GENETICS OF MULTIPLE DISEASE RESISTANCE IN MAIZE INBRED NY22613 AND SCIENCE COMMUNICATION OF QUANTITATIVE GENETICS

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

Dhyaneswaran Palanichamy

August 2019

GENETICS OF MULTIPLE DISEASE RESISTANCE IN MAIZE INBRED NY22613 AND SCIENCE COMMUNICATION OF QUANTITATIVE GENETICS

Dhyaneswaran Palanichamy, Ph.D.

Cornell University 2019

Abstract

Given unpredictable pathogen pressures caused by changing climatic patterns, plant breeders aim to breed crop varieties with durable resistance to multiple plant pathogens. Understanding the genetic basis of multiple disease resistance will aid in this endeavor. Maize inbred NY22613, developed at Cornell University, have shown resistance to northern leaf blight (NLB), gray leaf spot (GLS), common rust, and Stewart's wilt (SW). A BC_3S_3 bi-parental mapping population (resistant inbred NY22613 and susceptible inbred Oh7B) was used to map the QTLs responsible for disease resistance. The analysis revealed that 16 quantitative trait loci (QTL) were associated with NLB resistance, 17 QTL with GLS resistance and 16 QTL with SW resistance. No QTL were colocalized for all three diseases. Three QTL were shared for NLB and GLS and one QTL was shared for GLS and SW. To select individuals with multiple disease resistance, we demonstrated a selection method that uses phenotypic data, QTL data and high density marker information in a cluster analysis, designated the high density marker phenotype (HEMP) QTL selection strategy.

A differential expression study was conducted using susceptible inbred Oh7B and resistant inbred NY22613 in both field and greenhouse conditions, to identify genes that are differentially expressed when inoculated with *Setosphaeria turcica* (NLB). The *Zm00001d024772* gene (unknown function in maize) was differentially expressed between the uninoculated and inoculated

Oh7B in field and greenhouse conditions. *Zm00001d027691, Zm00001d011152, Zm00001d008951, Zm00001d033623, Zm00001d021770* and *Zm00001d034421* were differentially expressed in response to NLB inoculation in NY22613 in field and greenhouse conditions. None have a previously known function in maize, but *Zm00001d033623* plays a major role in rice disease immunity. QTL analyses implicates *liguleless1* to be associated with disease resistance to GLS and SW and the differential expression study implicates *liguleless1* gene to be associated with disease resistance for NLB. This suggests that *liguleless1* is an important candidate gene for multiple disease resistance.

Direct-to-consumer genetic testing companies conduct low cost genotyping and genome sequencing for humans. This has led to the public having access to their genomic data more than ever before. Quantitative genetics is essential to understand genomic data. Science communication of quantitative genetics to the public is an under-explored strategy to address this issue. The story of quantitative genetics in humans is ugly due to its eugenic origins, however, the story of quantitative genetics in agriculture is inspiring. Using the achievements of quantitative genetics in agriculture, key concepts can be communicated to a diverse audience. Further, the quantitative genetics methods used in plant and animal breeding are being used in human genomic data. This necessitates plant and animal breeders/geneticists to participate in the communication of quantitative genetic methods to the public, so that the public can make informed decisions with their genomic data.

BIOGRAPHICAL SKETCH

Dhyaneswaran Palanichamy was born in a tea plantation in Southern India. His parents were health workers and were instrumental in Dhyan getting a good education. Dhyan spent most of his childhood in catholic boarding schools, since the schools in tea plantations were usually very under resourced. Even at a very young age, he was fascinated by life and how it works. He learned about DNA during his seventh grade in a public library, and it has caught his imagination since then. Dhyan did his bachelors in biotechnology in Tamil Nadu Agricultural University. There he learned how molecular markers can be used to breed drought tolerant rice varieties. Realizing the potential of plant breeding he decided to pursue his graduate studies.

He learned disease resistance breeding in wheat from Dr. Mark Sorrells during his masters of professional studies in Cornell University. Later, he pursued his masters in biotechnology in TNAU and learned to use molecular markers in maize breeding to select for better nutrition. However, at Cornell, Dhyan understood that there is so much more to plant breeding than molecular markers. Determined to learn more, he asked around about pursuing a PhD. With the help of several mentors and teachers Dhyan applied and got accepted for a PhD position in plant breeding and genetics at Cornell University. He worked with Dr. Margaret Smith in the Cornell corn breeding program on multiple disease resistance in maize.

At Cornell, Dhyan learned about plant breeding and how it could be used to serve humanity. He also learned about the distance between the scientific community and the public through his visits to Museum of Natural History in NYC and the Creation Museum in Kentucky. Dhyan learned that

humanity needs to know about science and its consequences before they use it to better themselves, especially with food. He also realized the enormous potential of genomic technologies and the need for better regulation. It is at this juncture, Dhyan developed an interest in science communication. Through his mentors at Cornell University, Dhyan learned how to communicate science to a wide audience. With his unique skill set in plant breeding and communication, Dhyan hopes to be a part of the next generation of scientists who strive to help feed humanity through science.

DEDICATION

I dedicate this thesis to my wife, mother, father and sister. Their sacrifices for me made this work possible.

ACKNOWLEDGEMENTS

First, I would like to thank God for helping me to accomplish this goal. I thank my teachers and mentors, especially, Dr. Chandra Babu and Dr. Senthil Natesan in Tamil Nadu Agricultural University, India who believed in me and inspired me to be a scientist. I appreciate and acknowledge the scholarships and assistantships from Tata Cornell Fellowship, Cornell-Sathguru foundation, The Hatch Research Project, National Institute of Food and Agriculture, United States Department of Agriculture, Panzea project, the Graduate School and the Section of Plant Breeding and Genetics, Cornell University, I thank the Department of Neurobiology and Behavior and Investigative Biology Labs (BioG1500) for allowing me to teach undergraduates and improve my pedagogical and teaching skills, and Guterman Bioclimatic Labs for employing me on weekends when I desperately needed funding and teaching me how to care for research plants.

I thank Dr. Gebisa Ejeta, Dr. Peter Gregory, and Dr. K.V. Raman for their encouragement during my masters to pursue my Ph.D. I am grateful to Dr. Mark E. Sorrells for teaching and inspiring me with the potential of plant breeding and genetics to address problems in world food production and hunger during my masters. I thank Dr. Edward Buckler and Dr. Mike Gore, for inspiring me with their passion and teaching me quantitative genetics. I am grateful for Dr. Carlyn Buckler for her guidance and advice in science communication. I thank Dr. Rebecca Nelson and Dr. Judith Kolkman for their guidance and help with my experimental design and pathogen samples. I thank Dr. James Giovannoni and Dr. Yimin Xu for helping with my RNA Seq experiment. I thank Dr. Qi Sun and Dr. Deniz Akdemir for helping me with high performance computation and statistical analysis. I thank Dr. Mark A. Sarvary for helping me with the opportunity to teach and inspiring

me with his passion in pedagogy to conduct pedagogical research. I thank Dr. Jeff J. Doyle for his words for encouragement and guidance. I thank Mr. Dan Fisher, Mr. Keith Payne and Mrs. Sherrie Norman for their support and help in my field projects.

I thank the elder statesmen pickup basketball group, a 50 year old Cornell tradition, for teaching me the game and inspiring me to be a leader. I thank the Synapsis graduate student organization of the Section of Plant Breeding and Genetics, a 100 year old Cornell tradition for the book clubs, BBQs and friendships. I thank my friends in Cornell International Christian Fellowship for their support and kindness throughout my PhD. I thank my parents and sister for their love and support. I thank my wife Mrs. Lauryn Palanichamy for her love and support throughout my PhD.

I would like to thank Dr. Bruce Lewenstein for inspiring me and advising me in the field of science communication. I am grateful for his guidance in my professional development and publishing in the field of science communication. I thank Dr. Walter De Jong for teaching and advising me in molecular biology and his guidance in applying to the PhD program in Cornell University. I also thank him for trusting me enough to ride in my motorcycle in India. Finally, I am grateful to Dr. Margaret Smith for believing in me and teaching me maize breeding. I am thankful for her kindness, guidance and unwavering support during the most difficult times of this journey.

LIST OF TABLES

LIST OF FIGURES

PREFACE

Maize (*Zea mays* L.) is one of the major cereal crops and its production is affected by a number of plant pathogens. In the 21st century, the rise of genomics has enabled plant breeders to use genomic tools to understand the genetic basis of disease resistance and breed for this trait. These genomic tools have also been used in animal breeding. The genomic tools that are developed in animal breeding can be easily translated to plant breeding and *vice versa*. These developments in genomics make one wonder if these tools could also be used in human genetics. The thought of using genomic tools developed for efficient discrimination in plant and animal breeding in human genetics raises questions about ethics, regulation and public awareness.

In this dissertation, I describe use of a biparental mapping population to map the disease resistance loci present in maize inbred NY22613, which was developed using traditional breeding methods. I also discuss the modern genomic methods to understand the genetic basis of disease resistance in NY22613. Following the genomics of the disease resistance in maize, I delve into the science communication aspects of quantitative genetics for humans. This is important because the public is largely unaware of the developments in the field of genomics and understanding quantitative genetics will aid in correct interpretation of direct-to-consumer genetic tests.

The first chapter describes the rationale for this study and provides background on the genetic basis of multiple disease resistance. It discusses the complex nature of the plant immune system and then delves into the relationship between the maize genome and plant pathogens. Results are presented for three large QTL mapping studies that dealt with resistance to northern leaf blight, Stewart's wilt and gray leaf spot. A new method is described for selecting families from a mapping

population that are important for multiple disease resistance. This method uses the phenotypic data, high density SNP marker data and QTL data to select for families that are resistant to multiple diseases and is termed as HEMP (High dEnsity Marker Phenotype) QTL analysis.

The second chapter describes the results from differential expression analysis of northern leaf blight resistant maize inbred NY22613 and susceptible maize inbred Oh7B. The analyses were conducted both in greenhouse and field conditions, since disease phenotypes and resistances can vary from the greenhouse to the field. The differential expression studies were followed by a series of gene ontology studies in which functions of the differentially expressed genes were analyzed.

The third chapter delves into the need to communicate quantitative genetics to the public and how the success stories in plant genetics can contribute to communicating quantitative genetics to a wider audience. This chapter starts by describing advances in low-cost genome sequencing and its effects on genetic studies. This is followed by a discussion of the differences between using genome wide association studies as a research tool as opposed to a diagnostic tool. Genome-wide association approaches are used in direct-to-consumer genetic testing and can result in misdiagnosis of major diseases such as cancer and heart disease. Following this is a discussion of how quantitative genetics science communication could help in improving the experience of people using direct-to-consumer testing. Finally, the chapter closes with a discussion of regulation of genomic technologies.

