8	True	Upreg. in both		
553461	198	99250	8	False	Not DE		
118759	198	35592	8	False	Upreg. in both		
443577	198	98410	8	False	Upreg. in both		
590717	198	123805	8	False	Upreg. in both		
521071	198	NA	NA	NA	Not DE	SSP	100% identity in coding region, downstream polymorphism
444082	219	1390250	27	True	Not DE		
470764	219	152159	27	True			
140283	219	185070	27	False	Not DE		
533553	219	40207	27	True	Not DE		
521495	219	40206	27	False	Downreg. in both		
565643	219	1452275	27	True	Not DE		
394240	219	NA	NA	NA	Not DE		
521501	219	NA	NA	NA	Not DE		

444081	219	NA	NA	NA	Not DE		
404769	219	NA	NA	NA	Not DE		100% ident.
533599	232	39771	21	True	Not DE		
533600	232	NA	NA	NA	Not DE		Potential SSP? Secretion signal by 3/5 Nns
522177	245	NA	NA	NA	Not DE		
566873	245	112015	47	True	Downreg. in both		
444525	245	161957	68	True	Not DE		
566884	245	55644	93	False	Not DE		
496334	245	71967	9	True	Not DE		
388723	245	NA	NA	NA	Not DE		Potential SSP? Secretion signal by 3/5 Nns
533651	245	NA	NA	NA	Not DE		
444911	281	NA	NA	NA	Not DE		
444910	281	NA	NA	NA	Not DE		
444905	281	70449	25	True	Not DE		
444909	281	NA	NA	NA	Not DE		
522697	281	NA	NA	NA	Not DE		
444908	281	NA	NA	NA	Not DE		
599989	281	NA	NA	NA	Not DE		
554084	337	61335	8	True	Upreg. in both		
523198	337	166135	8	True	Upreg. in both		
189245	337	166136	8	False	Upreg. in both		
471237	337	98269	8	False	Upreg. in both		



Supplementary Figure S2.1: *S. turcica* gene models in gene-sparse regions do not tend to have higher estimated rates of dN/dS in St28A (left) or StNY001 (right). All gene models were binned by distance to nearest proximal gene in both the 3' and 5' directions. Color indicates geometric mean of log10(dN/dS) for all genes in that distance bin relative to predicted orthologs in related pathogenic fungi in Pleosporales.



Supplementary Figure S2.2: Distance from a given maize gene model to the nearest GWAS hit for NLB resistance is strongly influenced by the gene density in the region surrounding that gene model. Each point signifies a single maize gene model.

CHAPTER 3

DEEP LEARNING FOR AERIAL DETECTION OF NORTHERN LEAF BLIGHT

CHAPTER 3.1: IMAGE SET FOR DEEP LEARNING[3]

ABSTRACT

Objectives

Automated detection and quantification of plant diseases would enable more rapid gains in plant breeding and faster scouting of farmers' fields. However, it is difficult for a simple algorithm to distinguish between the target disease and other sources of dead plant tissue in a typical field, especially given the many variations in lighting and orientation. Training a machine learning algorithm to accurately detect a given disease from images taken in the field requires a massive amount of human-generated training data.

Data description

This data set contains images of maize (*Zea mays* L.) leaves taken in three ways: by a hand-held camera, with a camera mounted on a boom, and with a camera mounted on a small unmanned aircraft system (sUAS, commonly known as a drone). Lesions of northern leaf blight (NLB), a common foliar disease of maize, were annotated in each image by one of two human experts. The three data sets together contain 18,222 images annotated with 105,705 NLB lesions, making this the largest publicly available image set annotated for a single plant disease.

3 Wiesner-Hanks T, Stewart EL, Kaczmar N, DeChant D, Wu H, Nelson RJ, Lipson H, Gore MA. 2018. Image set for deep learning: field images of maize annotated with disease symptoms. BMC Research Notes 11:440.

OBJECTIVE

Globally, plant diseases are an enormous burden to farmers. Northern leaf blight (NLB), a foliar disease of maize, has become increasingly severe in the US (1). Screening a large area for early symptoms is time-consuming, and there is high intra- and inter-rater variation in NLB severity estimates (2).

Automated, field-based detection of plant disease symptoms would be valuable for plant breeders and growers. However, this is complicated by the “noisy” nature of field imagery. There may be many sources of dead tissue, along with obscured symptoms. This requires a computer vision approach that is specific to the target disease and insensitive to such variations.

Convolutional neural networks (CNNs) are a class of machine learning models that can be trained to accurately detect objects in images, making them the current standard for object recognition (3). CNNs must be trained on a large number of classified or annotated images, but unlike recognizable everyday objects, plant disease symptoms require expertise and experience to identify.

Very few large, expert-curated image sets of plant disease exist (4). PlantVillage contains over 50,000 images of numerous crops and diseases (5). However, these were taken with detached leaves on a plain background, and CNNs trained on these did not perform well on field images (6). Other image sets are much smaller (7), or not curated by experts (8).

We collected image data from several platforms and angles to help develop a system for real-time monitoring and phenotyping of NLB in maize fields using drones equipped with CNNs. The resulting data set exceeds 18,000 maize plant images annotated with more than 100,000 NLB lesions, which is the largest collection of images for any one plant disease. These annotated images are expected to be valuable for furthering the development of novel computer vision and deep learning approaches in agriculture.

DATA DESCRIPTION

The data consists of three image sets and their accompanying annotations. All images were taken in field trials of maize that had been inoculated with *Setosphaeria turcica*, the causal agent of NLB. All trials were planted at Cornell University's Musgrave Research Farm in Aurora, NY (https://cuaes.cals.cornell.edu/farms/musgrave-research-farm/). The trials consisted of maize hybrids from The Genomes to Fields Initiative (https://www.genomes2fields.org/resources/), arranged in two-row plots with a length of 5.64 m and inter-row spacing of 0.76 m. There was a 0.76 m alley at the end of each plot. The trials were rainfed and managed with conventional maize cultivation practices. Plants were inoculated at the V5–V6 stage with both a liquid suspension of *S. turcica* (isolate NY001) spores and sorghum grains colonized by the fungus as previously described (9). The first image set, the "handheld set," was taken by hand in summer 2015. This image set was described and analyzed previously (9), but is included here to make all images available in a single repository. The second, the "boom set," was taken by mounting the camera on a 5 m boom in summer 2015. This boom held the remotely triggered camera above the canopy with nadir view. The third, the "drone set," was taken by mounting the camera on a DJI Matrice 600 sUAS in summer 2017. The drone was flown at an altitude of 6 m and a velocity of 1 m/s, and images were captured with nadir view every 2 s.

For the handheld and boom sets, images were checked manually to ensure the image was in focus and otherwise adequate. For the drone set, images with a low total length of edges (as reported by canny edge detection) were filtered out, in order to remove blurry images. Images were then discarded during annotation if they were out of focus or otherwise unacceptable.

In each image, lesions were annotated by one of two human experts, as denoted in the annotation files. Annotators drew a line down the main axis of each lesion visible in the image, stretching down the entire length of the lesion. If a lesion appeared bent or curved from the camera's perspective, two or more intersecting annotation lines were drawn to form an angle or arc as needed. In the handheld set, this was done with the markup tools in Bisque (9). In the boom and drone sets, these steps were done using a custom ImageJ macro (**Table 3.1**, lesionCount_v2.1_dataNote.txt). Endpoint coordinates of each annotation line are recorded in the three.csv data files, each corresponding to a single data set. Images with 0 values for all four endpoint coordinates had no visible lesions.

The number of images and annotation lines are as follows:

- Handheld: 1787 images, 7669 annotations.
- Boom: 8766 images, 55,919 annotations.
- Drone: 7669 images, 42,117 annotations.

Some boom images are 1/4 slices of larger images, as a wider field of view made it difficult to annotate the entire image at once. These are denoted with suffixes, e.g., 'img01_00.jpg', 'img01_01.jpg.'

Label	**Name of data/file set**	**File type (extension)**	**Data repository and identifer**
Images	images_handheld	.tar.gz (folder with.jpg files)	https://osf.io/arwmy/
	images_boom	.tar.gz (folder with.jpg files)	https://osf.io/er3zb/

	images_drone	.tar.gz (folder with.jpg files)	https://osf.io/vfawp/
Annotations	annotations_handheld	.csv	https://osf.io/7ue84/
	annotations_boom	.csv	https://osf.io/u6mfb/
	annotations_drone	.csv	https://osf.io/25agh/
ImageJ macro	lesionCount_v2.1_data_Note	.txt	https://osf.io/av7dj/

Table 3.1: Overview of data files/data sets.

LIMITATIONS

- Lesion axis annotations do not indicate width or margins.
- There is no way to indicate confidence of annotations. Some lesions are easily visible, while others are partially occluded, out of the main focal plane, in heavy shade, or washed out by bright sunlight.
- Even experts may have a hard time distinguishing between NLB and similar-looking diseases, such as Stewart's wilt or anthracnose leaf blight, from a distance. While no similar-looking diseases were noted as we phenotyped fields on foot, this does not preclude the possibility of such false positives.

- All photographs were taken in a single field in central New York State. This limits the generalizability of the data, as symptoms of the same disease in other regions may present or develop differently.

REFERENCES

1. Mueller DS, Wise KA, Sisson AJ, Allen TW, Bergstrom GC, Bosley DB, *et al.* 2016. Corn yield loss estimates due to diseases in the United States and Ontario, Canada from 2012 to 2015. Pl. Health Prog. 17:211–22.
2. Poland JA, Nelson RJ. 2011. In the eye of the beholder: the effect of rater variability and different rating scales on QTL mapping. Phytopathol. 101:290–8.
3. LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. Nature. 521:436–44.
4. Kamilaris A, Prenafeta-Boldú FX. 2018. Deep learning in agriculture: a survey. Comput. Electron. Agric. 147:70–90.
5. Hughes DP, Salathé M. 2015. An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv 1511.08060.
6. Mohanty SP, Hughes DP, Salathé M. 2016. Using deep learning for image-based plant disease detection. Front Plant Sci. 7:1419.
7. Hallau L, Neumann M, Klatt B, Kleinhenz B, Klein T, Kuhn C, *et al.* 2017. Automated identification of sugar beet diseases using smartphones. Plant Pathol. 67:399–410.
8. Sladojevic S, Arsenovic M, Anderla A, Culibrk D, Stefanovic D. 2016. Deep neural networks based recognition of plant diseases by leaf image classification. Comput. Intell. Neurosci. https://doi.org/10.1155/2016/3289801.

9. DeChant C, Wiesner-Hanks T, Chen S, Stewart E, Yosinski J, Gore MA, *et al.* 2017. Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. Phytopathol. 107:1426–32.

CHAPTER 3.2: MILLIMETER-LEVEL PLANT DISEASE DETECTION FROM AERIAL PHOTOGRAPHS VIA DEEP LEARNING AND CROWDSOURCED DATA[4]

ABSTRACT

Computer vision models that can recognize plant diseases in the field would be valuable tools for disease management and resistance breeding. Generating enough data to train these models is difficult, however, since only trained experts can accurately identify symptoms. In this study, we describe and implement a two-step method for generating a large amount of high-quality training data with minimal expert input. First, experts located symptoms of northern leaf blight (NLB) in field images taken by unmanned aerial vehicles (UAVs), annotating them quickly at low resolution. Second, non-experts were asked to draw polygons around the identified diseased areas, producing high-resolution ground truths that were automatically screened based on agreement between multiple workers. We then used these crowdsourced data to train a convolutional neural network (CNN), feeding the output into a conditional random field (CRF) to segment images into lesion and non-lesion regions with accuracy of 0.9979 and F1 score of 0.7153. The CNN trained on crowdsourced data showed greatly improved spatial resolution compared to one trained on expert-generated data, despite using only one fifth as many expert annotations. The final model was able to accurately delineate lesions down to the millimeter level from UAV-collected images, the finest scale of aerial plant disease detection achieved to date. The two-step approach to generating training data is a promising method to streamline deep learning approaches for plant disease detection, and for complex plant phenotyping tasks in general.

4 Wiesner-Hanks T, Wu H, Stewart E, DeChant C, Kaczmar N, Lipson H, Gore MA, and Nelson RJ. 2019. Millimeter-level plant disease detection from aerial photographs via deep learning and crowdsourced data. Frontiers in Plant Science. doi:10.3389/fpls.2019.01550.

INTRODUCTION

Machine learning models for object detection require a large amount of training data, typically generated by humans. When the average person can identify the feature or object in question, such as a face, a stop sign, or an apple, these data can be generated through crowdsourcing, as was done for large datasets such as ImageNet (Deng *et al.* 2009) and Microsoft COCO (Lin *et al.* 2015). Even if the feature is unfamiliar to most people, crowdsourcing may be viable if the task is simple and the feature obvious. In a recent study on best practices for crowdsourcing plant feature annotation, Zhao *et al.* (2018) found that, with minimal instruction, anonymous online workers could accurately identify maize male flowers in images where they were clearly visible. Accurate identification of many plant features requires a certain level of expertise, however. If only a handful of human experts are qualified and willing to generate training data, the process takes much longer than if tasks could be reliably performed by hundreds or thousands of non-experts. This places a burden on those experts and creates a bottleneck in the model training process.

This dilemma has been addressed by many groups, particularly in the field of human medicine, wherein a model trained on low-quality data could endanger lives, but experts' time is limited and expensive. Different circumstances allow for distinct solutions to the problem. For some tasks, such as interpreting X-ray radiographs, large amounts of training data are already generated and archived under normal protocols, and these data can be used as is without need for additional annotations (Gale *et al.* 2017). When untrained workers perform moderately well, but not quite on par with experts, their annotations can be used to train a "first pass" model that identifies regions of interest (Park *et al.* 2017), or one that performs only those tasks that non-experts can do well (Heim *et al.* 2018). Researchers might have access to a community of knowledgeable, enthusiastic amateurs, such as those who enjoy

identification of birds (Van Horn *et al.* 2015) or aircraft (Maji *et al.* 2013). If nothing but expert annotations will suffice, data sharing lessens the burden on any one group. Multiple groups have used the International Skin Imaging Collaboration imageset of human skin diseases (Codella *et al.* 2015; Haenssle *et al.* 2018) or the PlantVillage imageset of plant diseases (Mohanty *et al.* 2016, Ghosal *et al.* 2018, Picon *et al.* 2018).

Identifying plant diseases via machine learning presents two challenges that limit the feasibility of the above solutions. First, qualified expert judgment is needed at some point in the annotation process, since there are often many causes for tissue death (e.g., disease, abiotic stress, physical damage, natural senescence) and the average person has no experience distinguishing among these. Second, there are hundreds of economically important plant diseases, each with unique considerations of host tissue appearance, plant architecture, symptomatology, etc. A group aiming to implement machine learning detection of a given disease for the first time will likely have to generate novel training data.

The identification of plant disease symptoms in an image might belong to one of three classes of tasks, per Liu *et al.* (2018): classification, detection, or segmentation. Object classification methods detect the presence or absence of features within an image on the whole, e.g., “this is an image of wheat stem rust.” Object detection methods identify the location and extent of symptoms within an image on a coarse spatial level, most commonly delineating them with bounding boxes. Semantic segmentation methods delineate the boundaries of features, assigning each pixel of an image to a given class, e.g. leaf, soil, or disease symptom. In this paper, we undertake this last task- identifying and outlining every diseased region in an image.

Aerial plant disease detection via machine learning has aroused much interest in the past few years, as evidenced in many reviews, letters, and prospectives (Singh *et al.* 2016, Tsaftaris *et al.* 2016, Shakoor *et al.* 2017, Yang *et al.* 2017, Araus *et al.* 2018, Ubbens and Stavness 2018, Maes and Steppe 2019, Chouhan *et al.* 2019). Compared to the level of interest, relatively few examples have been published. Machine learning classification has been used to classify entire plants as virus-infected or not (Ha *et al.* 2017, Sugiura *et al.* 2018). Object detection methods have been used to identify diseased regions of grape plants (Kerkech *et al.* 2018) and diseased leaves of soybean (Tetila *et al.* 2017). Semantic segmentation of UAV images, the task we undertake here, has been implemented in soybean (Tetila *et al.* 2019), tea plants (Gensheng *et al.* 2019), and maize (Stewart *et al.* 2019).

In the course of our previous work, we labeled over 100,000 examples of northern leaf blight (NLB), a fungal foliar disease of maize that causes grey-brown necrotic lesions (Wiesner-Hanks *et al.* 2018). Each of these annotations consisted of a line drawn down the principal axis of a lesion. With these line annotations, we trained convolutional neural networks (CNNs) to recognize NLB lesions in images taken by hand with 96.7% accuracy (DeChant *et al.* 2016) and in aerial field images with 95.0% accuracy (Wu *et al. 2019*). Delineating lesion boundaries with polygons would be ideal, as such annotations can ultimately yield much more precise image segmentation than lower-resolution annotations (Bell *et al.* 2015). Drawing such polygons is prohibitively time-consuming to do with only a small number of trained experts, however.

In this study we describe and implement a two-step approach for generating large amounts of high-resolution training data that has been vetted by qualified experts. First, experts identify disease symptoms, annotating them quickly at low resolution. Second, the more time-consuming task of annotating the lesion boundaries is outsourced to anonymous online workers through Amazon’s

Mechanical Turk platform. This two-step approach allows us to maintain the reliability of expert diagnosis while also exploiting the speed and scale of crowdsourcing, producing a model with high accuracy and spatial resolution (**Figure 3.1**) with only one fifth as many expert-generated annotations.

Figure 3.1. Examples of lesion segmentation on original images taken in the field by UAV. Regions classified as disease lesions by model outlined in magenta.

MATERIALS AND METHODS

Image annotation

All Mechanical Turk human intelligence tasks (HITs) consisted of one or more prompts to draw a single bounding polygon delineating the boundaries of a single lesion (**Figure 3.2**, top right), previously annotated with a line down the major axis by one of two human experts (Wiesner-Hanks *et al.* 2018). All images and annotations used, generated, or described herein are available in an Open Science Framework repository (https://osf.io/p67rz).

For each annotated lesion, a subimage was taken of the same width and height of the annotation line, plus 150 pixels padding on all four sides, so that workers had some context to the image. The annotation lines mostly spanned 400-1200 pixels in the x- and y-dimensions (depending on orientation), so this padding usually expanded the field of view by 25-75%. Workers were given basic instructions asking them to draw a polygon delineating the edges of the necrotic lesion with between 10 and 15 vertices, along with an example lesion thereof (Supplemental text, **Supplemental Figure 3.1**). The annotation lines drawn by experts were included in these subimages in red to make clear which lesion to annotate, as there was often more than one lesion in a single subimage.

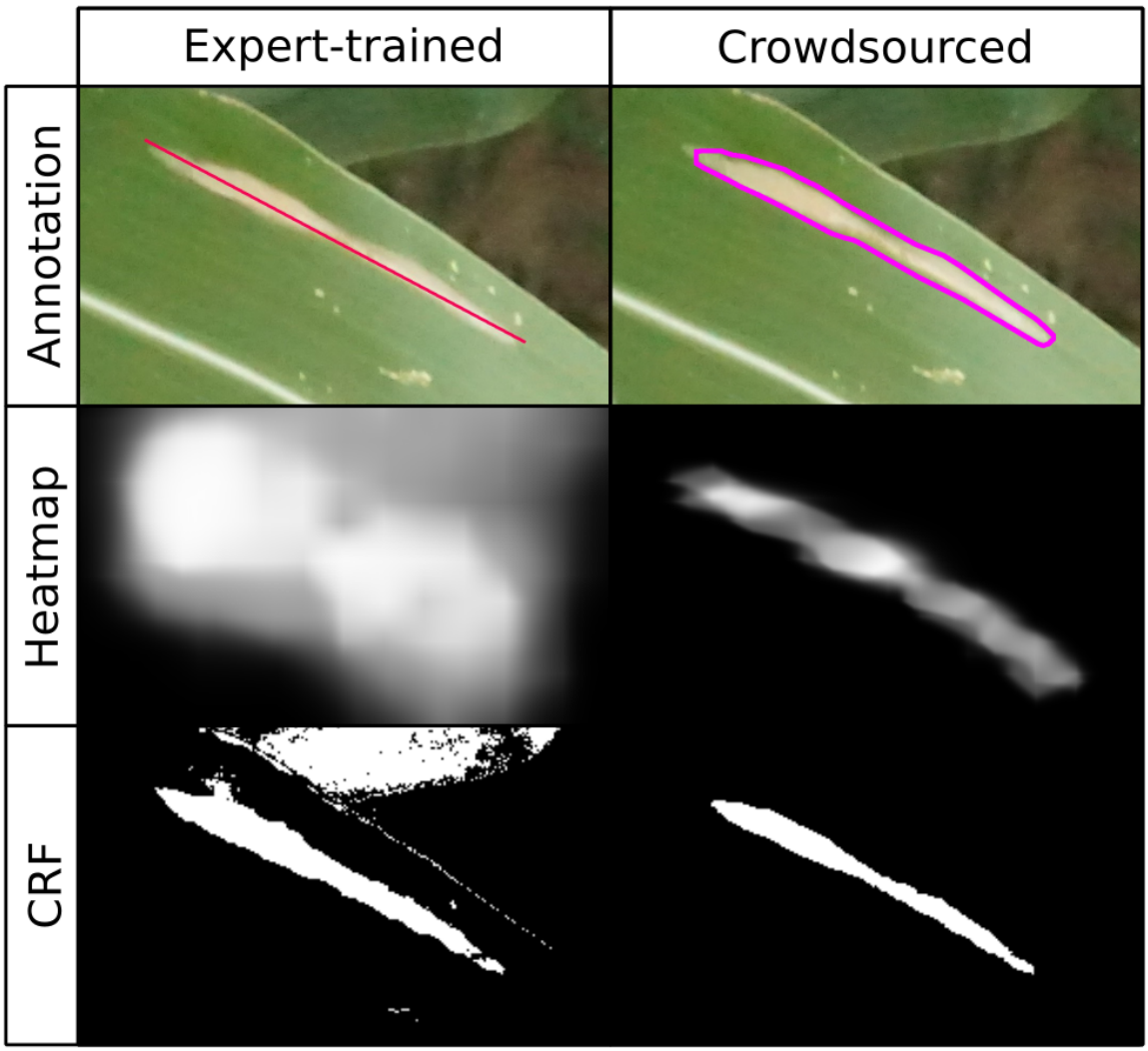


Figure 3.2. Comparison of annotations used and results of expert-trained model (left; Wu et al. in review) and crowdsourced model described here (right). Top row: Original image with annotations overlaid. Middle row: Heatmap created by applying CNN in sliding window across image, brightness indicating probability of lesion at a given point (white = lesion, black = non-lesion). Bottom row: Binary mask output of CRF segmentation using original image and heatmap.

HITs were deployed in three batches over the course of two months. Three unique workers were assigned to complete each HIT. Each worker was paid $0.03/HIT, an amount chosen to be similar to payment for comparable tasks deployed on Mechanical Turk at that time, adjusting for the fact that different HITs involve a different number of tasks. An additional $0.01 was paid to Amazon each time a worker completed a HIT, resulting in a total cost of $0.12/lesion (three workers per lesion, $0.03 per worker, $0.01 to Amazon per worker).

Annotations drawn by MTurk workers were first screened to see how much they agreed with the other annotations drawn on the same lesion. If a given worker drew polygons that rarely agreed with those drawn by other workers, their annotations were potentially suspect. After a batch was completed, the Intersection over Union (IoU), also called the Jaccard similarity, was calculated for each pair of polygons drawn on the same lesion by taking the area in pixels of their intersection divided by the area in pixels of their union. Each polygon was thus compared to the two other polygons drawn by other workers on the same lesion. If the mean Jaccard similarity between all annotations drawn by a given worker and those drawn by other workers was < 0.5, the worker was flagged for manual review. This threshold was set at 0.5 because the vast majority of workers had overall mean IoUs in the 0.5-0.8 range, while a small number, who mostly completed only a handful of HITs each, had mean IoUs in the 0-0.5 range (**Supplemental Figure 3.2**). Manual review was deemed necessary, since a worker may have drawn their high-quality annotations compared to low-quality annotations. If their work was found to be unacceptable, all of their annotations were rejected and lesion subimages were redeployed as needed until three unique workers had acceptably annotated each. In all cases, workers whose annotations were rejected appeared to be drawing polygons at random.

The IoU was also used to filter out low-quality lesions. Entire UAV images were filtered automatically prior to annotation and manually during annotation, as described previously (Wiesner-Hanks *et al.* 2018), but individual lesions in an image could still vary in how clear and defined the visible symptoms were. Preliminary manual inspection of MTurk annotations revealed that lesions on which otherwise well-performing workers drew lesions with low overlap with one another were often blurry, ambiguous, or otherwise unacceptable. Only lesions for which all three polygon annotations had an IoU > 0.6 with one another (a threshold chosen to filter out roughly the bottom 25% of lesions) were used to generate images for model training as described below. The mean and standard deviation of pixel RGB values, used for later normalization of images, were calculated on these whole images.

Training, validation, and test data were generated based on the method used with polygon annotations in the OpenSurface dataset (Bell *et al.* 2015). Multiple square subimages, hereafter referred to as "patches," were cropped from the entire UAV image and classified as "lesion" or "non-lesion" based on whether the exact center point of the image lay within a lesion. To generate positive patches (the "lesion" class), pixels lying within at least two of three annotation polygons were used as a search space. From these, random points were sampled via Poisson-disk subsampling (scipython.com/blog/poisson-disc-sampling-in-python/), with minimum distance of 200 pixels between each point. Negative patches (the "non-lesion" category), were chosen by randomly sampling points from the pixels in each image that were not included in any of the annotation polygons. Negative training images thus could contain a lesion, so long as they were not centered on one.

Because the original UAV images consisted mostly of non-lesion area, many more non-lesion patches could be extracted from the images than lesion patches. Preliminary model testing with sample images suggested that using a balanced dataset with an equal number of lesion and non-lesion patches

biased model predictions towards false positives, i.e. detecting lesions where there were none (data not shown). We thus used a moderately unbalanced dataset and accounted for the class imbalance using weights in the loss function, as described below.

In order to make the model more generalizable, training images were augmented via random transformations that preserved the image class, i.e. the location of the central pixel in a lesion or not. Images were horizontally and/or vertically flipped, rotated by 0 to 90 degrees either clockwise or counterclockwise, and scaled between 0.75x and 1.33x. As these images were taken from directly overhead, there was no need to preserve image orientation.

Network construction

We used a ResNet34 model (He *et al.* 2016) that had been pre-trained on the ImageNet dataset of several million labeled images (Russakovsky *et al.* 2015) as a generalized feature extractor, replacing the final fully-connected layer with a fully-connected layer of output dimension 2. The output tensor for each input image was a two-dimensional vector of scores for the two classes: centered on a lesion or not centered on a lesion (note that images containing a lesion but not exactly centered on it belong to this second class). A weighted cross-entropy loss function was used, which normalizes the scores into estimated probabilities via the softmax function, then takes the negative log of these probabilities and multiplied by the class weights to account for class imbalance. Class weights of 0.36 and 1.0 were used for lesion and non-lesion images, proportional to the number of images in each class. In order to determine which patch size and learning rate was most appropriate, we analyzed performance on a smaller sample set of images. For both image classes (lesion and no lesion), 5% of the training and validation sets were randomly sampled. The above network was trained and validated

on this 5% subsample with six patch sizes (square patches of size 200, 400, 500, 600, 800, or 1000 pixels, using the same centerpoints for each size) and seven learning rates (1e-5, 3e-5, 1e-4, 3e-4, 1e-3, 3e-3, and 1e-2). With each combination of patch size and learning rate, the network was trained for 10 epochs with a step size of 10 and gamma of 0.1, corresponding to a 10-fold decrease in the initial learning rate every 10 epochs.

The best-performing parameters were then used to train the network on the entire training and validation set for 20 epochs with a step size of 10 and gamma of 0.1. Patches were resized to 224 by 224 pixels and treated with a random horizontal flip, then normalized using the previously calculated mean and standard deviation of pixel RGB values. To compare learning rate dropouts, the model was also trained using step sizes of 5 and 20, maintaining a gamma of 0.1. Weights were optimized using stochastic gradient descent with weights of 1.0 and 0.36 for the non-lesion and lesion labels, respectively, proportional to the number of images in each category. All training was done on an Nvidia GTX 1070 Ti GPU with batch size 120, randomizing image input order.

To visualize the model-estimated probability of a given region containing a lesion or not, heatmaps were generated by applying the final CNN on a sliding window across whole UAV images, then applying softmax transformation to generate probabilities for the two classes (centered on a lesion or not). To account for varying lesion sizes, we used the resizing approach of Bell *et al.* (2015). The image was resized by three separate scaling factors: the original scale r used in model training (such that a 500x500 window was resized to 224x224 pixels), r*sqrt(2), and r/sqrt(2). At these scales, a window of size 500x500, 690x690, or 345x345 pixels, respectively, mapped to 224x224 pixels. Images were padded on all sides via reflectance padding, and the trained model was applied via a sliding window approach across the entire image with a stride of 50 pixels in both dimensions. The resultant output was

then resized to the original 4000x6000 pixels via bilinear interpolation. The three resultant heatmaps were then averaged, and this averaged heatmap was used for downstream analyses. For comparison, the trained model described by Wu *et al.* (2019) was applied in an identical manner. As the scaling used for training purposes was identical between these two models, the same scales were used for heatmap generation.

Image segmentation

Pixel-wise classification was performed using the fully-connected conditional random field (CRF) method of Krahenbuhl and Koltun (2011), implemented in Python via pydensecrf. CRF optimization was performed using three separate color spaces: the original, untransformed RGB values, RGB values transformed to maximize contrast between lesion and non-lesion pixel values, and L*a*b* color space. For the second method, the pixels surrounding each polygon annotation were found by dilating the polygon mask (expanding the mask along its edges to include pixels for which a kernel overlaps with the mask) for five iterations using a 20 pixel by 20 pixel square kernel, then subtracting the area created by performing only one dilation of the mask. The RGB values of pixels within these regions and those lying within polygon annotations were then downsampled by a factor of 10 and analyzed via linear discriminant analysis (LDA) to obtain a transformation maximizing between-group differences in Euclidean distance between values in the two regions. RGB to L*a*b* transformation was performed using OpenCV, producing 0-255 integer-valued L*a*b* coordinates.