The fourth and final chapter is a synthesis of results and understandings from the first three chapters.

CHAPTER 1

GENETIC BASIS OF MULTIPLE DISEASE RESISTANCE IN MAIZE INBRED NY22613 USING QTL ANALYSIS AND A NOVEL SELECTION STRATEGY FOR FINE MAPPING MULTIPLE DISEASE RESISTANCE

1.1 Rationale and Background

Plant diseases are a major threat to global food production. It was estimated that plant diseases caused a crop loss of about 4-14% worldwide (Oerke *et al.* 2006). The United States is the largest producer of maize (*Zea mays* L.) and contributes about 40% to the global maize production. In the year 2013, U.S. and Ontario in Canada produced a record amount of maize, estimated at 14.2 billion bushels (NASS, 2015). New York State produced about 334 million bushels. Despite this record production, 1.1 billion bushels of maize were still lost to plant diseases (Wise, 2014). Maize diseases can be classified as root rots, seedling blights, stalk rots, ear rots and mycotoxin-producing fungal diseases. Among these, the most devastating were the leaf diseases, collectively causing a yield reduction of about 479.2 million bushels in the year 2013 (Wise, 2014). Among the leaf diseases, northern leaf blight (NLB) caused by *Setosphaeria turcica* (Luttrell) Leonard and Suggs (anamorph *Exserohilum turcicum*) resulted in the highest yield reduction of about 132.3 million bushels (Wise, 2014).

The United States and Canada have faced significant economic losses due to plant diseases. However, current losses due to plant diseases in these regions are not as significant as in the

tropical and equatorial regions. In Africa, crop losses due to diseases varied between 15 and 50%, whereas the global average of losses due to plant diseases was estimated to be only 9% (Fajemisin, 1986). Plant pathogens thrive in the warmer regions since they are ideal for the survival and reproduction of plant pathogens. For example, *Setosphaeria turcica* sexually reproduces in Uganda, resulting in the generation of new pathogen races (Martin, 2011).

Our knowledge about global crop losses due to plant diseases is very limited for several reasons. Some of the most important reasons are different data collection techniques followed by different research institutions and most of the smallholder farmers in the developing world do not quantify their yield or losses since their produce is mostly used for personal consumption (Pinstrup-Andersen, 2000). This poses a serious problem when it comes to designing a global strategy to combat plant pathogens (Yudelman *et al.* 1999).

Understanding the genetic basis of disease resistance in maize is important to breed for maize varieties and hybrids that are resistant to diseases. Since chemical management of diseases is expensive, the most widely used strategy to manage maize diseases is using genetic resistance (Ferguson and Carson 2007). In order to study the genetic basis of inheritance of a trait, traditionally geneticists have generated mapping populations to identify the regions of the genome that are important for the specific trait using quantitative trait locus (QTL) mapping. This can be followed by fine mapping or messenger RNA expression studies, which will eventually lead to identification of alleles that contribute to disease resistance in plants. Due to the recent developments in reduced cost expression studies and reverse genetics studies like the CRISPR-Cas9 gene knockout technique, identifying genes responsible for a certain trait is much faster and

more efficient than in previous years. Thus understanding the biology of disease resistance can aid breeders to generate cultivars that have more durable resistance to plant pathogens for their respective breeding target locations. Additionally, understanding the genetic basis of multiple disease resistance in maize will help us obtain deeper insights into the molecular basis of plant pathogen interactions and genetic control of correlated traits.

1.1.1 Complex nature of plant immunity to pathogens

In contrast to vertebrates, plants do not have adaptive immunity where macrophages or T-cells circulate throughout the plant system and provide immunity against pathogens. Resistance to pathogens is less complex but more specific in plants. Plants achieve immune responses that are highly specific to pathogen targets without a circulatory system and mobile immune response cells. They also achieve minimal self-reactivity and often a lifelong immune memory. The plant immune system is comprised of innate immunity where each cell detects the signals that are emitted by pathogens. The innate immunity in plants can be classified into pathogen associated molecular patterns triggered immunity (PTI) and effector triggered immunity (ETI) (reviewed by Spoel & Dong, 2012). PTI is triggered when pattern recognition receptors in plants recognize microbe/pathogen associated molecular patterns (MAMPs/PAMPs) in pathogens and result in immune responses. However pathogens have evolved to suppress these immune responses by injecting effector molecules or virulence factors targeting proteins that contribute to these immune responses. Plants overcome these effector molecule initiated pathogenies using ETI (reviewed by Jones *et al.* 2006). Immune memory and systemic immunity are achieved by plants through several pathways involved in a phenomenon known as systemic acquired resistance (SAR). In SAR, plants

accumulate salicylic acid (SA) and result in local hypersensitive response when infected with plant pathogens (Ross, 1961).

Genetic resistance to plant pathogens was first explained by Flor (1956) through the gene for gene hypothesis, which stated that for every avirulence (Avr*)* gene in the pathogen (effector gene) there should be a resistance (R) gene in the plant to achieve a successful resistance reaction. However, there are a virtually infinite number of effector molecules in the plant pathogen microbiome and the finite genome size of plants does not support this hypothesis.

The guard hypothesis proposes that R genes in plants are capable of detecting the effector molecules by not recognizing them directly, but by recognizing the signals from the target molecules of these effector molecules. Thus a finite number of R genes in the plant genome can indirectly detect the virtually infinite number of effector molecules (Van Der Biezen *et al.* 1998). R genes have a distinct structure in the plant genome. They occur in clusters, are mostly formed of tandem genes and have higher recombination frequency than the rest of the genome (Baumgarten *et al.* 2003). It has been hypothesized that this high recombination frequency leads to the formation of new R genes, which results in the coevolution of plants with pathogens (Pryor, 1994). R genes provide highly effective disease resistance to plants and are associated with qualitative disease resistance. R genes play a major role in contributing to disease resistance of several biotrophic and hemibiotrophic pathogens. However, there is no evidence that resistance to necrotrophic pathogens is contributed by R genes with the exception of *Arabidopsis thaliana* (L.) Heynh. *RESISTANCE TO LEPTOSPHAERIA MACULANS 3 (RLM3).* This gene codes for a Toll/interleukin 1 receptor domain R-protein and is associated with resistance to several

necrotrophs. Reactive oxygen species, hypersensitive response and autophagy contribute to the resistance in necrotrophs but they have contrasting roles in response to biotrophs (Staal *et al.* 2008).

Resistance to necrotrophic pathogens is associated with production of various secondary metabolites, antimicrobial peptides, hormones, ethylene, host specific toxins, cell wall degrading enzymes (CWDE), salicylic acid (SA) , abscisic acid (ABA), jasmonic acid (JA), reactive oxygen species, callose and various other cell wall modifications. The kinetics and timing of these responses are complex and vary depending on pathogens (Reviewed by Mengiste, 2012). It is important to note that some of these immune responses are also associated with resistance to biotrophic and hemibiotrophic pathogens. Additionally, not all the immune responses are unfavorable to pathogens of all modes of actions. For example, cell death through hypersensitive response, while resulting in a resistance reaction to several biotrophic pathogens, improves the susceptibility of plants to *Botrytis cinerea*, (de Bary) Whetzel) a necrotrophic pathogen in arabidopsis (Govrin *et al.* 2002). High levels of SA cause resistance reactions when infected by most hemobiotrophic pathogen but resulted in susceptibility to a necrotrophic pathogen (Ferrari *et al.* 2003; Veronese *et al.* 2004).

Hemibiotrophic pathogens have an early biotrophic stage followed by a necrotrophic stage. Hypersensitive response is effective against hemibiotrophic pathogens, unlike necrotrophic pathogens (Jia *et al.* 2000; Vleeshouwers *et al.* 2000). Production of reactive oxygen species by plants are effective for both hemibiotrophic and necrotrophic pathogens (Gorvin *et al.* 2002). Increased production of SA causes resistance to hemibiotrophic pathogens and susceptibility to

necrotrophic pathogens (Veronese *et al.* 2004). These studies suggest that multiple virulence mechanisms in different modes of pathogens lead to complex immune mechanisms in plants. Despite this complexity, plants do exhibit a significant differentiation in immune response based on mode of pathogenesis.

Plants can also achieve transgenerational immunity. Since plant gametes arise from somatic cells, unlike the preset embryonic germ line in animals, stress induced somatic homologous recombination in a portion of plant cells can generate gametes that can be successfully transmitted to the next generation. This transgenerational immunity has been demonstrated in *Arabidopsis thaliana* (Ries *et al.* 2000).

1.1.2 The maize genome and pathogens

Maize has a medium plant genome size of ~ 2.3 GB, however its complexity is exemplified by the unusually high activity of transposable elements. About 85% of the maize genome is comprised of hundreds of families of transposable elements (Schnable *et al.* 2009). Maize also demonstrates immense diversity. Two maize lines chosen at random are more genetically diverse on average than humans and chimpanzees (Buckler, 2006). The B73 maize genome has been sequenced and more than ~ 32000 genes have been detected. However, sequencing of 503 maize lines indicated that the maize pan-genome has an additional 8681 expressed genes that were absent in the genome of maize inbred B73 (Hirsch *et al.* 2014).

Pathogens and host plants undergo a coevolutionary arms race. This alternation of resistance and susceptibility has been demonstrated in various plant pathogen interactions (reviewed by Anderson *et al.* 2010). In the case of maize, sexual reproduction might play a major role in the survival of the crop despite multiple pathogen pressures across generations. Evolutionary advantages of sexual reproduction have been discussed by several authors (Fisher, 1930; Muller, 1932, Smith, 1978). The red queen hypothesis proposes that sexual reproduction enables organisms like maize to keep pace with both changes in their environment and threats like pathogens by constantly altering their genomes (Van Valen 1973; Bell 1982).

Models on cycling of alleles in sexual reproduction suggest that host genotypes that are rare survive when pathogens demonstrate more virulence. In other words, when there is an increased pathogen pressure, only the resistant individuals survive and they will be a small portion of the general population. Fisher hypothesized that sexual reproduction helps in the spread of advantageous alleles (1930). Recombination can help plants to achieve disease resistance since sexual reproduction results in plants fixing advantageous mutations (R genes) and eliminating detrimental mutations (Avr or S genes) in finite populations. R genes have a distinct structure in the maize genome. They occur as tandem genes and have higher recombination rates than the rest of the genome. It has also been hypothesized that pathogen pressure has contributed to fixation of recombination itself (which can be considered as a quantitative trait like disease resistance).

Some of the R –genes that condition resistance to a given pathogen in maize have been found in other plant species as well. For example, the same pathogen that causes NLB in maize, *Setosphaeria turcica*, also attacks sorghum (*Sorghum bicolor* (L.) Moench) and Brassica species.

The *(St)* genes that cause resistance to the pathogen in sorghum were found to be evolutionarily conserved in multiple plant species including maize and brassica. In sorghum, three pairs of six *St* genes encoding nucleotide-binding leucine-rich repeats (NB-LRR) were found on chromosome 5 and were known to cause resistance reactions to *S. turcica* treatment. However, *S.turcica* seems to favor pathogenesis towards maize rather than sorghum or brassica. *S.turcica* that affected both brassica and maize were analyzed and 628 maize-specific protein groups, including a number of potential effectors, were found to be secreted by the NLB pathogen. Thus, plant resistance to pathogens goes beyond the genome of a single plant species.

Breeders use R gene mediated resistance widely to combat plant pathogens due to the strength of the resistance genes against the pathogens. This type of resistance is also referred to as qualitative resistance. Wisser *et al.* (2006) catalogued all the qualitative resistance genes that had been identified in maize and found that at least 17 qualitative resistance genes had been mapped in diverse maize lines for diseases including maize streak virus, NLB, southern rust (*Puccinia polysora* Underw.) and common rust (*Puccinia sorghi* Schwein). Despite their wide use, only six R genes that are known to impart genetic resistance to plant diseases in maize have been cloned. This is mainly because map-based cloning is usually a time consuming and an expensive research activity (Moose and Mumm, 2008). These genes are *Rp1, Rp3, Rcg1, HM1, Rxo1 and Htn1* (Collins *et al.* 1999; Webb, 2002; Frey *et al.* 2011; Johal & Briggs, 1992; Leach & Hulbert, 2005; Hurni *et al.* 2015).

Genetic resistance to many maize diseases is largely quantitative (intermediate) in nature. About 437 QTLs were identified to provide resistance to 19 different maize diseases in the study

conducted by Wisser *et al.* (2006). These disease QTLs (dQTLs) are spread across both arms of all ten maize chromosomes. These dQTLs might contain gene clusters that result in quantitative resistance to diseases. The molecular mechanisms that control quantitative resistance are poorly understood in maize.

Even though R-genes have been successfully used in breeding disease resistant crop varieties, the lack of durability of R-genes over time has been a major issue for breeders. Pathogens tend to overcome the host resistance conferred by R-genes more quickly than the host resistance conferred by dQTLs. Most R-genes are also race specific. When the Ug99 race of stem rust (*Puccinia graminis* f.sp. *tritici* Erikss. & Henning, Z. Pflanzenkrankh) in wheat was first discovered, the lack of R-genes for the pathogen race at that time generated an international panic (Singh *et al.* 2011). Similarly, in maize a number of NLB resistance-related genes such as *Ht, Ht2*, and *Htn1* have been demonstrated to be race specific. In an environment with unpredictable pathogen pressures, it is highly risky to use R-genes for disease resistance or multiple disease resistance.

In a meta-analysis of 50 publications that were related to various maize disease resistance studies, it was found that disease resistance QTL were distributed among all ten chromosomes in maize and covered about 89% of the genetic map to which the data were anchored (Wisser *et al.* 2006). Nested association mapping (NAM) for northern leaf blight showed that 29 QTL, some of them with multiple alleles, were responsible for NLB resistance in maize (Poland *et al.* 2011). Sixteen QTL were found to be associated with GLS disease resistance (Benson *et al.* 2015). QTL mapping using joint linkage and stepwise regression analysis in the NAM population in the US and China

showed that 49 QTL were associated with SLB resistance and 48 QTL were associated with NLB resistance (Li *et al.* 2018).

However, sometimes disease resistance seems to be governed by colocalization of both R-gene and dQTLs in multiple crop systems like rice, potato and maize (Wang *et al.* 1994; Gebhardt and Valkonen. 2001, Xiao *et al.* 2007). One of the hypotheses for this occurrence is that QRLs might be weaker forms of R-genes. Understanding the relationship between QRLs and R-genes in contributing to durable disease resistance is of major interest.

1.1.3 Multiple Disease Resistance in Maize

Multiple disease resistance (MDR) is defined as host plant resistance to two or more diseases. This definition includes both partial and complete resistance, but not non-host disease resistance. MDR can be either qualitative or quantitative resistance. MDR is a valuable but rare trait in plant breeding (Wiesner-Hanks and Nelson, 2016).

Genetic control of correlated traits like multiple disease resistance is hypothesized to be due to linkage of multiple genes or pleiotropy (Bernardo, 2010). Linkage of some 14 closely linked loci at the *Rp1* locus (*Rp1A to Rp1N*) has been identified to cause resistance to common rust in maize. These loci were initially considered to be allelic but later suspected to be closely linked loci (Hagan and Hooker 1965; Hooker 1985; Hulbert and Bennetzen 1991). A similar locus named *Rp3 (Rp3a to RP3f)* on Chromosome 3 is associated with rust resistance and also speculated to be associated with virus resistance in maize (Redinbaugh *et al.* 2003).

Pleiotropy refers to the effect of a gene on more than one phenotype. Pleiotropic QTLs that contribute to multiple disease resistance are less common or more difficult to identify than the multiple QTLs contributing to disease resistance. However, evidence for MDR in plants governed by pleiotropy has been demonstrated in various crops in different studies. The *Lr34* gene in wheat (*Triticum aestivum* L.) was found to be associated with resistance against leaf rust (*Puccinia triticina* (= *P. recondita* Roberge ex Desmaz. f. sp. *tritici*.)), stripe rust (*Puccinia striiformis* Westend.), and powdery mildew (*Blumeria graminis* (DC.) Speer f. sp. *tritici* (Em. Marchal)) (Sucher *et al.* 2017). A family of germin-like proteins were found to be associated with sheath blight (*Rhizoctonia solani* AG1-1A Kühn (Teleomorph: *Thanatephorus cucumeris* (A. B. Frank) Donk.)) and rice blast disease (*Magnaporthe oryzae* H. Barr, (anamorph: *Pyricularia oryzae C. Saccardo*)) in rice (*Oryza sativa* L.) (Manosalva *et al.* 2009). The *glutathione S transferase (GST)* gene was found to be associated with modest levels of resistance to southern leaf blight (SLB) (anamorph: *Bipolaris maydis* (teleomorph: *Cochliobolus heterostrophus*), NLB and gray leaf spot (GLS) ((*Cercospora zeae-maydis*) Tehon and Daniels) (Wisser *et al.* 2011). The *pan1* gene was implicated for resistance to NLB and Stewart's wilt (Jamann *et al.* 2014). *ZmREM6.3*, a remorin gene was shown to be associated with resistance to NLB, Stewart's wilt and common rust (Jamann *et al.* 2016). The *ZmCCoAOMT2* gene that encodes a caffeoyl-CoA O-methyltransferase confers quantitative resistance to both southern leaf blight and gray leaf spot (Yang *et al.* 2017). In maize, several QTLs have been associated with multiple disease resistance. For example, *qNLB1.06 Tx303* is associated with disease resistance to common rust, northern leaf blight and Stewart's wilt. *qNLB 1.02* showed resistance to Stewart's wilt and common rust (Chung *et al.* 2010). This evidence suggest that pleiotropy plays a role in governing multiple disease resistance in maize.

One hypothesis that explains multiple disease resistance due to pleiotropy is that the resistance may relate to genes that code for or interact with genes that code for transcription factors that are involved in vital physiological functions like regulating hormone signaling, cell death, hypersensitive response, and sugar signaling and partitioning.

1.2 Maize Inbred NY22613

Maize inbred NY22613, developed at Cornell University, is resistant to anthracnose leaf blight caused by *Colletotrichum graminicola* Ces. Wils., gray leaf spot (GLS), caused by ((*Cercospora zeae-maydis* Tehon & E.Y. Daniels. and *C. zeina))*, common rust caused by *Puccinia sorghi*, northern leaf blight (NLB) caused by *Setosphaeria turcica* and Stewart's wilt (SW) caused by *Pantoea stewartii*. The northern leaf blight disease resistance demonstrated by inbred NY22613 is depicted in Figure 1.1. Susceptible maize inbred Oh7B and resistant inbred NY22613 were planted in summer 2016 in Chemung, NY. The natural inoculum of NLB caused significant leaf damage to the susceptible plants and caused significantly less damage to the adjacent resistant plants.

Figure 1.1 *The susceptible Oh7B (left) and the resistant NY22613 (right) inbred infected with northern leaf blight (NLB).*

Setosphaeria turcica is an ascomycete, hemibiotropic fungal pathogen, *Cercospora zeae-maydis* is a haploid ascomycete fungus and *Pantoea stewartia* is a bacterium. Inbred NY22613 also demonstrated resistance to anthracnose leaf blight and anthracnose stalk rot, caused by an ascomycete, *Colletotrichum graminicola*, and Common rust caused by a biotrophic basidiomycete pathogen called *Puccinia sorghi* (Yang, 2007). The genetic basis of GLS resistance in NY22613 is known to be additive in nature through generation means analysis (Palanichamy, 2014).

1.2.1 Pedigree of maize inbred NY22613

The pedigree of maize inbred NY22613 is depicted in Figure 1.2. NY22613 was previously referred to as NYRD6613.

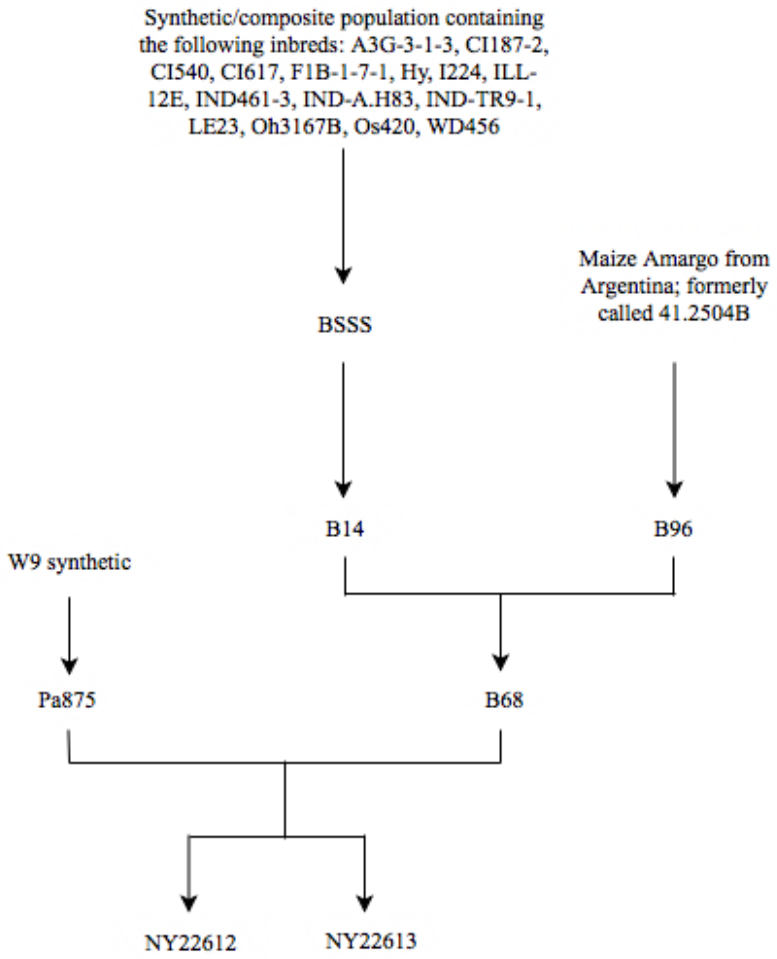

Figure 1.2 *Pedigree of maize inbred NY22613.*

NY22613 is an S4 line derived from a cross between Pa875 and B68. It was developed by alternating generations of selfing in New York with a generation of selection among selfed progeny rows by screening conducted collaboratively by Dr. J.E. Hawk (University of Delaware) and Dr. E.L. Stromberg (Virginia Polytechnic Institute and State University). The parent inbred

Pa875 was developed by using the W9 synthetic as the source material (Figure 1.2). The Pa875 parent was initially selected due to its GLS disease resistance. The parent B68 was developed by crossing B14 and B96. B14 was derived from the BSSS population and B96 was derived from Maiz Amargo from Argentina (formerly called 41.2504B).