CRF performance is controlled the θ parameters, which determine how strongly pixel classification is influenced by proximity (is it close to many pixels believed to be NLB lesions?) and color (is it the same color as pixels believed to be NLB lesions?). Because optimizing these is difficult

(Krahenbuhl and Koltun 2011), we used a simple grid search to find suitable parameters, evaluating CRF performance for all combinations of θ values on a set of 118 training images. These were selected from the entire set of training images by choosing images in which the annotation polygons of all three workers agreed fairly well (each one having IoU > 0.8 with the union of all three, a cutoff chosen to be fairly stringent) for all lesions in the image. CRF performance on each image was evaluated under each color space with slightly different parameters, as appropriate for each. For the RGB and LDA-transformed color spaces, the kernel width θα, corresponding to the spatial dimension of pixel correlation and deviation, was evaluated at values ranging from 10 to 600 by a step size of 10. For the untransformed RGB color space, θβ, corresponding to the color-space correlation and deviation of pixels, was evaluated at values ranging from 1 to 40, step size 1. For the LDA-transformed RGB values, θβ was evaluated at values ranging from 0.1 to 0.4, step size 0.1. For the L*a*b* color space, separate kernel widths were used for the distance along the L dimensions and distance in a-b dimensions. CRF performance was analyzed for θα (still the spatial kernel width, unrelated to the a* color dimension) ranging from 10 to 500 with step size 10, θL ranging from 1 to 25 with step size 1, and θab from 1 to 20 with step size 1. CRF performance on the model of Wu *et al.* (2019) was tested only in the RGB color space.

RESULTS

Mechanical Turk annotations

MTurk workers drew 15,240 polygon annotations on 5,080 lesions, cropped from 752 parent images collected by the UAV. Training data for the CNN were generated only from those images in which, for all lesions in the image, all three polygon annotations had an Intersection over Union (IoU) of at least

0.6 with one another, leaving us with 588 UAV images containing 3,834 annotated lesions. Poisson-disk subsampling of the lesion polygon annotations yielded 22,193 centerpoints that were used to generate 22,193 positive images (**Table 3.2**). From the same 588 UAV images, we sampled 58,800 negative images, 100 from each image. Both positive (centered on a lesion) and negative (not centered on a lesion) images were divided into training, validation, and test sets in a 70:15:15 ratio.

Most workers annotated only a few images, with a small number of workers annotating several hundred (**Supplemental Figure 3.3**). On average, it took an MTurk worker 32 seconds to annotate a single lesion (median 27 seconds, standard deviation 19 seconds). All sets of deployed HITs were fully annotated in under 2 hours. Workers generally performed fairly well, as shown by the fact that any two annotations drawn on the same lesion tended to overlap (**Figure 3.3**). Most pairs of polygons (83.2%) had an IoU of at least 50%. Manual examination found that many of the annotations with low IoU were on images that were blurry, ambiguous, or otherwise undesirable. Workers were paid $0.03/lesion, resulting in an average payment of only $3.75/hour for annotation.

Phase	**Number of images**	
	Lesion	**No lesion**
Training	14,783	41,160
Validation	3,168	8,820
Test	3,168	8,820

Table 3.2: Number of images sampled of each label (lesion vs. no lesion) and their division into training, validation, and test sets.

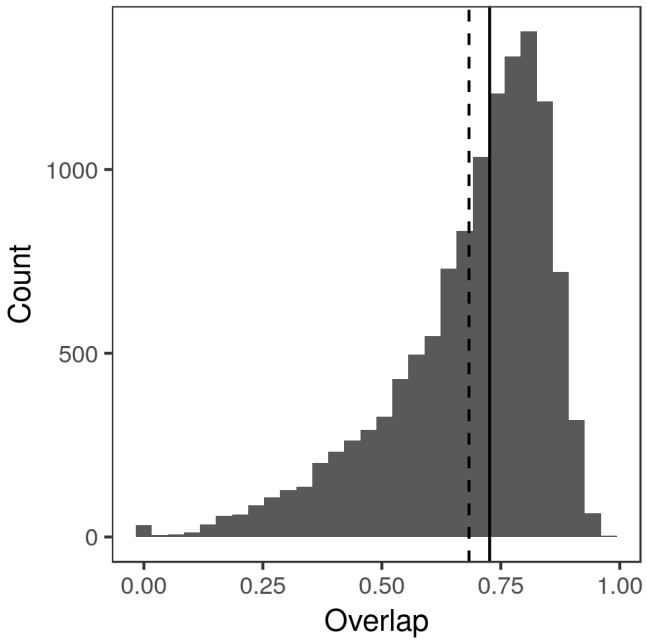


Figure 3.3. Histogram of Intersection over Union (IoU) between all pairs of polygon annotations drawn by MTurk workers, calculated as the area in pixels of intersection divided by the area in pixels of the union. Median IoU (0.7265) indicated by solid line, mean IoU (0.6832) indicated by dotted line.

Model performance

Testing classification accuracy of the crowdsourced CNN on a subsample of training and validation images, we found a learning rate of 3e-3 and a patch size of either 500 or 800 to be best (**Figure 3.4**). Though classification accuracy was slightly higher when using a patch size of 800 compared to a patch size of 500, we chose a patch size of 500 to be consistent with that used in the model trained on expert-drawn-lines (Wu *et al.* 2019) to facilitate comparisons between the two.

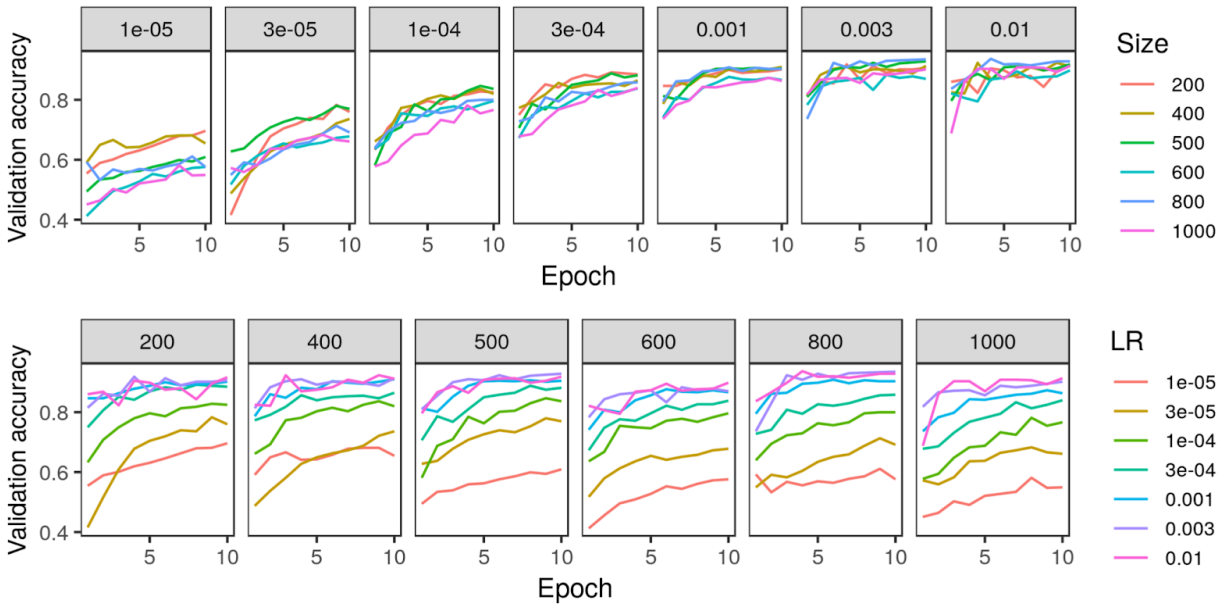


Figure 3.4. Comparison of crowdsourced CNN accuracy on 5% subsample of training/validation images across various parameters of learning rate (LR) and patch size in pixels.

The accuracy of the crowdsourced CNN on the validation set of image crops converged by 15 epochs (**Figure 3.5**). One concern with any machine learning model is the possibility of overfitting: training a model that performs well on the specific data set being used, but that is not generalizable and performs poorly on new data. Loss on the validation set did not tend to increase after reaching a global minimum, suggesting that overfitting was not a major concern (James *et al.* 2013), though the gap between training loss and validation loss suggested some overfitting (**Figure 3.5**). On the final held-out test set of image crops, the crowdsourced CNN performed well, achieving an overall classification accuracy of 0.9741, precision (TP/[TP+FP]) of 0.9351, recall (TP/[TP+FN]) of 0.9694, and F1 (harmonic mean of precision and recall) of 0.9520 (**Table 3.3**).

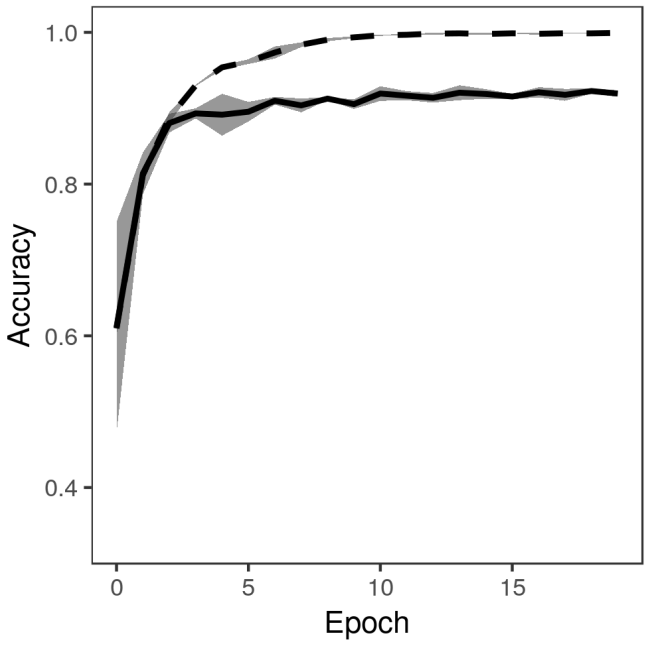


Figure 3.5. Accuracy of crowdsourced CNN on training images (dashed line) and validation images (solid line) converged by 15 epochs. Grey area shows standard deviation of accuracy over 5 replications of training on the same training/validation sets.

Prediction	**Image**	
	Lesion	**No lesion**
Lesion	3,071	213
No lesion	97	8,607

Table 3.3: Predictions of the final network on the held-out test set.

Image segmentation

Applying a fully-connected CRF to the heatmaps generated by the crowdsourced CNN and the held-out test images, we were able to accurately classify each pixel of an image as lesion or non-lesion with high spatial resolution (**Figure 3.2**, bottom row). Pixel-wise classification accuracy was high even when heatmaps were clearly not suitable, as the vast majority of most images is non-lesion, so a model that classified all pixels as non-lesion would still achieve an accuracy of 0.9940. For this reason, F1 was taken to be a more suitable metric for image segmentation performance than accuracy.

Exhaustive grid search found the best-performing θ parameters for each color space to be θα=50, θβ = 5 for the standard RGB color space, θα = 110, θL = 25, and θab = 1 for the L*a*b* color space, and θα=70 and θβ=0.7 for the LDA-transformed color space (**Figure 3.6**). Transforming images into the L*a*b* color space moderately increased segmentation accuracy. The best-performing CRF parameters segmented images with an accuracy of 0.9957 and F1 of 0.6695 in the RGB colorspace, compared to peak accuracy of 0.9977 and F1 of 0.6777 in the L*a*b* color space. Transforming the RGB values using the matrices obtained via LDA was the most effective, yielding a peak accuracy of 0.9981 and F1 of 0.7153. The parameters that segmented LDA-transformed images with the highest F1 score also did so with near-maximum accuracy (**Figure 3.7**).

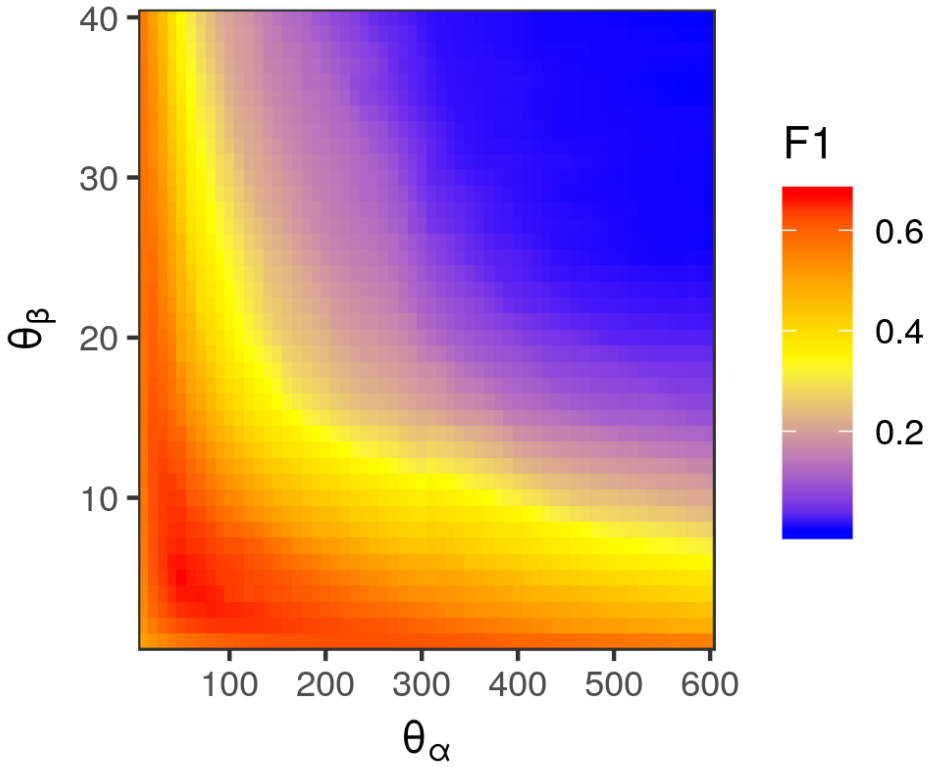


Figure 3.6. Heatmap of pixel-wise F1 score of CRF segmentation across different levels of θ_α, corresponding to the spatial scale of correlations between pixel color values, and θ_β, corresponding to the color space scale of correlations. Values were determined using images transformed with RGB values transformed via LDA-derived differentiation transformation, as this was the color space in which CRF segmentation performed best.

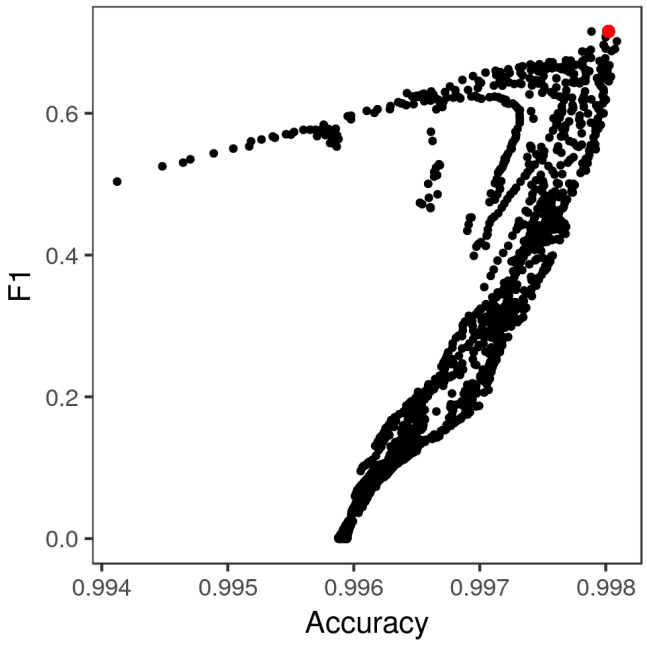


Figure 3.7. Pixel-wise F1 score of lesion/non-lesion segmentation vs. accuracy thereof across different levels of θ_α and θ_β. The CRF parameters that yielded the highest F1 score (red point) also yielded near-maximum accuracy of segmentation. Each point represents a single combination of θ_α and θ_β tested in the grid search (**Figure 3.6**).