1.2.2 Characteristics of maize inbred NY22613

NY22613 is a tall inbred with horizontal leaves and a tendency to tiller. The plants have yellow green leaves and sheaths are sun-red at the base of the plant. The leaves are long and slightly wavy and tips are slightly curved. The stalks are straight. The tassel shape is spiky and upright. About two to seven branches were observed in the tassels. The tassel is green and the anthers are yellow. The glumes were sun-red. Most plants had good pollen production. The plants flowered at about 86 to 88 days at Aurora, New York in 1996, which is similar to B73. Leaf stay green character during disease infestation is considered good. Ears are long with about 16 kernel rows, white cobs, and dull orange-yellow kernels. A few kernels have large red streaks. The kernel shape is small and square with deep dents.

The silk color is sun-red. Most plants had two ears. The ear fill was variable (half to almost full ear fill). The ear length was 12.7- 17.7 cm. General seed quality is considered as good. Table 1.1 shows NY22613 disease rating data, revealing that NY22613 is resistant to both ALB and NLB. NY22613 ranked second for NLB and third for ALB among the inbreds evaluated.

Table 1.1 *Resistance assessment for northern leaf blight (NLB), southern leaf blight (SLB) and anthracnose leaf blight (ALB) in 2006 under greenhouse conditions (IP = incubation period, AUDPC-PDLA = the area under the disease progress curve based on the percent diseased leaf area).* [a]

Line	IP				AUDPC-PDLA					
	NLB		ALB[b]		NLB		ALB		SLB	
	Day	Rank	Day	Rank	Score	R	Score	Rank	Score	Rank
73405	12.5	1	6.0	4	24.1	1	48.0	4	81.7	4
NY22613	**11.4**[c]	**2**	**6.7**	**3**	**42.6**	**3**	**50.3**	**5**	**95.6**	**6**
67099	11.3	3	5.7	5	40.8	2	68.4	8	98.3	7
IBM054	10.3	4	7.0	2	74.5	4	18.5	1	86.4	5
BRG3-34	10.0	5	4.3	10	117.9	5	92.6	10	105.4	8
86368	9.6	6	5.7	6	146.2	6	40.1	2	70.2	2
Mo17	9.5	7	5.3	8	173.1	7	42.3	3	56.7	1
E45A	8.8	8	4.5	9	164.1	8	87.4	9	79.2	3
B73	8.3	9	7.3	1	267.6	9	52.5	6	154.6	10
Checks	7.5	10	5.5	7	273.2	10	60.4	7	121.0	9

[a] Source: Yang (2007)

[b] IP data from the spray inoculation method was used for ALB. Other data are means of the two inoculation methods.

[c] Inbred NY22613 resistance scores are underlined (Yang 2007).

Table 1.2 *Inbred NY22613 and control inbred responses to GLS in field conditions in 2005 and 2006 and in greenhouse conditions in 2006.* [a]

Inbred line	Derivation	Disease Response	2005 Field	2006 Field	2006 Greenhouse
B73	BSSS C5 (Reid)	Generally s[b]	+[c]	+	+
Mo17	C.I 187-2 X C103	Generally r[b]	+	+	+
NY22613[e]	***Pa875 X B68***	***GLS r***	**+**	**+**	**+**
67099	CIMMYT[d] Pop.41 NTR-1	NLB r	+	+	+
73405	B59 X Oh43Ht2	SLB r	+	+	+
86368	CIMMYT MBR	ASR r	+	+	+
E45A	W153R X A632	Eyespot r	+	+	+
BRG3-34	Insect Resistance germplasm, CIMMYT	Common rust r	+	+	+
RD4058	Mo17 X W153R	ALB r	+	+	-
ECB6-863	A632Ht X 86409	B-	+	+	-
CO441	Jacques 7700 X CO298	Gibberella ear rot r	+	-	+
B52	Segregating material from a private breeder	NLB s	+	+	-
W64AHt	Wf9 X C.I. 187-2	SLB s	+	+	-
LH119	B73(2) X H93	ALB s	+	+	-
W117Ht	643 X Minnesota No.13	Eyespot and common rust s	+	+	+
B37	BSSS	ASR r	+	-	-
RD4503	VA59 X W153 R	Eyespot and common rust s	-	-	-
IBM 054	B73 X Mo17	NLB r	-	-	+
IBM 262	B73 X Mo17	NLB s	-	-	+

[a] Source: Yang (2007)

[b] s – susceptible ; r – resistant.

[c] + - present in that experiment; - - not present in the experiment

[d] CIMMYT – International Maize and Wheat Improvement Center; NTR = Northern Temperature Region; MBR: Multiple Borer Resistant Population

[e] Inbred NY22613 resistance is bolded.

Table 1.3 *Evaluation of the multiple disease resistant inbred NY22613 for disease resistance in 2005 and 2006.*

	GLS resistance ratings (0-5 Scale)			
Inbred	**Inbred Per Se**	**B73 test cross**	**Iodent test cross**	**Mo17 test cross**
NYRD6600	2.3	3.2	2.9	2.9
NY22613	1.6	3.4	3.2	2.9
Pa875 (Resistant Check)	1.2	3.5	3.1	3.3
B73 (Susceptible Check)	4.2	-	-	3.7
Hybrid Checks				
Agway 792	3.3			
Pioneer 3394	4.1			

Table 1.1, 1.2 and Table 1.3 show GLS rating data for NY22613 and other inbreds per se and for their testcrosses. Collectively, these data show that NY22613 is resistant to GLS across three different growing conditions. Additional evidence for GLS resistance is presented in Table 1.4 where inbred NY22613 exhibits resistance to GLS in multiple years across multiple environments.

Table 1.4 *Test cross ratings for inbreds PA875, NY22612 and NY22613 in a scale of 0-5 in two replications across two environments. Mo17 was the tester (Blacksburg, Virginia (VA) and Chemung, Pennsylvania (PA)) in years 2005, 2006, 2007 and 2008.*

Inbred	Weighted GLS Disease Rating (0-5 Scale)
PA875	3.3
NY22612	2.93
NY22613	2.77

Inbreds NY22612 and NY22613 have high resistance to GLS, equal to or better than the resistant check Mo17 (Yang 2007).

In addition to the resistance demonstrated by the MDR inbred NY22613 to leaf diseases, it also exhibits resistance to anthracnose stalk rot (ASR). Figure 1.3 depicts the resistance demonstrated by inbred NY22613 to ASR. These plant stalks were injected with a spore suspension of *Colletotrichum graminicola* with a metal syringe usually used to vaccinated large animals. As the figure shows, the resistant inbred had little ASR spread upwards from the basal injection site while the susceptible inbred had significant ASR lesions.

Figure 1.3 *Stalks of resistant inbred (NY22613) and susceptible inbred (Oh7B) inoculated with ASR.*

From these studies it is clear that maize inbred NY22613 has resistance to multiple diseases and can be considered as a multiple disease resistant inbred. The goal of this study was to determine the genetic basis of multiple disease resistance in NY22613.

1.3 Materials and Methods

1.3.1 OHNY mapping population

A bi-parental linkage mapping population titled OHNY mapping population was generated from the inbred NY22613 and Oh7B to study multiple disease resistance in maize. These near-isogenic lines were developed from the F_1 population generated from NY22613 and Oh7B that was backcrossed to the recurrent parent (Oh7B) three times and then selfed three times. The BC_3S_3 population is expected to have 6.25 percent of the donor genome (NY22613). Thus the population is not very heterozygous and due to the selfing is fairly homogenous. The OHNY population is a BC_3S_3 mapping population of Oh7B and NY22613 and was genotyped using genotyping by sequencing (GBS) at the 384-plex (ApeKI) platform. Planting, inoculation and phenotyping information can be found in Table 1.5.

Table 1.5 *Planting, inoculation and phenotyping dates for NLB, GLS and SW in the years 2016 and 2017.*

Disease	**Year**	**Planting**	**Inoculation**	**Phenotyping**		
				Day-1	**Day-2**	**Day-3**
NLB	2016	25-May	7-Jul	3-Sep	20-Sep	29-Sep
GLS	2016	11-May	Natural inoculum	9-Sep	21-Sep	6-Oct
SW	2016	28-May	26-Jul	5-Sep	11-Sep	22-Sep
NLB	2017	10-Jun	26-Jul	17-Sep	5-Oct	12-Oct
GLS	2017	2-Jun	Natural inoculum	29-Aug	4-Sep	12-Sep
SW	2017	10-Jun	26-Jul	26-Sep	4-Oct	17-Oct

1.3.2 Pathogen preparation and inoculation

Spore suspension for northern leaf blight was prepared as described by Chung *et al.* (2010). Plants were inoculated with NLB at the mid-whorl stage using 180,000 spores per ml liquid inoculum and one ml of NLB-infected sorghum grain solid inoculum. Both liquid and sorghum grain inoculum were deposited into the whorl of the plant. For Stewart's wilt, pathogens were prepared as bacterial cultures and inoculated with the pinprick method using a special device designed at Cornell University (See Figure 5.6 in Appendix). For GLS, natural inoculum was used for screening.

1.3.3 Phenotyping and QTL mapping

Field trials for NLB and SW were conducted during 2016 and 2017 at Cornell University at the Robert B. Musgrave Research Farm in Poplar Ridge, NY. Due to severe drought conditions in 2016, research farm fields were irrigated with 2.5 cm and 3.8 cm of overhead irrigation on 30 June and 25 July, respectively. Field trials for GLS were conducted during 2016 and 2017 in a farmer's field in Chemung, NY. All trials were laid out in an augmented incomplete block design with two replications for each disease in each year. Plants were hand planted due to limited availability of seed in 2016. Seeds were increased in 2016 and the plants were machine planted in 2017. Twenty plants per row were planted but only the first ten plants were evaluated for disease resistance.

NLB trials in 2016 were planted on May 25 and were inoculated with *Setosphaeria turcica* (race 1) on 7 July as described by Chung *et al.* (2010). The plants were phenotyped three times during the growing season (3 September, 20 September and 29 September). In 2017, trials for NLB were

planted on 10 June and were inoculated on 26 July. The plants were phenotyped on 17 September, 5 October and 12 October.

GLS trials in 2016 were planted on May 11. Disease occurrence was due to natural inoculum of *Cercospora zeae-maydis*. The plants were phenotyped three times during the growing season: 9 September, 21 September and 6 October. In 2017, trials for GLS were planted on 2 June. Disease occurrence was due to natural inoculum of *Cercospora zeae-maydis*. The plants were phenotyped on 29 August, 4 September and 12 September.

SW trials in 2016 were planted on May 28 and were inoculated on 26 July using a bacterial culture of *Pantoea stewarti*. The plants were phenotyped three times during the growing season: 5 September, 11 September and 22 September. In 2017, trials for SW were planted on 10 June and were inoculated on 26 July. The plants were phenotyped on 26 September, 4 October and 17 October.

Percentage disease leaf area (PDLA) was estimated by observing the diseased leaves of the plants. First ten plants in a row were considered for estimating the PDLA. Leaves that are present over the ear were observed and based on the lesions on the leaves, the PDLA was estimated. Variation attributable to different raters was minimized by using one individual to phenotype all the plants in the mapping population for all the diseases in both years.

Area under the disease progress curve (AUDPC) was calculated using the formula:

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

where y_i = percent diseased leaf area i, $(t_{i+1} - t_i)$ = number of days between two ratings, n = total number of ratings. Best linear unbiased predictors (BLUPs) of the lines were calculated for each disease by fitting a mixed model with AUDPC values as the response, 'year' as a fixed effect and 'line' as a random factor using the lme4 package in R statistical software (Version 1.1.447) (Bates *et al.* 2004; Rstudio Team 2015). These BLUPs were used to conduct a stepwise regression for QTL mapping. This analysis was conducted using TASSEL 5 software (version 20161020) (Glaubitz *et al.* 2014).

Total genotypic variance explained for each trait by the corresponding QTL was obtained by calculating the percentage variance captured by these QTL in a mixed model. More specifically, we constructed a relationship matrix using only the QTL and measured its variance contribution to the trait through a mixed effect model.

1.3.4 Allele analysis

R statistical software (Version 1.1.447) was used for allele analysis (Rstudio Team 2015). Clustering analysis was conducted using K-means algorithm and GAP statistics (MacQueen *et al.* 1967, Tibshirani *et al.* 2000). GAP statistics were used to determine the number of clusters and K-means was used to group the SNPs based on the AUDPC values. Individuals with SNPs in the lowest disease resistance group were considered more important for multiple disease resistance. Figure 1.4 shows the steps involved in the clustering analysis.

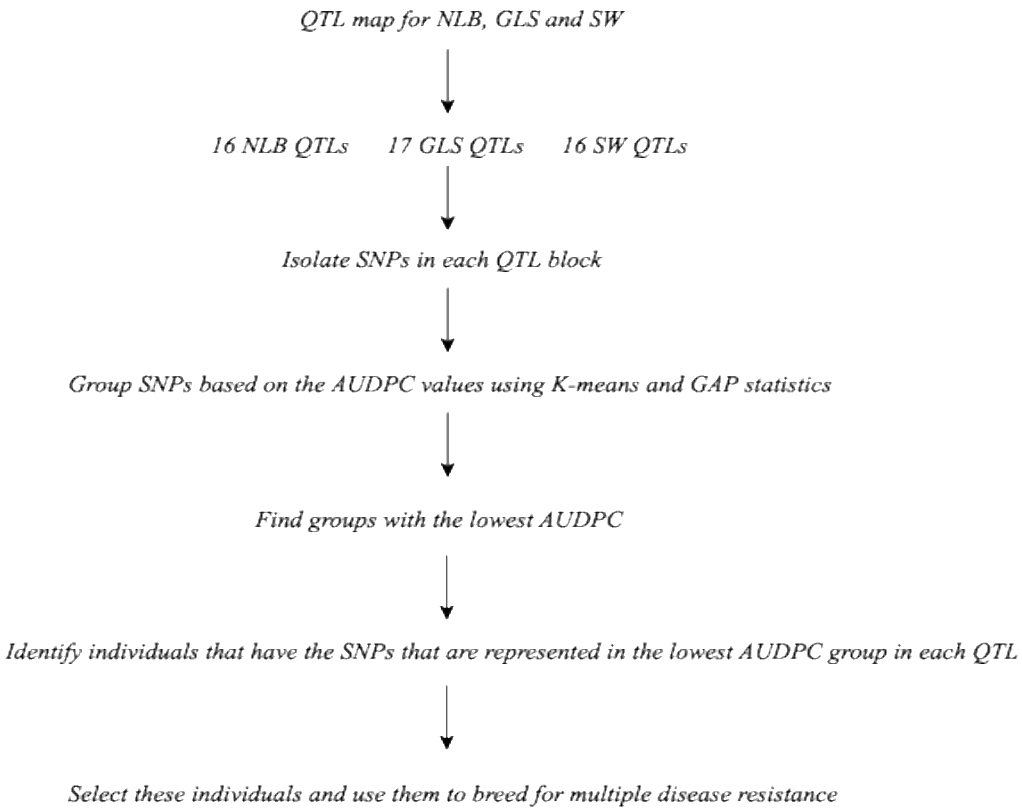

Figure 1.4 *Clustering analysis to select for MDR families using SNPs, AUDPC and QTL data.*

Using the variances calculated from the AUDPC values, broad sense heritability (H) was calculated using the formula:

$$\frac{V_G}{\frac{V_E}{(y * r)} + V_G}$$

where V_G is the genetic variance, V_E is the residual variance, y is the number of years and r is the number of replications for each disease. Gene ontology analysis was conducted using the protein annotation through evolutionary relationship (PANTHER) classification system described by Mi *et al.* (2013).

1.4 Results

The broad sense heritability for northern leaf blight in the mapping population was 0.55, gray leaf spot was 0.39 and Stewart's wilt was 0.49. The narrow sense heritability (h^2) was calculated based on the additive genomic relationship between the SNP markers and the traits. Narrow sense heritability for NLB was 0.37, GLS was 0.30 and SW was 0.15.

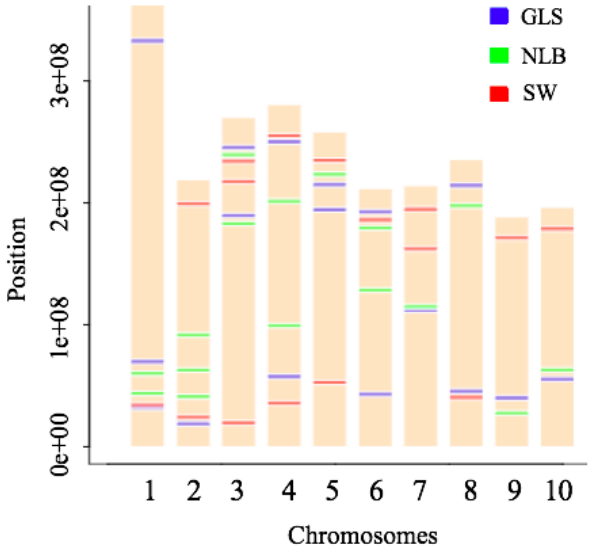

Figure 1.5 *QTL map showing disease resistance QTL on the ten chromosomes of maize.*

QTL mapping was conducted using stepwise regression and 49 QTLs were detected for NLB, GLS and SW. Sixteen QTL were associated with NLB resistance, 17 QTL were associated with GLS resistance and 16 QTL were associated with SW resistance. No QTL were colocalized for all the three diseases. The percentage of phenotypic variance explained by the 16 QTL for NLB was 0.29, for the 17 QTL for GLS it was 0.27 and for the 16 QTL for SW it was 0.23.

Table 1.6. *QTL that were identified for NLB, GLS and SW disease resistance in the NY22613 X Oh7B biparental mapping population.*

Disease	Locus	Physical map	Chromosome location	P-value
NLB	1	37720906	1:37717388..37725221	9.64E-10
NLB	1	19979634	1:19978997..19979751	1.29E-08
NLB	2	2484151	2:2476120..2492538	1.43E-07
NLB	2	49057160	2:49016939.. 49057266	4.54E-11
NLB	2	55560345	2:55560318..55560440	3.56E-06
NLB	3	223740910	3:223740841..223745642	7.92E-07
NLB	3	171900654	3:171900438..171900685	2.18E-06
NLB	4	185566213	4:185529728..185649204	2.55E-09
NLB	4	79449676	4:79162068..79451618	1.40E-06
NLB	5	215567451	5:215567100..215620081	1.24E-10
NLB	6	148650827	6:148633043..148654966	6.75E-11
NLB	6	98010324	6:98007087..98010327	2.45E-09
NLB	7	100081042	7:100035505..100170373	9.55E-08
NLB	8	172365897	8:172365300..172365909	2.32E-10
NLB	9	8974761	9:8778978..8974937	3.25E-20
NLB	10	36380036	10:36259906..36380814	4.09E-17
GLS	1	46073541	1:46009342..46073629	5.65E-08
GLS	1	2694428	1:2694098..2694439	2.77E-07
GLS	1	300744863	1:300744838..300799862	3.42E-06
GLS	2	4257737	2:4211875..4257983	4.06E-11