CRF segmentations could be used to accurately estimate the proportion of an image covered by lesions (**Figure 3.8**). The proportional lesion coverage estimated by CRF was highly correlated to ground truth estimates. The heatmaps themselves could also be used to estimate proportional lesion coverage in an image, bypassing the CRF step. Thresholding probability heatmaps at 0.5 produced binary images, in which pixels had a value of 1 if the interpolated predicted softmax probability of the "lesion" prediction was higher and a value of 0 if that of "non-lesion" was higher. However, the lesion coverages estimated by CRF segmentation were proportional to the ground truth areas in an

approximately 1:1 manner, while the areas generated from thresholding probability heatmaps were artificially inflated (**Figure 3.8**).

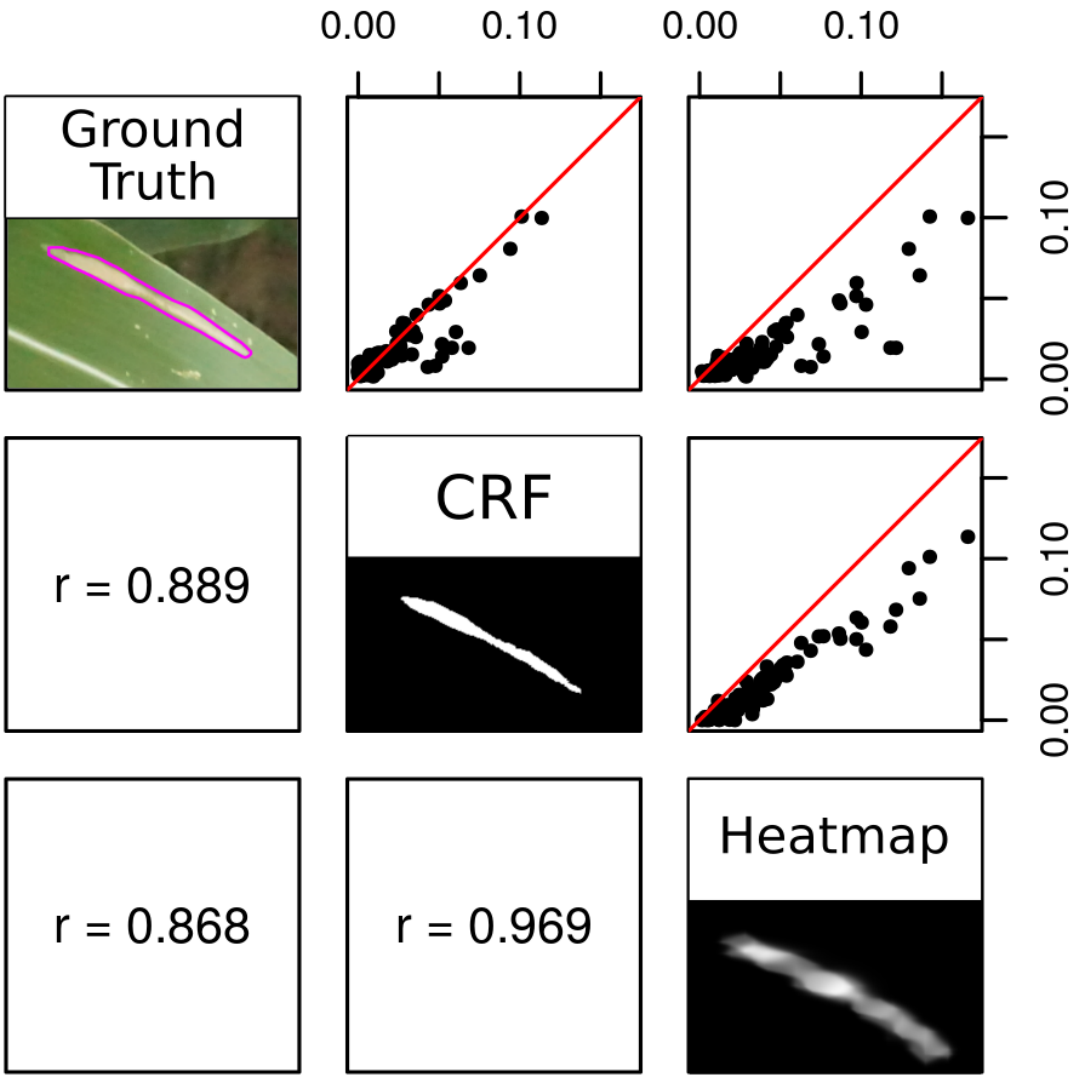


Figure 3.8. Correlation between the proportion of a test image classified as lesion in ground truth (consensus polygons of three high-quality MTurk annotations), CRF segmentation, and heatmap thresholded at 0.5. Red lines depict 1:1 ratio.

Image segmentation using the crowdsourced CNN and CRF tended to outperform human experts. There were seven instances in which the proportion of a test image classified as “lesion” diverged highly between CRF segmentation and ground truth (**Figure 3.8**, outliers lying off of the red 1:1 line). This was surprising, as precision of CRF segmentation was higher than recall (0.7388 vs. 0.6937) on a pixel-wise basis. Examining these seven cases more closely, we found that five of them were due to the model correctly locating lesions missed by the experts, while only two were due to the model misidentifying senescent leaves as lesions (**Figure 3.9**). Excluding the five images in which the CRF outperformed human experts, the Pearson’s correlation between the proportion of pixels in an image labeled as lesions in the ground truth masks and the proportion classified as lesions by the CRF segmentation rose from 0.8893 to 0.9428. Thus, while there is room to improve the model by addressing false positives, it was more often than not outperforming trained human experts.

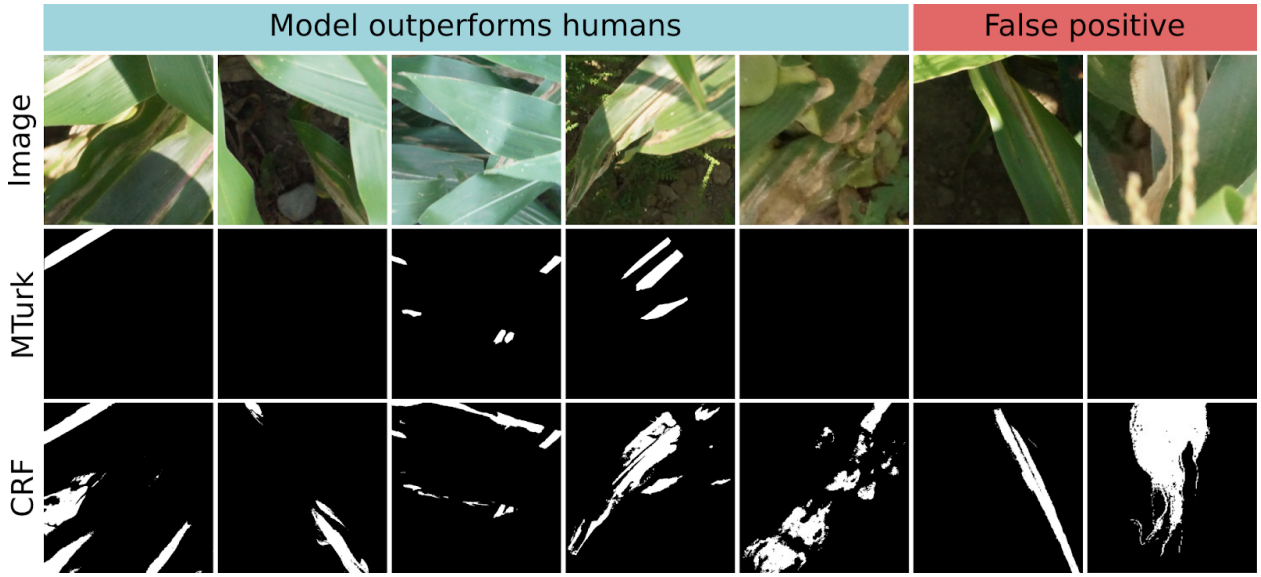


Figure 3.9. Original image (top row), ground truth annotations drawn by MTurk workers (middle row), and CRF segmentation (bottom row) for all seven test images in which CRF segmentation and ground truth diverged highly. In left five images, crowdsourced model outperformed humans by identifying lesionated areas where humans had missed them. In the right two images, the model falsely classified senescent leaf tissue as lesions. White = lesion, black = non-lesion.

The two-step image segmentation process was fairly slow, however. Heatmap construction by the sliding-window approach using three different scales took a mean of 38.1 seconds on a 4000x6000 image: 10.8 seconds at original scale r, 6.2 seconds at scale r/sqrt(2), and 21.1 seconds at scale r*sqrt(2). CRF segmentation of a 4000x6000 image took 2.8 seconds on average. Newer end-to-end segmentation methods should be able to improve on this, as discussed below.

Benefits of crowdsourcing

Using crowdsourced polygon annotations greatly improved the spatial resolution of the final model with far less time investment from experts. We compared two CNNs of similar structure and implementation: one trained on lines drawn by experts (Wu *et al.* 2019) and the one trained on crowdsourced polygons, described here. These models were used to perform semantic segmentation using the same approach, via applying the CNN on a sliding window across images to generate probability heatmaps, then feeding these heatmaps into an optimized CRF to perform the final segmentation. Using the same approach with both model outputs isolated the effects of using the more information-rich crowdsourced polygons, rather than differences in segmentation methods.

Using the crowdsourced annotations provided three key benefits. First, the greater spatial resolution of polygon annotations allowed us to reliably delineate individual lesions with millimeter-level accuracy (**Figure 3.10**), which could not be done with line annotations alone. CRF segmentation using the crowdsourced CNN output was able to segment images into lesion and non-lesion pixels with a maximum F1 of 0.76 on the validation image set, while segmentation using expert-drawn lines achieved a maximum F1 of only 0.21.

Figure 3.10. Image segmentations performed by CRF using heatmaps generated by the crowdsourced CNN. Magenta outline shows lesion boundaries from twelve randomly-selected images in the test set.

Second, the crowdsourced-polygon model was able to achieve this higher spatial resolution using only one-fifth as many annotations. The crowdsourced CNN was trained, validated, and tested on only 5,080 expert-drawn lesions, compared to the 25,508 used for the expert-drawn-lines model (Wu *et al.* 2019).

Third, crowdsourcing allowed us to generate these polygons more quickly than would be possible using only a handful of experts. Drawing a line took far less time than drawing a polygon. Examining the

timestamps of the annotations, we found that experts took 4.38 seconds on average to annotate a lesion with a line, while it took an MTurk worker a mean of 32 seconds to draw a polygon. An expert could thus annotate 1,000 lesions with lines in 73 minutes on average, while a non-expert would take 533 minutes to annotate 1,000 lesions with polygons. The more complex non-expert task thus took 7.3x more time than the simpler expert task. Due to the parallel nature of crowdsourcing, however, all 5,080 lesions were annotated by MTurk workers in less than 15 hours.

A comparison of the total time needed to generate training data at the scale used in this study shows the benefits of a two-step crowdsourcing approach. A single human expert can annotate 5,000 lesions in roughly 6 hours, which could then be completely annotated with polygons by MTurk workers in one to two days. Assuming this expert worked as fast as the average MTurk worker (including locating lesions, which MTurk workers were not required to do), drawing these polygons would take them roughly 44 hours. Crowdsourcing the more laborious part of the task as described here is a more efficient use of plant scientists' time and expertise.

DISCUSSION

Our full method, combining a CNN applied across a sliding window and image segmentation via a fully-connected CRF, was able to identify and delineate disease lesions at the millimeter level, the smallest spatial scale reported so far for aerial plant disease detection. The two-step approach for generating training data, in which experts annotate symptoms in low detail and non-experts annotate them further in high detail, was critical to achieve high spatial resolution. Without the non-expert polygon annotations, our previous effort was able to identify lesions with high accuracy at a sub-leaf scale (Wu *et al.* 2019), but not at sufficient resolution to accurately segment an image and delineate

individual lesions. With them, we were able to segment images down to the millimeter with sensitivity surpassing that of human experts: in five out of seven cases in which human ground truth and model predictions diverged, the model had correctly identified disease symptoms where experts had missed them (**Figure 3.9**).

Using Mechanical Turk, thousands of images could be annotated in only a few hours, reducing what was until then a major bottleneck in the model training process. Despite the fact that these workers (presumably) have no experience in plant disease diagnosis, their annotations were generally of high quality and could be used to train the model without the need for an expert to look over each one. With three annotations for each image, we were able to identify and filter both low-performing workers, whose annotations tended not to agree with others, and low-quality images, on which multiple (otherwise well-performing) workers drew annotations that did not agree. There are several possibilities for improving the MTurk annotation process. Increasing the number of workers per image could increase the quality of annotation polygons or the ease of identifying low-quality images.

The cost of crowdsourcing via MTurk was quite low, at $0.03/lesion, implying a wage of $3.75/hour based on the average time to annotate a lesion. Future studies would ideally compare different payment structures in order to maximize worker payment, minimize overhead, and maintain or increase annotation quality. Restructuring the HIT so that each consists of annotating multiple lesions, rather than just a single lesion, would decrease the payment to Amazon per image while paying workers the same per HIT. Many HITs posted on MTurk require a short qualification test to vet workers. In our case, workers could be asked to annotate three lesions adequately in order to be approved to complete HITs. Increasing worker payment in tandem with this could attract and retain better-performing

annotators, providing workers with a higher wage while decreasing the amount of post-processing needed to filter out low-confidence annotations.

We used a two-step method for semantic segmentation, first training a model to classify lesions, then using a sliding window approach and CRF to turn these classifications into semantic segmentation of a full image. This allowed us to make a useful comparison to a model trained on coarse, expert-generated annotations, since the same segmentation method could be used with both models' output, isolating the impact of the annotation data rather than the segmentation approach used. However, newer methods for semantic segmentation, such as region proposal networks (Ren *et al.* 2016) or atrous convolution (Chen *et al.* 2017) might well perform the task better and faster.

A chief limitation of this method is the difficulty of acquiring field images at high enough resolution and clarity such that individual lesions can be discerned. Capturing images in which each pixel represented a millimeter or less at canopy level required slow flights at low altitude with a high-zoom lens (Wiesner-Hanks *et al.* 2018), not ideal for comprehensively imaging a large area. This challenge would be even greater when working with a disease with small or inconspicuous symptoms- chlorosis, leaf curling, lesions only a few millimeters in diameter- as opposed to the large, obvious lesions of NLB. Targeted sampling of a field, rather than attempting to image every plant, can still give growers a large amount of information with which to make decisions regarding disease management. Acquiring images and diagnosing lesions every 10 meters or so would only analyze a very small proportion of a field's total area, but it would provide much more information compared to the zig-zag walking paths commonly used when scouting for pests and diseases (Doll *et al.* 2016).

UAVs are now a common part of many US growers' field operations, and interest continues to grow (Luck *et al.* 2018, Purdue Extension 2018, Miller and Adkins 2018). The use of UAVs for disease

diagnosis is still in its infancy, however. We predict that UAV-based disease phenotyping will be most readily adopted in crops with a high value per acre where fungicide usage is common, such as grapes or almonds. In such crops, the added benefit of fast, frequent, reliable disease screening is most likely to outweigh the time and monetary costs needed to develop the diagnostic platform. As UAV and imaging technology progress, and more and more image datasets are generated and freely shared among researchers, we believe that UAV-based deep learning will become simpler to implement and will soon be a useful tool for growers and geneticists across many crops and pathosystems.