GLS	3	165128802	3:165128785..165128833	3.26E-09
GLS	3	215437040	3:215437034..215453883	5.65E-07
GLS	4	234908817	4:234908733..234911336	5.80E-08
GLS	4	43590931	4:43590715..43591420	4.78E-07
GLS	5	199649696	5:199649599..991649708	6.55E-12
GLS	5	165854752	5:165655778..165978258	1.06E-06
GLS	6	17002706	6:17002694..17165743	1.44E-09
GLS	6	164646494	6:164646431..164675590	5.52E-07
GLS	7	100081042	7:100035505..100170373	6.45E-12
GLS	8	171615123	8:171613342..171617688	4.64E-07
GLS	8	8860232	8:8860221..8887120	1.79E-05
GLS	9	8974761	9:8778978..8974937	2.00E-22
GLS	10	36380036	10:36259906..36380814	1.41E-15
SW	1	10996800	1:10995565..11170650	8.26E-10
SW	2	4257737	2:4211875..4257983	2.25E-06
SW	2	176239002	2:176083016..176240477	5.40E-05
SW	3	220747659	3:220747650..220751436	2.98E-06
SW	3	196347378	3:196346322..196351670	5.92E-06
SW	3	2720513	3:2720421..2720554	4.85E-05
SW	4	238121510	4:238121295..238123025	7.35E-10
SW	4	10455637	4:10250168..10485810	8.16E-06
SW	5	34098699	5:33968688..34135549	5.37E-07
SW	5	213052061	5:213051764..213056452	1.02E-05

SW	6	162141592	6:162108496..162152305	2.15E-06
SW	7	166382910	7:166382883..166383776	4.72E-07
SW	7	133357660	7:133271978..133376044	2.59E-06
SW	8	14594610	8:14390220..14728266	5.51E-07
SW	9	153303317	9:153302987..153303323	1.00E-10
SW	10	140985791	10:140985769..140992671	1.63E-08

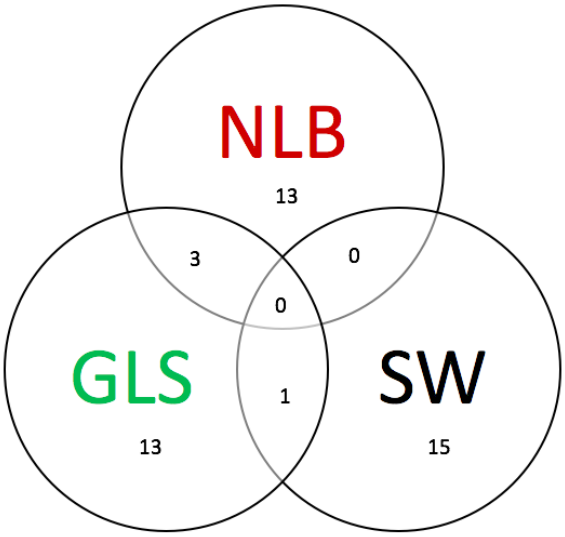

Figure 1.6 *Number of quantitative trait loci (QTL) shared among northern leaf blight (NLB), gray leaf spot (GLS) and Stewart's wilt (SW) resistance.*

One QTL was colocalized for GLS and SW and three QTL were colocalized for GLS and NLB. No QTL were colocalized for NLB and SW. The colocalized QTL for GLS and SW is located on chromosome 2 between markers S2_4211875 and S2_4257983. The colocalized QTL for GLS and NLB are located on chromosome 7 between markers S7_100035505 and S7_100170373, on

chromosome 9 between markers S9_8778978 and S9_8974937 and on chromosome 10 between markers S10_36259906 and S10_36380814.

Table 1.7 *QTL that are colocalized for disease resistance*

Disease	**Chromosome**	**Chromosome Location**
GLS and SW	2	2:4211875..4257983
GLS and NLB	7	7:100035505..100170373
GLS and NLB	9	9:8778978..8974937
GLS and NLB	10	10:36259906..36380814

Ensembl Genomes (https://plants.ensembl.org/Zea_mays) database was used to search for genes located in the individual QTL regions (Kersey *et al.* 2018). QTL regions that were associated with NLB contained 21 unique genes. Gray leaf spot associated QTLs contained 979 unique genes. QTLs associated with Stewart's wilt contained 2217 genes. (See Appendix for table 5.3 for gene lists). We used the same stringency for NLB, GLS and SW are still we ended up getting a disproportionate number of genes for NLB than the GLS and SW stepwise regression analysis. This reason for this disproportionate number of genes for NLB when compared to GLS and SW is unknown at the moment.

1.4.2 Selection for multiple traits based on BLUP values and high density marker information

Table 1.8 *Lines that were selected for disease resistance based on BLUP values. These lines had more disease resistance than the susceptible inbred Oh7B. Bolded underlined lines were selected for more than one disease resistance.*

Disease	**Lines**
Gray leaf spot	OHNY05, OHNY06, OHNY71, OHNY83, OHNY95, OHNY109, OHNY140, OHNY146, OHNY169, OHNY171, OHNY184, OHNY212, OHNY277
Northern leaf blight	OHNY97, OHNY128, OHNY169
Stewart's Wilt	OHNY02, OHNY05, OHNY06, OHNY15, OHNY24, OHNY26, OHNY28, OHNY30, OHNY34, OHNY35, OHNY56, OHNY57, OHNY63, OHNY65, OHNY72, OHNY75, OHNY80, OHNY82, OHNY92, OHNY96, OHNY103, OHNY111, OHNY126, OHNY148, OHNY155, OHNY167

Individual families in a mapping population can be selected for single traits such as NLB, GLS or SW disease resistance using phenotypic data such as best linear unbiased predictors (BLUPs) of the AUDPC values or just AUDPC values.

However, as observed in table 1.8, the disease resistant individuals in the mapping population are rarely correlated for all the three diseases. Only family OHNY803 was found to be resistant to more than one disease in our study. Here we demonstrate a High dEnsity Marker Phenotype (HEMP) QTL selection strategy that uses marker, QTL and phenotypic data for selection.

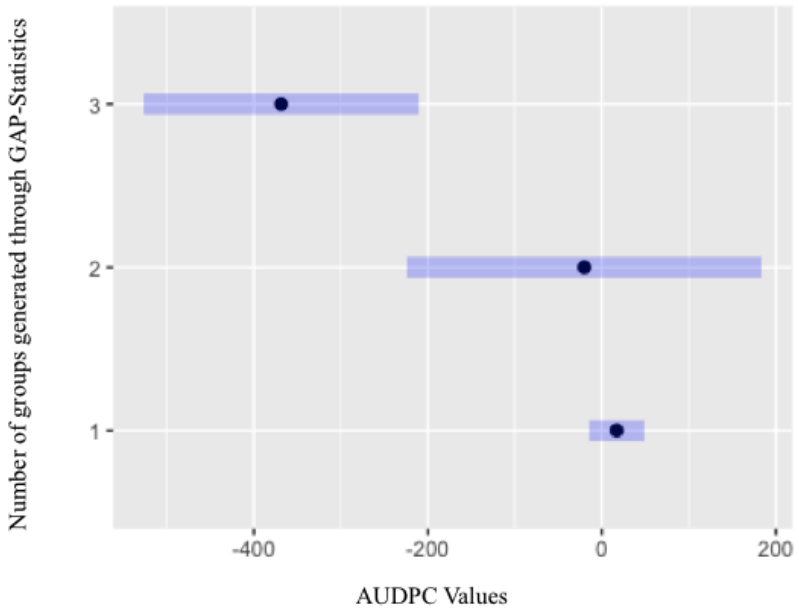

Figure 1.7 *Clustering of SNPs in one example block (QTL) that is associated with NLB resistance. Individuals with SNPs in group three (group with lowest AUDPC values) can be selected for breeding for MDR.*

This analysis was repeated for all 49 QTLs and by comparing the low disease resistance group for each disease, we were able to identify selected BC_3S_3 families and the alleles that are most important for multiple disease resistance. In HEMP QTL analysis, SNPs that are present inside the QTL regions are clustered using GAP statistics and K-means based on their AUDPC scores (See Fig. 1.4). Figure 1.7 shows the clustering in one example QTL bock that is associated with NLB

disease resistance. Plants from the most resistant clusters are selected for multiple disease resistance.

Table 1.9 *Families selected for multiple disease resistance using the best linear unbiased predictors (BLUP) selection strategy and using HEMP QTL selection.*

Families selected for all three correlated traits (NLB, GLS and SW resistance) using disease BLUPs	NONE
Families selected for three correlated traits (NLB,GLS and SW resistance) using HEMP QTL selection.	OHNY07, OHNY09, OHNY17, OHNY19, OHNY25, OHNY31, OHNY36, OHNY43, OHNY59, OHNY60, OHNY134, OHNY135, OHNY150, OHNY159, OHNY163, OHNY171, OHNY182, OHNY192, OHNY193, OHNY197, OHNY217, OHNY238, OHNY245

NLB: Northern Leaf Blight, GLS: Gray Leaf Spot, SW: Stewart's Wilt

Through HEMP QTL analysis, alleles from the group with the lowest AUDPC values can be considered more important for multiple disease resistance. This strategy helps to select more individuals that are associated with multiple disease resistance and thus increases the probability of selecting individuals with more beneficial alleles. Table 1.9 shows that HEMP QTL analysis was able to select 23 families that have the alleles that are responsible for multiple disease

resistance. This in turn might accelerate the development of genotypes with multiple quantitative disease resistance traits.

1.5 Discussion

1.5.1 Phenotyping for disease resistance

Phenotyping through estimating percentage diseased leaf area by manual observation for disease resistance has been the traditional way to phenotype for leaf diseases in maize (Chung *et al.* 2010). However this technique is a high skill task and the skill set is not scalable (Poland and Nelson 2011). For this study, phenotyping of NLB, GLS and SW were all conducted by one person during both growing seasons. We used the same environments for two years: Musgrave Research Farm in Poplar Ridge, NY for NLB and SW evaluation and Chemung NY for GLS evaluation. The BLUPs were calculated for these environments and were used for the QTL analysis.

New advances in imaging and image analysis have increased the potential to phenotype large populations accurately (Palanichamy and Cobb 2015). An automated high-throughput plant phenotyping system, that uses a computational pipeline of convolutional neural networks (CNNs) was used to classify images of NLB infected maize plants acquired in the field (DeChant *et al.* 2017). Phenotyping for multiple diseases in multiple crop species has been achieved by using images of diseased plant leaves collected in controlled conditions using CNNs algorithms (Mohanty *et al.* 2017). Using high throughput phenotyping and automated image analysis systems to phenotype multiple diseases in field conditions is not far away. These could improve the quality

of new phenotyping and reduce the time and effort spent on phenotyping populations for multiple disease.

Association mapping using increased marker densities can yield relatively more QTL with higher resolution for disease resistance (Li *et al.* 2018). However both association mapping populations and biparental mapping populations have been successfully used to identify genes responsible for disease resistance in maize (Chung *et al.* 2010;Jamann *et al.* 2014; Poland *et al.* 2011) In our study, unlike association mapping populations, all the individuals in the BC_3S_3 mapping population were closely related. Thus, the population structure and consequently the flowering time (~86 days) was almost the same and flowering time was not used in the mapping of QTL.