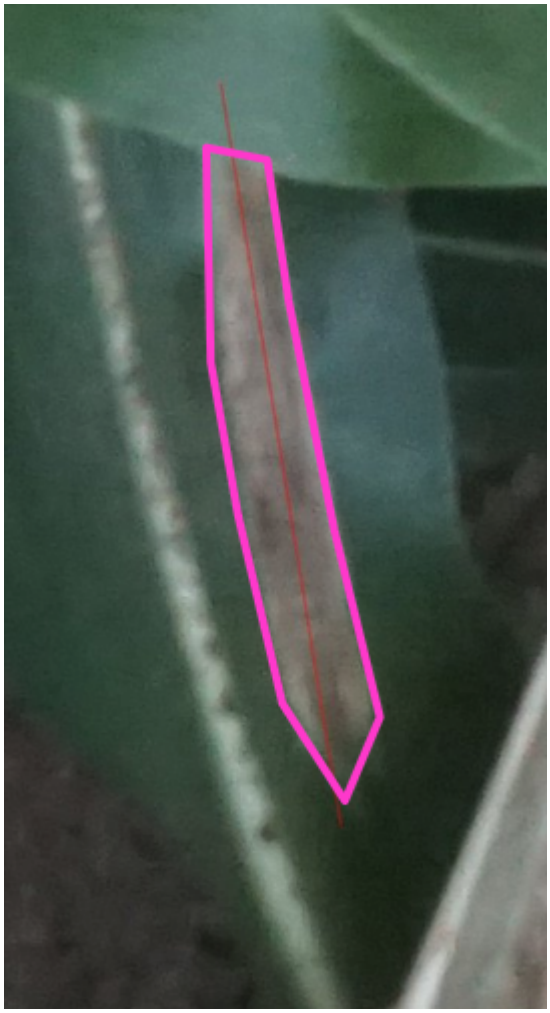
Supplemental text: **worker instructions**

Place points around the edge of the lesion, where the dead brown tissue meets the healthy green tissue, as shown in this example. There may be multiple lesions in the image- only label the one with a red line down the center.

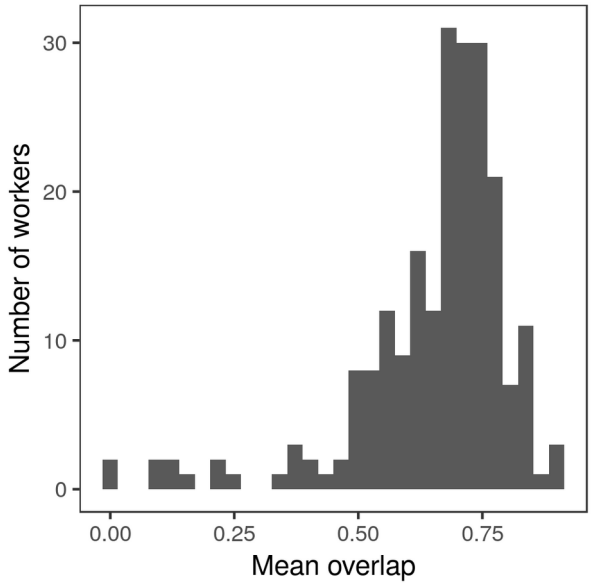
You must place at least 10 points, but no more than 15. When done placing points, right click to close the polygon. *Once you close a polygon, you cannot delete the most recent point*.

To move a point, click "Edit", then drag the point. To delete a polygon, click "Edit", select it, then click "Delete Selected Polygon".

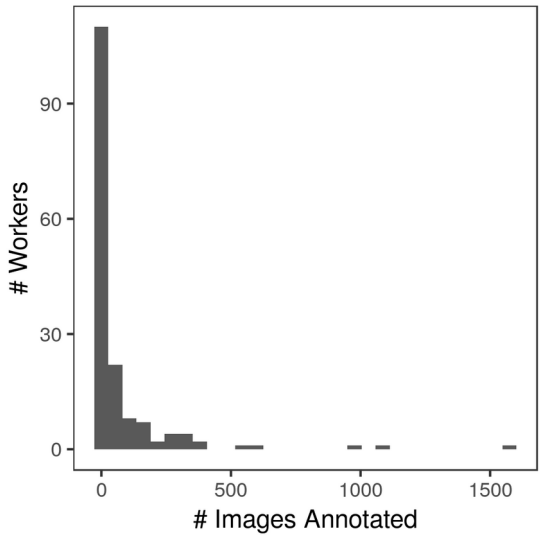
If the image is very blurry, please draw the polygon as best you can and place a comment (e.g. "blurry") in the box below, then submit.



Supplemental Figure 3.1. Example lesion annotation shown to MTurk workers.



Supplemental Figure 3.2. Histogram of mean overlap (Intersection over Union) of all polygons drawn by a worker.



Supplemental Figure 3.3. Histogram of number of lesions annotated by individual MTurk workers.

REFERENCES

Bell, S., Upchurch, P., Snavely, N. and Bala, K. (2015). Material recognition in the wild with the materials in context database. In Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition, 3479-3487.

Chen, L.-C., Papandreou, G., Schroff, F., and Hartwig, A. (2017). Rethinking atrous convolution for semantic image segmentation. ArXiv. arxiv.org/abs/1706.05587.

Codella, N., Cai, J., Abedini, M., Garnavi, R., Halpern, A., and Smith, J.R. (2015). Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images. Machine Learning in Medical Imaging, 118-226.

DeChant, C., Wiesner-Hanks, T., Chen, S., Stewart, E.L., Yosinski, J., Gore, M.A., Nelson, R.J. and Lipson, H. (2017). Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. Phytopathol. 107(11), 1426-1432.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In Proceedings of IEEE Conference On Computer Vision And Pattern Recognition, 248-255.

Doll, J., Grau, C., Jensen, B., Wedberg, J., and Meyer, J. (2016). Scouting corn: A guide to efficient pest scouting. University of Wisconsin Cooperative Extension Service, A3547.

Gale, W., Carneiro, G., Oakden-Rayner, L., Palmer, L.J., and Bradley, A.P. (2017). Detecting hip fractures with radiologist-level performance using deep neural networks. ArXiv. arxiv.org/abs/1711.06504v1.

Gensheng, H., Haoyu, W., Yan, Z., and Mingzhu, W (2019). A low shot learning method for tea leaf's disease identification. Comp. Elec. Agric. 163, 104852.

Ghosal, S., Blystone, D., Singh, A.K., Ganapathysubramanian, B., Singh, A. and Sarkar, S. (2018). An explainable deep machine vision framework for plant stress phenotyping. Proc. Nat. Acad. Sci. 115(18), 4613-4618.

Haenssle, H.A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., Kalloo, A., Hassen, A.B.H., Thomas, L., Enk, A. and Uhlmann, L. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Annals Oncol. 29(8),1836-1842.

Ha, J.G., Moon, H., Kwak, J.T., Hassan, S.I., Dang, M., Lee, O.N., and Park, H.Y. (2017). Deep convolutional neural network for classifying Fusarium wilt of radish from unmanned aerial vehicles. J. Appl. Remote Sens. 11(4), 042621.

He, K., Zhang, X., Ren, S. and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition, 770-778.

Heim, E., Roß, T., Seitel, A., März, K., Stieltjes, B., Eisenmann, M., Lebert, J., Metzger, J., Sommer, G., Sauter, A.W. and Schwartz, F.R. (2018). Large-scale medical image annotation with crowd-powered algorithms. J. Med. Imag. 5(3), 034002.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2017). An introduction to statistical learning with applications in R. Springer Texts in Statistics.

Kerkech, M., Hafiane, A., and Canals, R. (2018) Deep learning approach with colorimetric spaces and vegetation indices for vine diseases detection in UAV images. Comp. Elec. Agr. 155, 237-243.

Krähenbühl, P. and Koltun, V. (2011). Efficient inference in fully connected CRFs with gaussian edge potentials. In Advances In Neural Information Processing Systems, 109-117.

Luck, B., Drewry, J., and Nelson, J. (2018) Unmanned aerial vehicles: What you need to know for use in production agriculture.

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C.L. (2014) Microsoft COCO: Common objects in context. In European Conference On Computer Vision, 740-755.

Liu, L., Ouyang, W., Wang. X., Fieguth, P., Chen, J., Liu, X., and Pietikäinen, M. Deep learning for generic object detection: a survey. ArXiv. arxiv.org/abs/1809.02165.

Maji, S., Rahtu, E., Kannala, J., Blaschko, M. and Vedaldi, A. (2013). Fine-grained visual classification of aircraft. arXiv. arxiv.org/abs/1306.5151.

Miller, J.O. and Adkins, J. (2018). Types of drones for field crop production. University of Delaware Cooperative Extension, extension.udel.edu/factsheets/types-of-drones-for-field-crop-production/.

Mohanty, S.P., Hughes, D.P. and Salathé, M. (2016). Using deep learning for image-based plant disease detection. Front. Plant Sci. 7, 1419.

Park, J.H., Mirhosseini, S., Nadeem, S., Marino, J., Kaufman, A., Baker, K. and Barish, M. (2017). Crowdsourcing for identification of polyp-free segments in virtual colonoscopy videos. In Medical Imaging 2017: Imaging Informatics for Healthcare, Research, and Applications, 101380V.

Picon, A., Alvarez-Gila, A., Seitz, M., Ortiz-Barredo, A., Echazarra, J. and Johannes, A. (2019). Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. Comp. Elec. Agr. 161, 280-290.

Purdue Extension Annual Report (2018). UAVs for field diagnosis.

Ren, S., He, K., Girschick, R., and Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. ArXiv. arxiv.org/abs/1506.01497.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M. and Berg, A.C. (2015). Imagenet large scale visual recognition challenge. Int. J. Comp. Vis. 115(3), 211-252.

Sugiura, R., Tsuda, S., Tsuji, H., and Murakami, N. (2018). Virus-infected plant detection in potato seed production field by UAV imagery. 2018 ASABE Ann. Int. Meeting, 1800594.

Tetila, E.C., Machado, B.B., Belete, N.A.d.S., Guimarães, D.A., and Pistori, H. (2017). Identification of soybean foliar diseases using unmanned aerial vehicle images. IEEE Geosci. Remote Sens. Let. 14(12), 2190-2194.

Tetila, E.C., Machado, B.B., Menezes, G.K., Oliveira Jr., A.d.S., Alvarez, M., Amorim, W.P., Belete, N.A.d.S., da Silva, G.G., and Pistori, H. (2019). Automatic recognition of soybean leaf diseases using UAV images and deep convolutional neural networks. IEEE Geosci. Remote Sens. Let., early access.

Van Horn, G., Branson, S., Farrell, R., Haber, S., Barry, J., Ipeirotis, P., Perona, P. and Belongie, S. (2015). Building a bird recognition app and large scale dataset with citizen scientists: The fine print in fine-grained dataset collection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 595-604.

Wiesner-Hanks, T., Stewart, E.L., Kaczmar, N., DeChant, C., Wu, H., Nelson, R.J., Lipson, H. and Gore, M.A. (2018). Image set for deep learning: field images of maize annotated with disease symptoms. BMC Res. Notes 11(1), 440.

Wu, H., Wiesner-Hanks, T., Stewart, E.L., DeChant, C., Kaczmar, N., Gore, M.A., Nelson, R.J., and Lipson, H. (2019). Autonomous detection of plant disease symptoms directly from aerial imagery. Plant Phenome J. 2(1).

Zhou, N., Siegel, Z.D., Zarecor, S., Lee, N., Campbell, D.A., Andorf, C.A., Nettleton, D., Lawrence-Dill, C.J., Ganapathysubramanian, B., Kelly, J.W., Friedberg, I. (2018). Crowdsourcing image analysis for plant phenomics to generate ground truth data for machine learning. PLoS Comp. Biol. 14(7), e1006337.

CHAPTER 4

MULTIPLE DISEASE RESISTANCE IN TWO BIPARENTAL MAIZE POPULATIONS DERIVED FROM RECURRENT SELECTION

INTRODUCTION

There are many mechanisms by a which a plant might be bear resistance to multiple diseases (Wiesner-Hanks and Nelson 2016). As qualitative and quantitative resistance are better thought of as the far ends of a spectrum rather than distinct, unambiguous categories (Poland and Nelson 2011), many of these mechanisms apply similarly to the loci conditioning qualitative resistance (R genes) and those conditioning quantitative resistance (quantitative resistance loci; QRL). Multiple unlinked loci affecting single diseases might simply be pyramided in a single line. Such disease-specific loci might be closely linked to one another, a distinct possibility given the tendency of R genes to occur in clusters. A resistance locus may be truly pleiotropic, either to an even degree or a very uneven degree. If the effects of a resistance locus on two diseases are quite uneven (e.g. a 50% reduction in disease A but only a 5% reduction in disease B), the effect might only be statistically detectable for one disease, making the pleiotropic nature of the functionally invisible. Whatever the case, genetic mapping efforts can begin to characterize the nature and extent, if any, of multiple disease resistance (MDR).

This study characterizes maize resistance to three fungal foliar diseases: northern leaf blight (NLB), southern leaf blight (SLB), and gray leaf spot (GLS). The three diseases all produce grey-brown necrotic lesions but of quite different form and with quite different disease progression. NLB lesions are large and oblong, generally appearing at least a week after inoculation. SLB lesions are small and localized, typically appearing only a few days after inoculation. GLS lesions are intermediate

in size between those of NLB and SLB, with a characteristic rectangular shape caused by the fact that they do not cross the major veins of the maize leaf. In 2013, these diseases accounted for a combined estimated yield loss of 217.8 million bushels in the US Corn Belt and Ontario (Mueller *et al.* 2016).

Several collections of maize germplasm have been screened for resistance to NLB, SLB, and/or GLS. Resistance to these three diseases is highly correlated in the maize 282 diversity panel, but only a single potentially pleiotropic locus was identified by multivariate association mapping in this panel (Wisser *et al.* 2011). In the maize nested association mapping (NAM) panel, though resistance to the three diseases is highly correlated among the 25 founder lines, resistances are much less correlated within the biparental populations, suggesting that population structure is a stronger driver of these correlated resistances than pleiotropic or closely linked loci (Poland 2011). Accordingly, QTL or SNP associations for these three resistances do not overlap in the NAM population any more than would be expected by pure chance (Wiesner-Hanks and Nelson 2016).

This study examines the genetic architecture of resistance to NLB, SLB, and GLS in two biparental populations derived from a single line with high resistance to all three diseases and B73, which is moderately susceptible to all three. Ceballos *et al.* (1991) generated 411 full-sib families using parents from CIMMYT Pool 30, originally sourced from diverse subtropical lines from four continents. They then subjected this population to four cycles of reciprocal half-sib/S1 recurrent selection for NLB and rust resistance, as well as general vigor, decreasing mean area under disease progress curve (AUDPC) by 70%. This improved population (C4) was also more resistant to SLB and GLS, despite the fact that it was not selected for these resistances at all (Poland 2011).

MATERIALS AND METHODS

A single maize line from C4 of selection ($C4_{32-24}$) had been crossed to B73. Two F1 progeny were then selfed for four generations to create two population of $F_{4:5}$ RILs, P05 and P10, consisting of 197 and 235 lines respectively (**Figure 4.1**; Jesse Poland and Randall Wisser, personal communication). As the single C4 line was derived from a bulk selection scheme, it was expected to have high levels of heterozygosity, thus making the two F1 populations and resultant RIL populations distinct at some regions of the genome and isogenic at others.

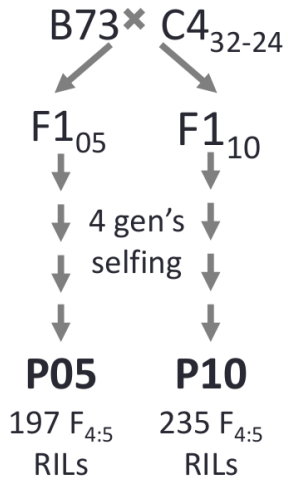


Figure 4.1: Creation of P05 and P10 RIL populations. Note that $C4_{32-24}$ was highly heterozygous, thus $F1_{05}$ and $F1_{10}$ were not identical.

The two populations had been evaluated for NLB resistance in Aurora, NY in 2009, 2010, and 2011 (Jesse Poland, personal communication), for SLB resistance in Clayton, NC in 2009 and 2010 (Peter Balint-Kurti, personal communication), and for GLS resistance in Blacksburg, VA in 2009, 2010, and 2011 (Jacqueline Benson, personal communication). NLB resistance was evaluated on a per-row

basis as percent diseased leaf area (DLA) three times at roughly ten days intervals. GLS resistance was scored on a 1-9 scale, with 1 = not diseased, 9 = fully diseased. SLB resistance was originally scored on a 1-9 scale, with 9 = not diseased, 1 = fully diseased, but this was later inverted to match the GLS scale. With the exception of the 2010 GLS experiment, days to anthesis (DTA) was scored on a per-row basis as the number of days after planting until anthers had emerged on >50% of the plants in a given row.

The standardized area under disease progress curve (sAUDPC) was calculated for each disease as:

$$sAUDPC = \frac{1}{t_n - t_1} \sum_{2}^{n} (t_j - t_{j-1}) \frac{x_j x_{j-1}}{2}$$

Where n is the number of evaluations, t_j is the days since planting at evaluation j, and x_j is the diseased leaf area (DLA) or 1-9 score at evaluation j. sAUDPC scores were square-root transformed to achieve normality as determined by Wilcoxon rank-sum test.

Lines were genotyped with genotyping-by-sequencing (GBS). SNPs were called using the TASSEL GBSv2 pipeline (Glaubitz *et al.* 2014). SNPs were filtered for minor allele frequency (MAF) >= 0.05 and presence in >= 10% of taxa with TASSEL and imputed with FSFHap (Swarts *et al.* 2014). SNPs called from GBS were merged with SNPs from the maize HapMapv3 (Bukowski *et al.* 2017), and the resultant intersecting SNPs were filtered for LD > 0.9 to the neighboring 5 SNPs. To call the parental origin of SNPs, haplotypes were pulled from GBS-called SNPs using FSFHap. A kinship matrix was calculated using the centered IBS method (Endelman and Jannink 2012).

Best linear unbiased predictors (BLUPs) for square-root transformed sAUDPC were calculated as:

$$\sqrt{(AUDPC)} = u + P_i + DTA_{jkl} + Y_j + R(Y)_{jk} + B(R)_{jkl} + e_{jklm} + Z_u$$

Where P_i and DTA_{jkl} are fixed effects of population and DTA, Y_j, $R(Y)_{jk}$, and $B(R)_{jkl}$ are random effects of year, rep, and block, u is the vector of RIL breeding values, and Z is an incidence matrix with covariance matrix K.

QTL were then determined by stepwise regression of imputed marker genotypes as predictors of BLUP value, using F-test of nested models in R to determine significance of marker addition or removal. The appropriate p-value threshold for addition or removal was determined by separate bootstrapping for each. For 1000 iterations for each trait in each population, BLUPs were randomized, F-tests were performed for each marker to determine significance of the addition of adding the marker as a predictor to the model with no markers, and the p-value of the most significant marker was recorded. The resultant 1000 p-values were then sorted and the 50th smallest p-value was taken as the appropriate p-value cutoff for alpha < 0.05.

After QTL were identified by stepwise regression, QTL confidence intervals were determined by finding the furthest flanking marker from the initial marker for which, when the initial and flanking markers were used as the only predictors in a GLM, the initial marker was no longer significant at the appropriate p-value threshold. Left and right boundaries to the confidence intervals were determined separately using this method.

RESULTS

As expected, given the fact that they were derived from a single open-pollinated line crossed to an inbred line, the P05 and P10 populations inherited identical allele contrasts at roughly half of the genome and unique allele contrasts at the other half (**Figure 4.2**). Most of the regions of identical

alleles in both populations spanned across most or all of a chromosome, namely chromosomes 6, 8, 9, 10.

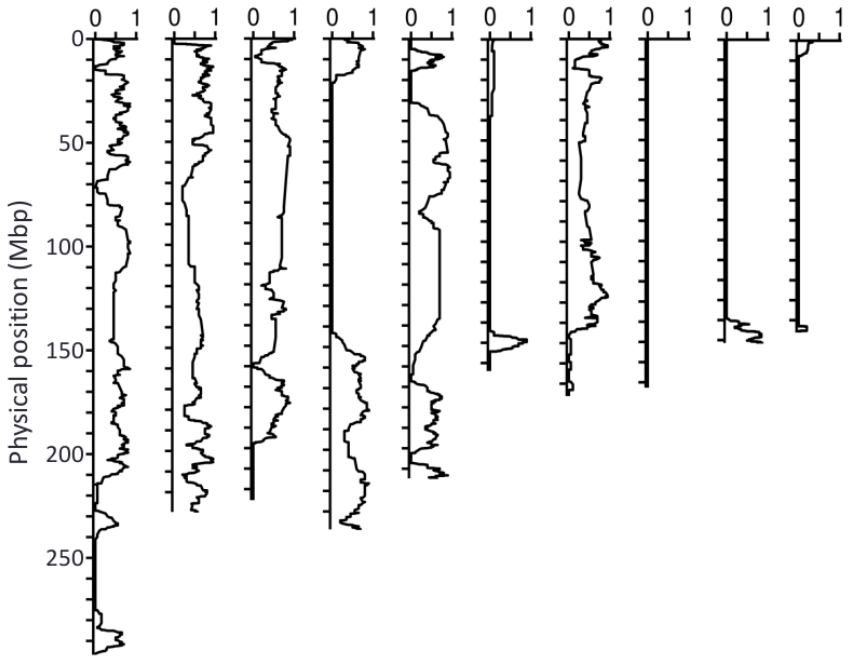


Figure 4.2: P05 and P10 have unique allele contrasts in roughly half the genome. Lines show proportion of SNPs for which P05 and P10 contained unique non-B73 alleles and thus presumably inherited different alleles from $C4_{32\text{-}24}$. Proportions averaged over 20-site sliding window. Stretches of 0 delineate regions in which P05 and P10 inherited identical haplotypes.

BLUPs for resistance to the three diseases were moderately correlated (**Figure 4.3**). The correlations observed (between r = 0.2 and 0.3) were similar to those observed between NLB, SLB, and GLS resistance within individual RIL families of the NAM population (Poland 2010). For each disease, between 5 and 8 QTL were identified in each population, explaining 50-75% of the additive

variance for resistance to the disease in question (**Table 4.1**). The confidence intervals of these QTL varied widely, with some pericentromeric QTL spanning over 50 Mb and more telomeric QTL generally under 5 Mb (**Figure 4.4**).

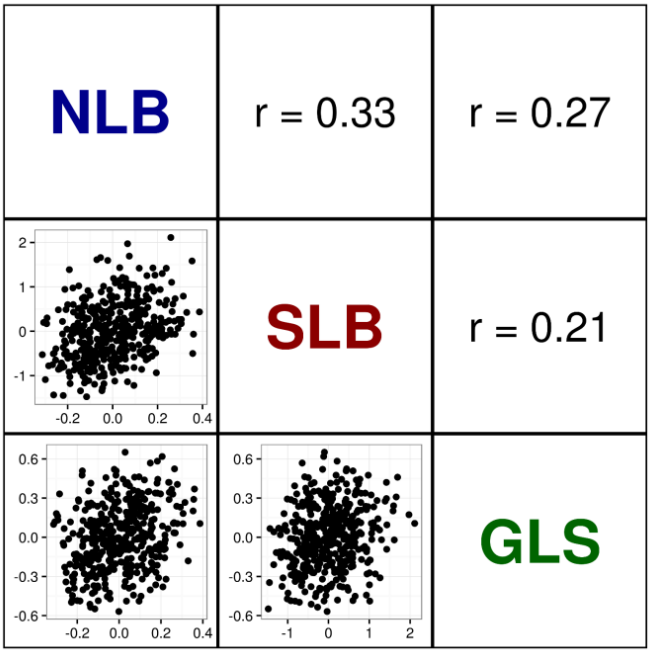


Figure 4.3: Correlations and biplots between BLUPs for NLB, SLB, and GLS resistance in both populations.

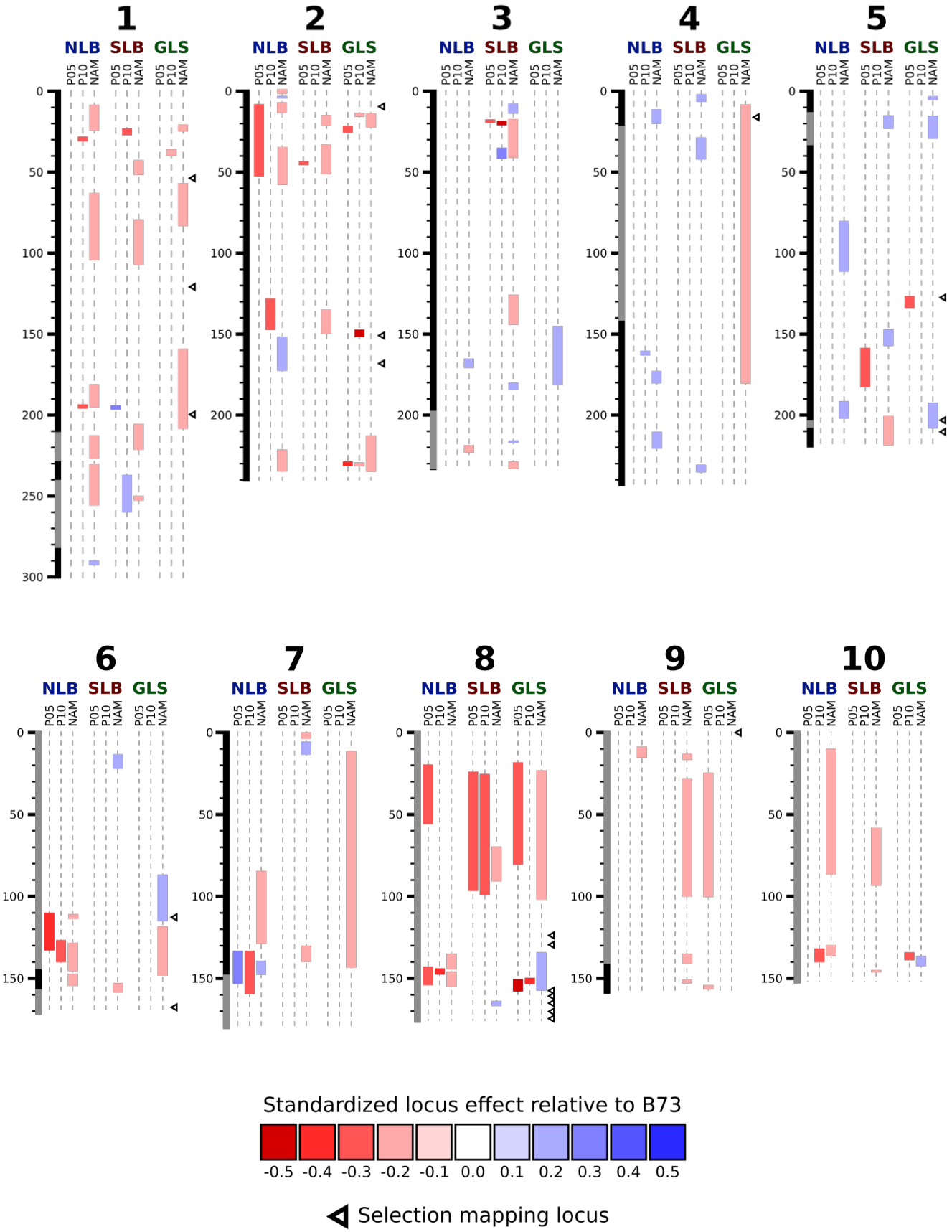

1
2
3
4
5
6
7
8
9
10
NLB
SLB
GLS
P05
P10
NAM
Standardized locus effect relative to B73
-0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5
Selection mapping locus

Figure 4.4: QTL confidence intervals for NLB, SLB, and GLS resistance. Locus effects are negative if the non-B73 allele reduces disease relative to B73 allele. Physical positions in Mb shown on left. Genomic regions for which P05 and P10 share haplotypes shown in grey on scale bar and genomic regions for which they were unique shown in black. For P05 and P10 QTL, effect estimates were standardized: β st = β ys marker. In the case of QTL from the NAM panel, QTL effects were not standardized and were coded light red if the average allele effect for families in which the marker effect was significant was negative to B73.

For most QTL, the P05- or P10-derived allele was the more resistant allele relative to that of B73. This was to be expected as a general trend, as B73 is a moderately susceptible line (Poland 2011), while most individuals of the RS C4 population were fairly resistant. One exception to this trend was very notable. Both P05 and P10 bore a QTL with fairly strong effect on chromosome 7, from roughly 135 Mb to 151-159 Mb. However, in P05, this QTL conferred susceptibility relative to B73, while in P10, it conferred resistance relative to B73. Thus, the original C4 line used to generate the populations apparently carried both alleles. Other such cases included three SLB QTL in which the P05-/P10-derived allele was more susceptible and one P10-derived QTL for NLB susceptibility relative to B73 on chromosome 4, 161-163 Mb. In all of these cases, it cannot be determined at the moment whether one parent is donating a susceptibility allele, one parent is donating a resistance allele, or both are true.

Trait	Population	# QTL	% Variance explained
NLB	P05	5	57.5
	P10	8	72.2
SLB	P05	5	61.3
	P10	5	72.1
GLS	P05	6	52.4
	P10	8	72.7

Table 4.1: Number of QTL identified for each disease in each population and total percent variance explained by QTL.

While some of the SSR loci previously identified as candidate resistance loci by selection mapping did colocalize with resistance QTL identified in this study (Wisser *et al.* 2008; **Figure 4.4**, triangles), a Fisher's exact test found no evidence to reject the null hypothesis of independence between whether an SSR was identified as positively selected (24 of 151 SSR) and whether it lay in a QTL or not for a given population and disease trait ($p >= 0.59$ in all cases). There were more assayed SSR loci inside NLB resistance QTL (19 in P05 QTL; 7 in P10 QTL) than inside SLB resistance QTL (0 in P05, 2 in P10) or GLS resistance QTL (0 in P05, 6 in P10), but further investigation showed that this was due to selection bias when SSRs markers were designed: a large number of SSRs were designed to lie near regions in which QTL had already been found in many populations, namely chromosome bins

8.06 and 6.05. Thus, on the whole, there was no statistical relationship between the loci identified by selection mapping and those identified by biparental QTL mapping.

For each disease, roughly two-thirds of QTL identified in P05 and P10 overlapped with QTL identified by joint linkage mapping in the NAM population (**Figure 4.4**). Despite the fact that the improved C4 population was also significantly more resistant to both GLS and NLB (Poland 2011), pleiotropic (or potentially pleiotropic) QTL were fairly rare. Of the 37 QTL identified for the three diseases in both populations, QTL for two diseases overlapped in only 4 instances (**Figure 4.4**).

DISCUSSION

Overall, the genetic architecture of disease resistance was similar for all diseases in both populations and generally comparable to that observed in the maize NAM population, the most relevant population with which to compare. For all disease/population combinations, between 5 and 8 QTL were detected, explaining between 50 and 75% of genetic variance. That the number of NLB resistance QTL and relative effects thereof were similar to those for SLB and GLS in this population was somewhat surprising. The original recurrent selection population from which the resistant parent was drawn was not selected for GLS or SLB resistance, but resistance to those diseases was higher in the final population nonetheless (Poland 2011). One would expect the four cycles of recurrent selection to break up the population structure present in the original, non-improved C0 population, leaving pleiotropic or closely linked loci as the main cause of MDR observed in the improved C4 population, or at least diminishing the effects of unlinked QTL for different resistances. Such pleiotropic loci or closely linked loci for different resistances would be apparent as overlapping QTL for different diseases in these populations. This did not appear to be the case, however. We found only a potential instances of

pleiotropic or closely linked loci, namely large pericentromeric QTL for NLB and GLS resistance on chromosome 8 and nearby, but not overlapping, QTL for NLB and GLS resistance in bin 8.06.