1.5.2 Selection for multiple traits using HEMP QTL selection strategy

Several types of phenotypic selection methods have been effective in breeding crop varieties with multiple desirable traits (Suwantaradon *et al.* 1975; Smith *et al.* 1981; Kumar *et al.* 2016; Bernardo 2010). Phenotypic selection methods have especially worked when genetic and genomic resources are not available. Due to the genomic resources available in maize, breeders can make use of genetic and phenotypic selection to breed for multiple disease resistance. Breeding for a maize inbred with multiple quantitative disease resistance traits whose inheritance is controlled by several hundred genes throughout the maize genome using only phenotypic selection is challenging. In our case, maize inbred NY22613 was already resistant to multiple diseases. However, selection of families within the NY22613 X Oh7B mapping population for multiple disease resistance using phenotypic scores alone was challenging. We demonstrated HEMP QTL

selection strategy which uses QTL data, high density marker data and phenotypic data to select for families with multiple disease resistance. This methodology can be used to select for individuals with multiple quantitative traits in any similar mapping population with QTL data, high density marker data, and phenotypes. With relatively rapidly changing weather patterns and unpredictable pathogen pressures, it is important to breed for crop plants with multiple favorable traits such as multiple disease resistance. In HEMP QTL analysis, utilizing high density SNP marker data along with phenotypic data and QTL data in the selection process will increase the probability of finding genotypes with multiple favorable alleles for multiple traits.

1.5.3 Gene Ontology Analysis

PANTHER is a comprehensive system that facilitates analysis of multiple genomes. PANTHER combines gene function, ontology, biological pathways and statistical analysis tools and aids in the analysis of gene expression experiments (Mi *et al.* 2013). The genes that were identified in the QTL regions for NLB, GLS and SW resistance were analyzed to see the nature of the genes identified. Unique genes that are located in the NLB, GLS and SW resistance QTLs that are listed in Table 5.1, 5.2 and 5.3 respectively in the appendix.

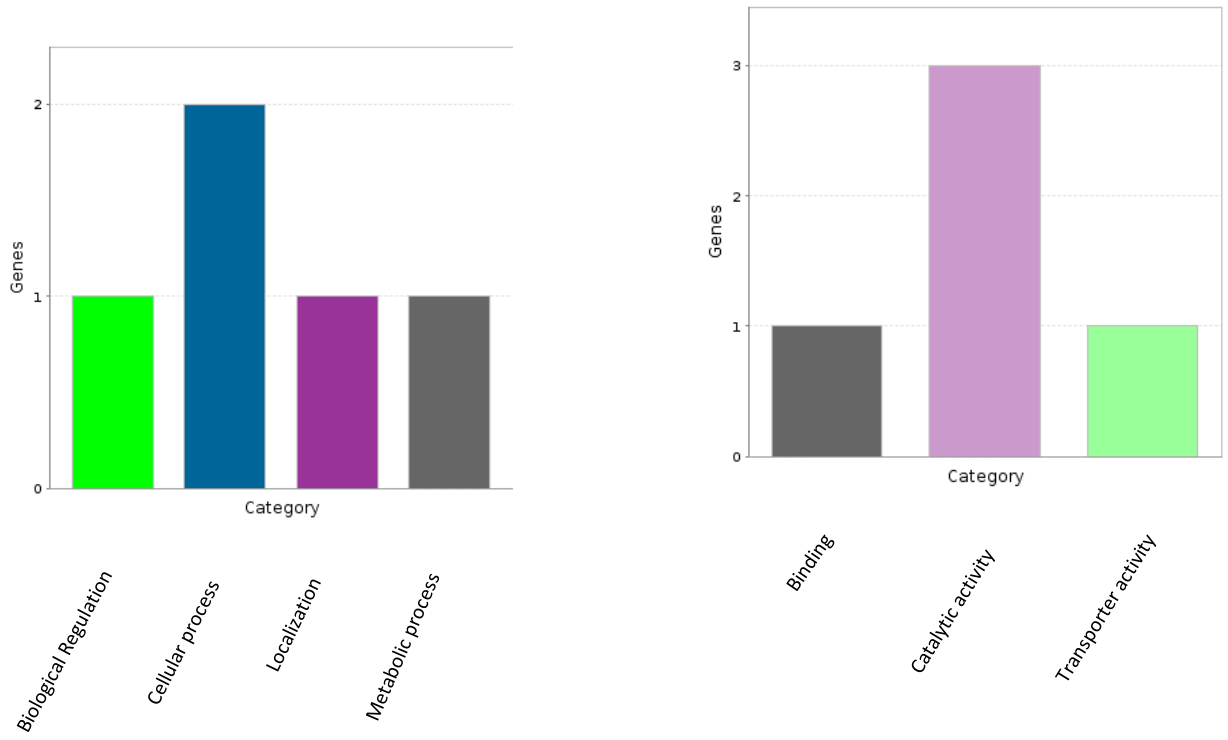

Figure 1.8a and b. *Gene ontology analysis of the genes under the QTLs of the NLB region. Figure a shows the genes involved in biological processes and Figure b shows the genes involved in molecular functions.*

For NLB, 13 genes were categorized based on their role in biological process and molecular function (Figure 1.8). Under biological processes, two out of five genes under the NLB QTLs were categorized to be involved in cellular processes. Under molecular function, most genes were categorized to be involved in catalytic activity. One gene was involved in transporter activity and one gene was involved in binding according to Figure 1.8b.

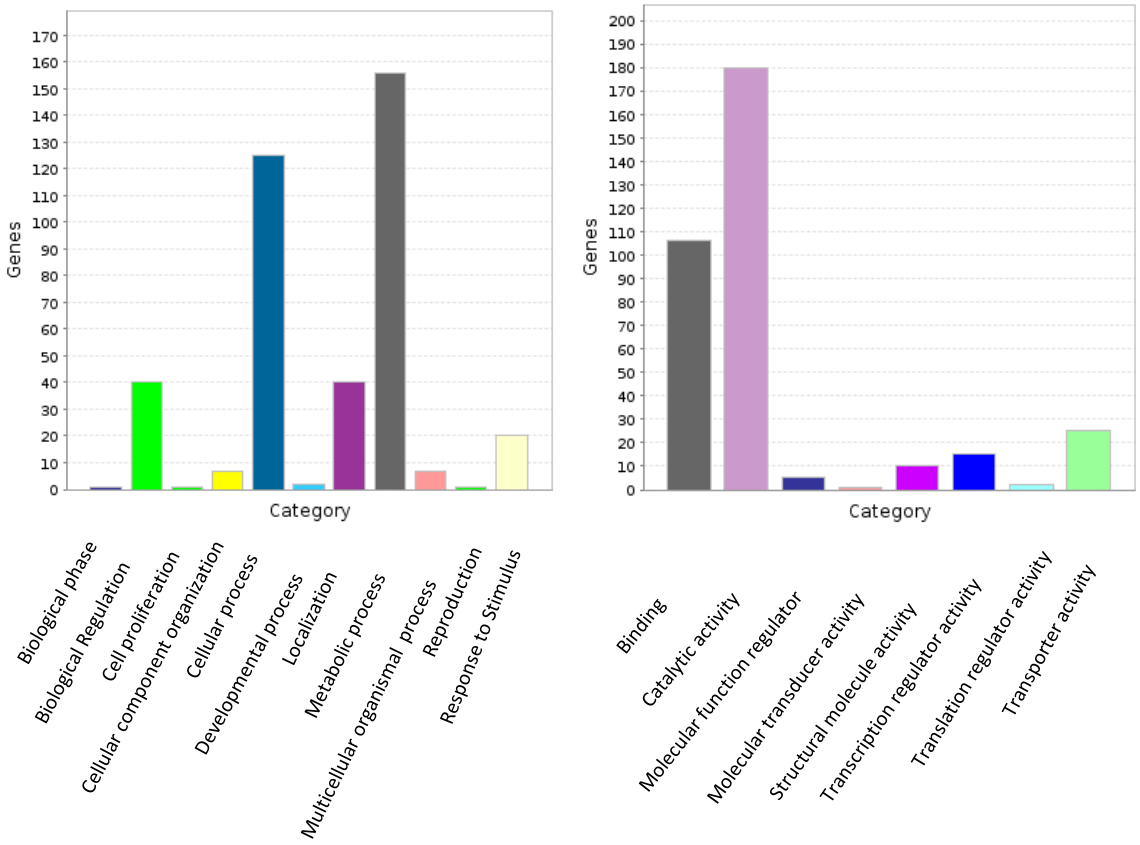

Figure 1.9 a and b *Gene ontology analysis of the genes under the QTLs of the GLS region. Figure a shows the genes involved in biological processes and Figure b shows the genes involved in molecular functions.*

For GLS, 845 genes under the GLS QTL regions were classified based on their activity in biological processes and molecular function (Figure 1.9). Most genes under biological processes were involved in metabolic processes. Under molecular function, most genes were categorized to be involved in catalytic activity. For Stewart's wilt, 1956 genes under the SW QTL regions were classified based on their activity under biological process and molecular function (Figure 1.10). Most genes under biological processes were involved in metabolic processes. Under molecular function, most genes were categorized to be involved in catalytic activity similar to the genes under

GLS QTLs. Most genes for GLS and SW resistance were involved in catalytic activity, metabolic and cellular processes. These genes could be involved in some well-known disease resistance mechanisms such as systemic acquired resistance (SAR) and hypersensitive response.