The lack of association between the location of the SSR loci identified by selection mapping and the QTL identified by linkage mapping in this population was surprising upon first glance. However, the nature of the populations used in these two mapping projects explains this fairly well. Selection mapping mostly identified loci that went from being relatively rare to being somewhat common during the multiple rounds of recurrent selection, rather than common loci that were pushed to fixation (Wisser et al. 2008). Thus, only a subset of the resistance alleles identified by selection mapping will be present in any one line from the improved C4 population, including presumably the $C4_{32\text{-}24}$ line used as a parent here. The relative precision of biparental mapping via GBS markers vs. that of selection mapping via SSRs also explains this somewhat. For instance, seven of the 24 SSRs identified as positively selected were located near bin 8.06, but outside of the bounds of the large-effect QTL identified in both populations. The positive selection of markers may very well have been caused by selection for this QTL.

Several of the QTL for NLB resistance lay in known “hotspots” for maize NLB resistance loci (Wisser et al. 2006), namely chromosome bins 8.06 (chromosome 8, ~145-165 Mb), 6.05 (chromosome 6, ~ 120-150 Mb), and 1.06 (chromosome 1, ~175-200 Mb). For each disease, roughly one-third of the QTL identified had no overlap with QTL identified in the maize NAM population. The number of putatively novel QTL identified here demonstrates that, despite the wide genetic basis of the maize NAM population, even a single population can provide numerous new promising resistance loci.

REFERENCES

Benson JM, Poland JA, Benson BM, Stromberg EL, Nelson RJ. 2015. Resistance to gray leaf spot of maize: genetic architecture and mechanisms elucidated through nested association mapping and near-isogenic line analysis. PLOS Genet. 11:e1005045

Bukowski R, Guo X, Lu Y, Zou C, He B, Rong Z, Wang B, Xu D, Yang B, Xie C, Fan L. 2017. Construction of the third-generation *Zea mays* haplotype map. GigaScience 7:gix134.

Ceballos H, Deutsch JA, Gutierrez H. 1991. Recurrent selection for resistance to *Exserohilum turcicum* in eight subtropical maize populations. Crop Science. 31:964-71.

Endelman JB, Jannink JL. 2012. Shrinkage estimation of the realized relationship matrix. G3 2:1405-13.

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES. 2014. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PloS One 9:e90346.

Mueller DS, Wise KA, Sisson AJ, Allen TW, Bergstrom GC, Bosley DB, *et al.* 2016. Corn yield loss estimates due to diseases in the United States and Ontario, Canada from 2012 to 2015. Pl. Health Prog. 17:211–22.

Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC, Nelson RJ. 2009. Shades of gray: the world of quantitative disease resistance. Trends Plant Sci. 14:21-9.

Poland J. 2010. The Genetic Architecture of Quantitative Resistance in Maize. PhD Thesis, Cornell Univ., Ithaca, NY

Swarts K, Li H, Romero Navarro JA, An D, Romay MC, Hearne S, Acharya C, Glaubitz JC, Mitchell S, Elshire RJ, Buckler ES. 2014. Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. Plant Genome 7.

Wiesner-Hanks T, Nelson R. 2016. Multiple disease resistance in plants. Annu. Rev. Phytopathol. 54:229-252.

Wisser RJ, Balint-Kurti PJ, Nelson RJ. 20016. The genetic architecture of disease resistance in maize: a synthesis of published studies. Phytopathol. 96:120-9.

Wisser RJ, Kolkman JM, Patzoldt ME, Holland JB, Yu J, Krakowsky M, Nelson RJ, Balint-Kurti PJ. 2011. Multivariate analysis of maize disease resistances suggests a pleiotropic genetic basis and implicates a GST gene. Proc. Nat. Acad. Sci. 108:7339-44.

Wisser RJ, Murray SC, Kolkman JM, Ceballos H, Nelson RJ. 2008. Selection mapping of loci for quantitative disease resistance in a diverse maize population. Genet. 180:583-599

APPENDIX A

THE ETYMOLOGY OF *ZEA*[5]

Introduction

As a personal project over the past few years, I have been researching the etymology of the name *Zea*. Though the history of the name is fascinating, I could not find a single summary in the literature of any era, and so decided to write one myself. This summary not being suitable for any journal I could think of, I submitted it to the Maize Genetics Cooperation Newsletter. This newsletter is a fairly informal avenue for distribution, as it does not appear in any academic database searches. I have included my summary as an appendix here in the hopes that such a fascinating story might find its way to more readers.

Background

In short, the name *Zea* referred to spelt or emmer for thousands of years. Linnaeus originally named maize *Thalysia* in his very first publication, but he used this name only once, deciding shortly thereafter to abandon it and requisition the name *Zea*. Though he gave clear reasons for rejecting the other extant names for the genus, his reasons for choosing *Zea* are unclear, since he knew the genus to be exclusively American and that it did not resemble spelt in any way. Given the justifications he made for similar choices, I believe that he chose *Zea* not due to any similarity between maize and spelt, but out of a desire to preserve a name with a millennia-long history and a prominent place in Greco-Roman sources.

5 Wiesner-Hanks, T. 2018. The etymology of *Zea*. Maize Genetics Cooperation Newsletter 92.

The Greek word ζεία (*zeia*) dates back to the oldest known works of Greek literature, appearing in Homer's *Iliad* and *Odyssey* and Hesiod's *Works and Days* as part of the phrase ζείδωρος ἄρουρα (*zeidoros aroura*, "corn-giving earth"). The etymology of this ancient phrase was thereafter the subject of occasional bickering for a few thousand years. Though *zeidoros* most directly translates to "*zeia*-giving" (euphemistically, "corn-giving" or "grain-giving"), it could also be derived from the Greek word ζάω (*zao*, to live) and thus mean "life-giving." As *zeidoros* appears before *zeia* in the written record, this would imply that the grain's name was similarly derived from its life-giving properties. Pliny the Elder asserted in *Naturalis Historia* (77-79 CE) that *zeidoros aroura* came from the "very considerable celebrity" of the hearty grain *zeia* and most certainly did *not* mean "life-giving," as was apparently widely believed by his contemporaries. This did not stop the the Greek grammarian Hesychius from reiterating the "life-giving" etymology in his *Alphabetical Collection of All Words* (c. 5th century CE). Various authors have sided with either authority since then: Leonhard Fuchs repeated Pliny's argument in *De historia stirpium commentarii insignes* (1542), while the French botanist Matthieu Bonafous supported Hesychius' etymology in his *Traité du maïs* (1833).

The three core texts of the Greco-Roman botanical canon- Dioscorides' *De Materia Medica*, Theophrastus' *Historia Plantarum*, and Pliny the Elder's *Naturalis Historia*- all described *zeia* as a rustic grain similar to wheat, traditionally translated as spelt or, more rarely, emmer. Theophrastus described it as the strongest among the grains that were not wheat or barley but one that "exhausts the ground," while Dioscorides noted single-seeded and double-seeded forms, both more nourishing than barley but less so than wheat. Pliny distinguished between *zea* (the Latin spelling) and *far*, which has also been used for both grains.

The short-lived *Thalysia*

Linnaeus originally named maize *Thalysia*, after a festival or harvest offering to the Greek goddess Demeter, enshrined in a pastoral poem of the same name by Theocritus. He used this name only once, in his very first publication, the first edition of *Systema Naturae* (1735). Over the next twenty years, Linnaeus would revise and expand *Systema Naturae* into an exhaustive 2300-page tome, the foundation of zoological taxonomy. This first edition, however, was a scant twelve pages, certainly an ambitious length in which to lay out his classification schema for every living thing (and all the world's minerals to boot). Linnaeus was bound to make a few missteps in his first publication, and it seems *Thalysia* was one of these. He did not use the name again in any other work, except to include it among other deprecated names.

The usage of *Zea* for maize has no precedent before Linnaeus' *Hortus Cliffortianus* and *Genera Plantarum*, both published in 1737. In 1735, Linnaeus took a post as curator at the estate of the wealthy Dutch banker George Clifford. Clifford's estate boasted an extensive herbarium and several greenhouses, replete with species from across the world. For the next two years, Linnaeus cataloged and categorized the plant genera therein and refined his system of generic and specific names, ultimately producing *Hortus Cliffortianus* and *Genera Plantarum*. The two texts work in concert, with *Hortus Cliffortianus* furnishing each genus with a list of pre-existing names, justification for the name Linnaeus settled on, and a cross-reference to its entry in *Genera Plantarum.* That entry in turn describes the morphology of the genus' flowers (the foundation of Linnaeus' classification schema) in great detail. In both works, and in every Linnaean work that followed, the genus of maize was listed as *Zea*.

Dated correspondence shows that Linnaeus decided at the last possible minute to drop *Thalysia* in favor of *Zea*. From 1735 to 1737, he corresponded frequently with Johan Frederik Gronovius, a Dutch botanist who was his friend and benefactor, to discuss the printing progress of *Genera Plantarum*. The work was printed in batches as Linnaeus completed the entries, and most of Gronovius' letters from this time revolve around updates on printing and additions or edits to be made. From the dates on these letters, we have a fairly detailed timeline of Linnaeus' writing process, one that shows that Linnaeus renamed maize *Zea* very shortly before printing.

In a 1736 letter, Gronovius asked Linnaeus to make some final edits so that another batch could be printed. He noted that Linnaeus had left blank space for *Thalysia* (entry 702), but still had not written the entry itself, and so requested that he send the text to be inserted. Though this letter was not dated, the entries discussed therein fall between those mentioned in two other letters, both dated. Thus Linnaeus decided to rename maize *Zea* between 15 June, when Gronovius had just received the manuscript for entries up to 507, and 26 September, when the entries up to number 717 had been printed. From the numbering, Linnaeus most likely made the decision to drop *Thalysia* and reassign *Zea* very shortly before printing. At the risk of overly dramatizing history, I find some comfort in the fact that history's most eminent botanist, when under a deadline, makes last-minute decisions.

Rejecting the alternatives

In the entry for maize in *Hortus Cliffortianus*, Linnaeus noted a bevy of extant names. To understand why he rejected these, we must look to his rules for nomenclature, which he laid out (in true Linnaean fashion) in painstaking detail. Each decision to accept, reform, or reject a name was ultimately founded on *Fundamenta Botanica* (1736), a collection of 365 aphorisms giving a philosophical framework for

why and how living things should be classified, as well as criteria for dividing bad names (lengthy, difficult to pronounce, based on color or size) from good ones (succinct, describing some consistent morphological feature, Greek or Latin root). When rejecting or accepting names in *Hortus Cliffortianus*, he often cited the relevant aphorism(s) from *Fundamenta Botanica*, occasionally clarifying with brief comments. For example, when he retracted *Thalysia*, his own name for maize, he made no explanation apart from citing aphorism 244: “New generic names should not be contrived, so long as adequate synonyms are readily available.”

Most of the extant names noted in *Hortus Cliffortianus* fell into two groups: derivatives of the Taíno word *mahiz* and demonyms such as *Frumentum indicum* (Indian corn) or *Triticum turcicum* (Turkish wheat). Two of Linnaeus’ most notable predecessors, Cesalpino and Tournefort, used derivatives of *mahiz* (Mays and Maiz) for the genus as a whole. Linnaeus, however, rejected these under aphorism 229 of *Fundamenta Botanica*: “Generic names that do not have a root in the Greek or Latin languages are to be rejected.” No such rule existed for specific names, towards which Linnaeus took a much more lax view, and so he relegated *mays* to a mere species name.

Many of the demonyms used for maize correctly noted the American origin of the crop, though using the adjective “Indian,” e.g. *Milium indicum* (Indian millet) or *Triticum indicum* (Indian wheat). Others, such as *Frumentum turcicum* (Turkish corn), purported an Asian origin, as discussed below. Linnaeus cited no specific reasons for rejecting these, though he used *Triticum* and *Milium* for other genera in *Hortus Cliffortianus* and disliked demonyms as a rule (see *Philosophia Botanica* 235).

The fascinating misnomer “Turkish corn” warrants a digression. Several other American species bore such misnomers in European sources in the 1500s, e.g. *Cucurbita pepo* (called “Turkish cucumber” or “Turkish melon” in several prominent herbals of the time) and the eponymous meaty

bird. A putative Turkish origin for maize was repeated by numerous sources. The first European herbals to describe maize, Tragus' *Kreüter Buch* (1546) and Rembert Dodoens' *Cruyde Boeck* (1554), both referred to it as a Turkish crop. The extreme similarity of their illustrations and several phrases in the text suggests that Dodoens borrowed heavily from Tragus' entry, likely repeating this false origin in the process. Caspar Bauhin's *Pinax Theatri Botanici* (1623) documented over twenty names for maize, many purporting an origin in Turkey or elsewhere in western Asia, such as *Frumentum asiaticum* (Asiatic corn) or *Triticum bactrianum* (Bactrian wheat).

Judging from successive editions and translations of Dodoens' highly influential *Cruyde Boeck*, the myth of a Turkish origin of maize was dispelled some time in the late 1500s. The original Dutch editions (1554, 1563) called maize *Frumentum turcicum* or *Bled sarrazin* ("Saracen wheat," Saracen being a generic term for Arab, Middle-Eastern, or Muslim) but did not mention its provenance. A later English translation by Henry Lyte (1578) added the name "Indian wheate," but maintained that it "groweth in Turkie." Not long after this, the first Latin translation (1583) correctly recognized that maize was "by no means from Asia" (*haudquaquam ex Asia*), but rather from the Americas, ostensibly Hispaniola. This misnomer has nevertheless survived as the modern *grano turco*, a colloquial Italian name for maize.

Linnaeus' justification for *Zea*

So, having ruled these other names out, why did Linnaeus choose *Zea* for an American genus that looks very little like spelt or emmer, the historical bearers of the name? His explanation in *Hortus Cliffortianus* is terse:

> *Zea, peculiaris frumenti species a veteribus adscriptum nomen, huc usque vagum, recepimus ad designandum hoc genus loco Barbari istius vocabuli Mays. F.B. 229. Thalysia F.B. 244. Zea F.B. 242*
>
> We have accepted this genus to be designated *Zea*, the name given by the ancients given to a specific species of grain, hitherto vague, in place of that barbarian name Mays. F.B. 229. Thalysia F.B. 244. Zea F.B. 242

The relevant aphorism, *Fundamenta Botanica* 242, states:

> *Nomen genericum antiquum (241) antiquo generi convenit.*
>
> An ancient (241) generic name is appropriate for an ancient genus.

What exactly did Linnaeus mean by "ancient"? The cited aphorism 241 discusses names given by the Greek and Roman "fathers" of botany (*nomina generica Patrum Botanices graeca vel Latina*), whom Linnaeus revered. It is reasonable to think that "ancient genus" would have a similar connotation, i.e. ancient Greek or Roman, not simply old. However, with *Zea*, Linnaeus gave an ancient Greek/Roman name to a genus he knew to be American and thus unknown to ancient Greek and Roman sources.

To resolve this apparent contradiction, we can look to Linnaeus' *Critica Botanica* (1737), which clarifies and expands on many of his aphorisms. Though he did not mention *Zea* in his discussion of aphorism 242, he discussed *Cactus* at great length, a genus with many relevant parallels to *Zea*. Both genera were of American origin, but widely known in Europe in the 1700s, and both names were originally Latin names for plants that were placed into genera named after their more famous relatives (*Zea* into *Triticum spelta*, *Cactus* into *Cynara cardunculus*), leaving the names unused. Several eminent botanists had named the genus of cacti *Opuntia*, a name Linnaeus rejected. This gave him the option to

devise some new name, one which conveyed the very distinctive physical characteristics of cacti. However, the genus was widely known in Linnaeus' day as *Melocactus* (loosely, "melon thistle"), and in order to not upset the "common people" (*ne vulgo displicerem*), he decided to simply repurpose the ancient name *Cactus*, used by the ancient Romans for the similar-looking cardoon. The contradiction is thus resolved by a double standard: while an "ancient name" is strictly one from the ancient Greeks or Romans, a genus widely known for only a few hundred years can qualify as an "ancient genus."

Though this explains why Linnaeus had no problem repurposing an ancient Greek/Latin name for an American genus, the question remains as to why he chose *Zea* in particular, given that maize and spelt look quite different. Given the frequently poetic dimension to his work (partitioning his system into exactly 365 rules was certainly more of an aesthetic choice than a scientific one), I believe that his choice was not founded in morphology, but rests more on his esteem for the name itself. His main justifications for reassigning *Cactus*, which had been left unused after Linnaeus placed the cardoon with artichokes in the genus *Cynara*, were that it was both very ancient (*antiquissimum*) and very widely-known (*vulgatissimum*). This outweighed the fact, noted by Linnaeus himself, that designing a new name that was unambiguous and founded on invariant morphological features (the ideal construction according to *Philosophia Botanica*) would be very easy. In the case of *Cactus*, Linnaeus felt that the cultural and historical considerations of leaving an ancient Latin name unused outweighed his own scientific criteria. Given the ancient history of the name *Zea*, I believe this was also his sentiment when renaming maize *Zea*.

Genaust's Hypothesis

To my knowledge, only a single author has put forward a hypothesis as to why Linnaeus chose *Zea* for maize. The German lexicographer Helmut Genaust addressed the question in his *Etymologisches Wörterbuch der Botanischen Pflanzennamen* (*Etymological Dictionary of Botanical Plant Names*, 1976):

> The long unanswered question, why Linnaeus now transferred the pre-Linnaean designation of spelt as a generic name to maize, finds its answer in the observation that the variety that is likely the oldest, *Zea mays* convar. t*unicata* (pod corn), has closed grains, as compared to dent corn (convar. *dentiformis*), today's most cultivated and highest-yielding variety, in much the same way that spelt can be contrasted to the higher-yielding, free-threshing common wheat; moreover, the male flowers of maize are only two-flowered.

I find this explanation dubious. Though we now know that *Tu1* was a later mutation, the question is whether Linnaeus himself believed pod corn to be maize's ancestral form. This was almost certainly not the case. First, it was unlikely that Linnaeus had ever seen pod corn. There are no reliable European accounts of pod corn predating 1809, when Spanish officer Felix de Azara described a variety called *abatý-guaicurú* in his *Voyages dans l'Amérique méridionale*. Not long after this, the French botanist Auguste Saint-Hilaire announced his discovery of pod corn in an 1829 letter to the French Academy of Sciences. Both authors felt they had encountered something unprecedented: de Azara described *abatý-guaicurú* as "singular," while Saint-Hilaire announced his as "a remarkable variety." Such breathless descriptions would be unlikely if pod corn was known to Europe's most eminent botanist a century prior. Second, in *Genera Plantarum*, Linnaeus describes the calyces and corollae of female maize flowers as being especially short (*brevissimus*), clearly not the long, enveloping structures found on

pod corn. The type specimen at Clifford's estate, of which only the tassel was preserved, was thus certainly not pod corn.

Final thoughts

A newly discovered genus had heart-shaped marks on its seeds, and so Linnaeus named it *Cardiospermum*, or "heart-seed." This etymology is clear, unambiguous, easily summarized, and boring. I find the muddled, winding history of the name *Zea* to be much more fascinating. Greeks and Romans debated for centuries whether "*zeia*-giving earth," found in the earliest works of Western literature, was a metaphor for the life-giving properties of Mother Earth or simply a nod to a useful cereal. Europeans compiled a laundry of list of mistaken homelands for maize, from the Arabian peninsula to the steppes of central Asia. History's most eminent botanist, under the gun from his publisher, made a last-minute decision to rename the plant which would become the world's most widely cultivated crop. Is the very messy, and thus very human, story of the name *Zea* not far more interesting than what can be said about so many other genera?

As Helmut Genaust said, the question of why Linnaeus transferred the name is a "long unanswered question." The simplest (and most cynical) answer is that if you are Carolus Linnaeus, you can name a genus whatever you darn well please. This explanation, however, is not a very satisfying one. Linnaeus spent his entire life creating and refining a nomenclature system, one that fused an obsession with clarity, poetic notions of a sublimely ordered universe, and a sense of inheritance from venerated ancients. His decisions may have had dubious justifications, such as his distaste for "barbarian" names, but they had justifications nonetheless. I believe his choice of *Zea* was not founded

in any meaningful parallels between maize and spelt, but rather rested on his notion that names with an illustrious history (to Linnaeus, synonymous with a lengthy history in Greek or Latin writing) must not be lost.

APPENDIX B: TRANSCRIPTION AND TRANSLATION OF PASSERINI'S 1876 DESCRIPTION OF *HELMINTHOSPORIUM TURCICUM*

The species *Exserohilum turcicum*, originally called *Helminthosporium turcicum*, was first described by Italian pathologist Giovanni Passerini in the October 1876 issue of the *Bolletino del Comizio Agrario Parmense* (*Bulletin of the Agrarian Committee of Parma*). Though this report is cited frequently (indeed, any time the species is described in detail), its contents are not available to the Anglophone scientific world. A scan or digital copy is not available in any database of scientific or historical publications, and hard copies of the *Bolletino*, a regional agricultural journal with limited circulation, can be found in only a few Italian libraries. I obtained a scan of the report from the Biblioteca Civica dell'Ospedale Vecchio in Parma, transcribed it, and made my best attempt to translate it. Because the report is now in the public domain under Italian copyright law, and because I could think of no other suitable avenue to distribute it, I have included it here as an appendix. Passerini's observations and impressions are fascinating, and I hope that they are of scientific and historical interest to other researchers of NLB, *S. turcica*, or the history of plant pathology and fungal taxonomy in general.

LA NEBBIA[6] *DEL GRANO TURCO*

Discorrendo nel no. 7 (Luglio 1876) di questo Bolletino sulla Nebbia dei Cereali accennai come il Grano turco sia soggetto alla ruggine (*Puccinia Sorghi* Schweinz) la quale ne deturpa e fa seccare anzi tempo le foglie.

In quest' anno, memorabile per le nebbie, ossia per l'invasione de' fungilli parassiti su quasi ogni sorta di piante tanto coltivate che selvatiche, questa Puccinia del Grano turco, benché non abbia mancato di mostrarsi qua e là, non ha preso però l' estensione né spiegato l' intensità che ebbi altre volte ad osservare. Con tutto ciò i lamenti sulla Nebbia del Grano turco si fecero sentire assai per tempo, non iscompagnati da timori sulla prosperità della raccolta, quali pur troppo si sono più o meno verificati.

Era un fatto del quale molti non sapevano rendersi ragione, che cioè, persistendo pur sempre la stagione piovosa, già fino dal mese di Luglio, al tempo della fioritura della pianta, e quindi in quello del maggior vigore di essa, le foglie inferiori e gradatamente le successive andassero d'improvviso seccando in tutto od in parte, non altrimenti da ciò che sarebbe avvenuto per opera di estrema siccità. E ciò accadeva sopratutto nella bassa pianura e nelle località umide ove le piante intere persino ne perirono.

Al primo sentore che n'ebbi, stimai che si trattasse della ruggine, ma vedute appena le piante affette, conobbi che doveva essere altra cosa, non trovando di ruggine che scarsissime traccie, od anche nessuna.

6 *Nebbia*, typically meaning "fog" or "mist," is an archaic Italian term used for many blights, leaf spots, etc. See Merino-Rodriguez 2013, *Lexicon Of Plant Pests and Diseases*

Cominciava il male con delle macchie giallicce che presto inaridivano, estendendosi a tutta la lunghezza della foglia od a gran parte di essa. Alla superficie di tali macchie, pressochè nulla di rimarchevole riesciva visibile all' occhio nudo, e soltanto coll' aiuto della lente potevasi scorgere una fina peluria di colore olivastro, la quale non di rado mancava per qualche tratto od anche per tutta quanta l' estensione della macchia. L' esame microscropico di tale peluria rendeva poi manifesta la presenza di un fungillo microscopico, formato di lunghi fili articolati diritti o flessuosi di color giallo-bruno e producenti grandi spore fusiformi divise da quattro o più tramezzi, giallo-olivastre e trasparenti.

Simigliante struttura è evidentemente quella di un fungillo demaziaceo del genere *Helminthosporium*, nel quale però non trovandosi, per quanto io ne sappia, alcuna specie nota che si attagli a questa del Grano turco, debbo di nenecessità [sic] presentarla sotto il nuovo nome di *H. turcicum*. (*)

(*) Gli esemplari che ne conservo nell' erbario portano la seguente scheda:

Helminthosporium turcicum. Passer. hb. Hyphae longae rectae vel flexuosae articulatae luteo-fuscae, sporae magnae fnsiformes [sic, fusiformes] multiseptatae, luteo-olivaceae.

Ad folia languida, mox arescenda Zeae Maydis in campis Provinciae Parmensis. Aestate 1876.

La peluria suaccennata costituisce !a [sic] parte fruttificante del fungo, la cui porzione vegetativa nascosta nel tessuto delle foglie è quella che produce in esse le macchie, cagionandone il successivo disseccamento. Quanto alle macchie sulle quali non iscorgesi la lanugine, ossia la fruttificazione del fungillo, si può ritenere che questo è rimasto sterile, forse per la troppo rapida morte

del tessuto, dal quale soltanto può trarre il necessario alimento. È il caso di chi non avendo che una data provvisione per vivere si trovi ridotto a morire di fame per averla esaurita anzi tempo.

Le macchie sterili però potrebbero anche appartenere ad un altro fungillo che ne produce delle somiglianti sulle foglie del Grano turco. È questo l' *Epicoccum neglectum* Desmaz. che talvolta ho incontrato da solo e tal'altra frammisto coll' *Helminthosporium* di cui emula gli effetti.

L'azione pertanto dei due accennati fungilli spiega abbastanza il fatto in apparenza paradossale del precoce inaridimento delle foglie ed anche di tutta la pianta, non ostante l' umidità della stagione; nella quale anzi trovano i fungilli medesimi la condizione più favorevole per prosperare e per estendere il loro influsso nocivo.

Tale essendo quindi la cagione e l' essenza della Nebbia del Grano turco, vano sarebbe il pensare a rimedi finché la causa perduri, mentre col cessare di questa anche il danno può arrestarsi od almeno scemare da sè. Questo poi, come facilmente s' intende, riesce tanto maggiore quanto più giovani sono le piante che ne vengono colte.

Che l' umidità eccessiva sia stata veramente la causa di questa nebbia lo dimostra anche il modo con cui si andò estendendo nella nostra Provincia. Nel mese di Luglio limitavasi dessa alla bassa pianura ed ai terreni di non facile scolo; soltanto verso la fine d' Agosto, persistendo sempre le piogge, aveva guadagnato l' alta pianura e le colline; e finalmente nella seconda metà di Settembre era salita sulle montagne, ov' ebbi ad osservarla, sebbene non molto intensa, presso Borgotaro e Bedonia in occasione della gita degli Alpinisti al Monte Penna.

Vigheffio 18 Ottobre 1876

Prof. G. PASSERINI

BLIGHT[7] OF MAIZE

While discussing cereal blight in number 7 (July 1876) of this bulletin, I mentioned that maize can be subject to rust (*Puccinia sorghi* Schweinz), which scars the leaves and dries them prematurely.

In this year, notable for blights- that is, for the invasion of parasitic fungi on almost every kind of cultivated and wild plant- this *Puccinia* of maize, although it has not failed to appear here and there, has not reached the range nor the intensity that I had observed on previous occasions. Given all this, reports[8] of maize blight arose in due time, not unaccompanied by fears over the success of the harvest, which have so far been more or less verified.

It was a fact of which many were not aware, that is to say, that persisting throughout the rainy season, until the month of July, at the time of the plant's flowering, and hence in the time of its greatest vigor, the lower leaves, and gradually those above it, suddenly began withering in whole or in part, not unlike what would occur under extreme drought. And this was happening particularly in the lowlands and humid areas, where entire plants died.

Upon first impression, I thought that this might pertain to rust, but as soon as I saw the affected plants, I realized that it had to have been something else, not finding even the scarcest trace of rust.

The disease began[9] with pale yellow spots that soon dessicated, extending down the entire the leaf or a large part of it. At the surface of these lesions, nothing remarkable was visible to the naked eye, and only with the help of the lens was it possible to see a fine olive-colored down, which was rarely lacking for a few parts or all the extent of the extent of the lesions. Microscopic examination of this down then revealed the presence of a microscopic fungus composed of long, articulated hyphae,

7 All instances of *nebbia* have been translated as "blight," per footnote above.
8 The original describes "laments" (*lamenti*), but a literal translation would not fit the tone of a scientific report.
9 Verbs in this paragraph were originally in mostly imperfect tense, but this translates awkardly to English. I have changed several verbs to simple past tense, but it must be remembered that Passerini is describing something he observed several times, not just once.

straight or flexuous and of a yellow-brown color, producing large fusiform spores divided by four or more septa, yellow-olive in color and transparent.

Such a form is evidently that of a dematiaceous[10] fungus of the genus *Helminthosporium*, which to my knowledge does not yet contain any known species corresponding to this pathogen of maize, and thus I must necessarily present it under the new name of *H. turcicum*. (*)

(*) The samples I keep in the herbarium bear the following index:
Helminthosporium turcicum. Passer. Hyphae long, erect or flexuous, articulated, yellow-brown. Spores large, fusiform, septate, yellow-olive. On diseased leaves, soon afterwards withered, of *Zea maydis* in the field in the Province of Parma. Summer 1876.

The aforementioned down constitutes the reproductive part of the fungus, while the vegetative portion concealed in the leaf tissue is the portion which produces lesions, leading to the subsequent dessication. As for the lesions on which the down- that is, the fruiting body of the fungus- is not found, one might think that the fungus has remained sterile, perhaps because of the overly rapid death of the tissue, the exclusive source from which it can extract its necessary nourishment.

The sterile lesions might instead belong to a different fungus that produces similar symptoms on maize leaves. This would be *Epicoccum neglectum* Desmaz., which I have encountered at times by itself and at other times intermixed with *Helminthosporium*, of which it emulates the appearance.

The action of these two fungi thus adequately explains the seemingly paradoxical fact of the early dessication of the leaves, and even of the whole plant, in spite of the humidity of the season; in

10 Of defunct family Dematiaceae.

such conditions, these fungi actually find more favorable conditions in which to prosper and extend their noxious influence.

This being the cause and the essence of Maize Blight, it would be pointless to think of remedies as long as the cause persists, whereas if this causal agent is stopped, the damage may stop or at least diminish on its own. This then, as is plain to see, is more successful the younger the plants being cultivated are.

That excessive humidity was truly the cause of this blight also is also demonstrated by the way in which it spread throughout our province. In the month of July it was limited to the lowland plains and to soils with poor drainage; only towards the end of August, with constantly persistent rain, did it reach the high plains and hills; and finally in the second half of September I climbed the mountains, where I observed it, though not very intense, at Borgotaro and Bedonia on the Mountaineers' trip to Monte Penna.

Vigheffio, 18 October 1876.