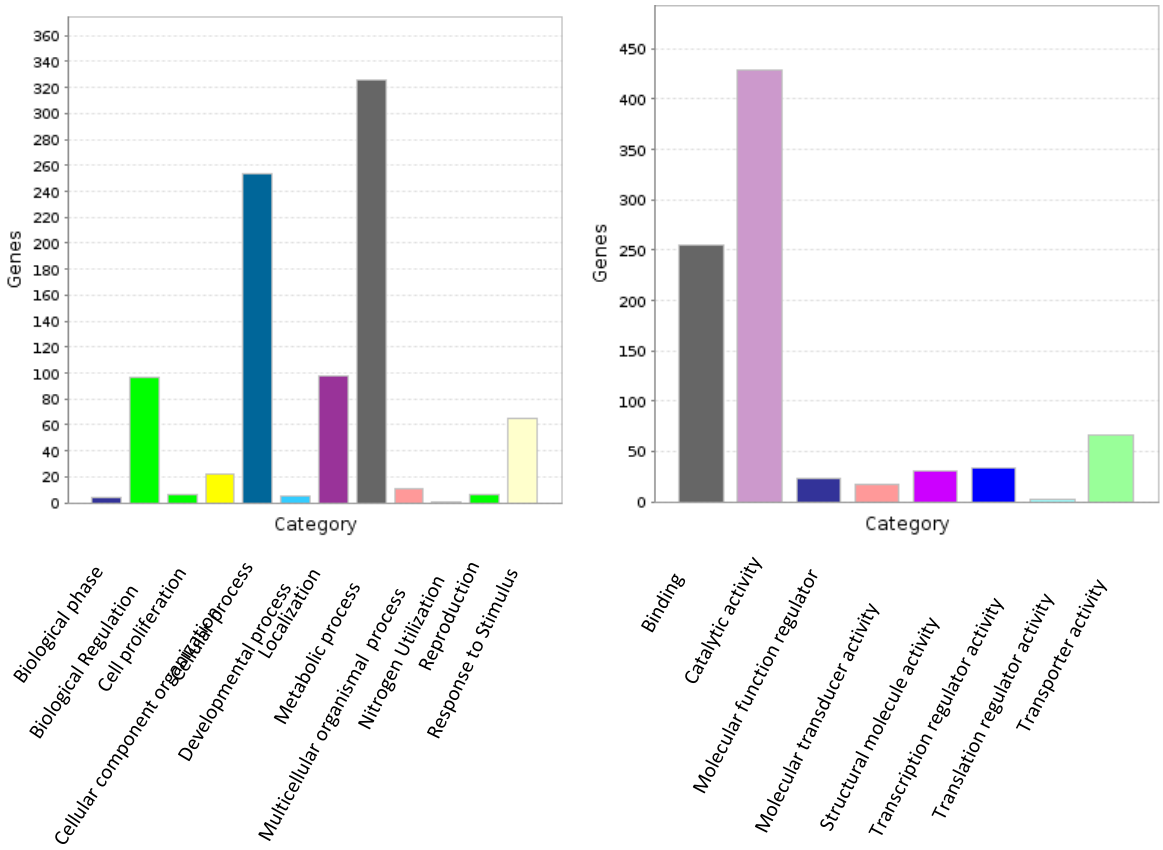

Figure 1.10 a and b *Gene ontology analysis of the genes under the QTLs of the SW region. Figure a shows the genes involved in biological processes and Figure b shows the genes involved in molecular functions.*

The gene ontology analysis of the genes under the QTLs for GLS and SW look similar (See Figure 1.9 and 1.10). The similar functional categorization of genes suggests, that similar pathways are involved in the disease resistance of SW and GLS. NLB resistance seems to have a different pathway than the resistance to GLS and SW.

Genes that contain the NBS-LRR region are widely associated with disease resistance in multiple crop species. In maize inbred B73, these genes were found in all ten chromosomes (Song *et al.* 2015). In our study, no NBS-LRR genes were found in the QTL regions mapped for disease resistance. This might be because the genes involved in disease resistance in NY22613 could have a different mode of action than R gene mediated disease resistance, like salicylic acid-mediated disease resistance. Thus based on our data, it appears that the disease resistance in maize inbred NY22613 is attributed to a number of genes involved in various biological pathways, but not to NBS-LRR genes.

1.5.4 QTL and colocalization analysis

In this QTL analysis study, a total of 49 QTL associated with disease resistance were found in all ten chromosomes of maize. Each of these regions might have several candidate genes that might be associated with disease resistance. The role of these candidate genes in disease resistance can be explored further through fine mapping, expression analysis, mutant analysis and cloning studies (Jamann *et al.* 2014; Hurni *et al.* 2015).

Candidate genes in the QTL that are shared for two or more diseases might be important for multiple disease resistance. In our study, three QTL were shared for NLB and GLS and one QTL was shared for GLS and SW (Table 1.7). In chromosome 7, the colocalized GLS and NLB QTL is located between markers S7_100035505 and S7_100170373. In chromosome 9, the colocalized GLS and NLB QTL is located between markers S9_8778978 and S9_8974937. *LOC541757*,

Zm00001d044974, Zm00001d044975, LOC109941121 are genes whose functions have been previously unidentified.

Table 1. 10 *Candidate genes in the colocalized QTLs.*

Disease	Chr	Region	Gene name
GLS and SW	2	2:4211875..4257983	*Zm00001d002005 (lg1 - liguleless1)*
GLS and NLB	7	7:100035505..100170373	*LOC109941121*
GLS and NLB	9	9:8778978..8974937	*Zm00001d044974*
GLS and NLB	9	9:8778978..8974938	*Zm00001d044975*
GLS and NLB	9	9:8778978..8974939	*LOC541757*
GLS and NLB	10	10:36259906..36380814	*Zm00001d024007*
GLS and NLB	10	10:36259906..36380815	*Zm00001d024008*
GLS and NLB	10	10:36259906..36380816	*LOC103642379*
GLS and NLB	10	10:36259906..36380817	*LOC103641009*

1.5.5 Multiple disease resistance and *Aux/IAA-transcription factor 42* gene

In chromosome 10, the colocalized GLS and NLB QTL is located between markers S10_36259906 and S10_36380814. This region has the *Auxin/Indole-3-Acetic Acid* (*Aux/IAA)-transcription factor 42* gene. Auxin is a phytohormone that is vital for plant growth and development. It plays an important role in apical dominance and differentiation of vascular tissues. *Auxin/Indole-3-*

Acetic Acid (*Aux/IAA)-transcription factor 42* are a class of auxin responsive genes. They are related to the *auxin response factor* (*ARF*) family, *small auxin upregulated RNA* (*SAUR*), and the *auxin-responsive Gretchen Hagen3* (*GH3*) family of genes. The *Aux/IAA transcription factor 42* gene encodes a nuclear protein that is short lived. Aux/IAA family of proteins are known to interact with auxin response factors (ARFs) and regulate genes in multiple ways (Luo *et al.* 2018). They have been known to regulate root development, leaf morphogenesis and fruit ripening (Wang *et al.* 2005). They are also known to be involved in cell wall biosynthesis and might show functional redundancy in some cases (Overvoorde *et al.* 2005) (Ludwig *et al.* 2015).

The maize gene *ZmAuXRP1* is known to condition resistance to gibberella stalk rot (*Gibberella zeae* (Schw.)) and fusarium ear rot (*Fusarium verticillioides* Sheldon) diseases by promoting the biosynthesis of indole-3-acetic acid (IAA). However, it is also known to suppress the formation of benzoxazinoid defense compounds (Ye *et al.* 2018). The *SGT1* gene has been expressed in yeast and is known to be regulated by auxin in multiple plants. *SGT1* is known to be involved in R-gene mediated and non-host resistance to plant pathogens (Wang *et al.* 2010). As a member of this gene family, we hypothesize that the *Aux/IAA transcription factor 42* gene might be involved in the regulation of *SGT1* in maize, which in turn might be involved in the activation of multiple R genes and resistance to NLB and GLS.

1.5.6 Multiple disease resistance and *liguleless1* gene

A colocalized QTL for GLS and SW was observed in chromosome 2 between markers S2_4211875 and S2_4257983. This region includes the *liguleless1* gene that encodes a nuclear-

localized protein that is needed for induction of ligules and auricles during maize leaf organogenesis (Moreno *et al.* 1997). This gene is required for the proper development of the ligules, auricles and the blade-sheath boundary of maize leaves. The LG1 protein encoded by the *liguleless1* gene is mostly novel, however, it contains a domain of 77 amino acids with significant similarity to SQUAMOSA PROMOTER-BINDING (SPL) proteins 1 and 2 identified in *Antirhinum majus* L. (Klein *et al.* 1996). Further functional studies conducted in arabidopsis revealed that SPL proteins are associated with the regulation of leaf initiation, juvenile to adult phase transition, flowering time, shoot branching and inflorescence form (Wang *et al.* 2008; Schwarz *et al.* 2008; Wang *et al.* 2009; Wu and Poethig 2006).

The *liguleless1* gene was also found to be closely associated with *liguleless2*, which encodes a basic leucine rich zipper protein (bZIP), which is a class of transcription factor proteins (Walsh *et al.* 1998). The LG2 protein encoded by *liguleless2* shows sequence homology to the bZIP transcription factor class of proteins. The LG2 protein is very similar to a subclass that has been found in multiple plant species such as wheat (Tabata *et al.* 1991), potato (*Solanum tuberosum* L.) (Feltkamp *et al.* 1994), arabidopsis (Kawata *et al.* 1992; Schindler *et al.* 1992a; Zhang *et al.* 1993; Miao *et al.* 1994; Xiang *et al.* 1995), tobacco (*Nicotiana tabacum* L.) (Katagiri *et al.* 1989) and fava bean (*Vicia faba* L.) (Ehrlich *et al.* 1992). In maize, the bZIP transcription factor class of proteins is hypothesized to be involved in transcriptional regulation (Foley *et al.* 1993). Some of the functions that are hypothesized to be transcriptionally regulated by the bZIP class of proteins are auxin, salicylic acid, and methyl jasmonate responses (Kim *et al.* 1994; Liu and Lam 1994; Qin *et al.* 1994; Zhang and Singh 1994; Xiang *et al.* 1996). They are also hypothesized to be

involved in transcriptional regulation of plant histone (Tabata *et al.* 1991) and *GST* genes (Ulmasov *et al.* 1994; Zhang *et al.* 1995; Xiang *et al.* 1996).

Disease resistance and the bZIP class of proteins are related through the involvement of the bZIP class of proteins in salicylic acid and methyl jasmonate regulation. Salicylic acid (SA) plays a major role in plant defense against plant pathogens through systemic acquired resistance (SAR) (Ross 1961). Salicylic acid or acetyl salicylic acid (aspirin) treatment of tobacco leaves demonstrated heightened resistance to tobacco mosaic virus and increased pathogenesis-related (PR) protein accumulation (White 1979; Antoniw and White 1980). Since then, a number of studies conducted in multiple plant species have shown that SA is a major component of plant defense to plant pathogens. SA plays a major role in multiple plant defense responses such as containment of pathogen spread, activation of cell death and the induction of local and systemic disease resistance. SA does this by altering expression or activity of various enzymes, potentiating the osmotic balance, cell death, expression of various defense genes or potentially even through generation of free radicals (Durner *et al.* 1997).

In maize, SA treatment showed the activation of expression of *PR-1* and *PR-5* genes which resulted in resistance to downy mildew *Peronosclerospora maydis* (Racib.) C.G. Shaw (Morris *et al.* 1998). SA treatment was shown to induce the expression of pathogenesis related (*PR)* genes in rice and barley (*Hordeum vulgare* L.) (Matsuta *et al.*, 1991; Vallélian-Bindschedler *et al.*, 1998). In maize, transcription factors are known to negatively regulate cell death and result in hypersensitive response (HR). Lesions simulating disease resistance (*lsd)* phenotype is associated with a recessive gene named *lsd1* and it codes for a zinc finger protein that is shown to negatively

regulate cell death in maize (Dietrich *et al.* 1997) . *Lls1* (lethal leaf spot 1) is another transcription factor in maize that was shown to negatively regulate cell death (Gray *et al.*, 1997).

The jasmonic acid (JA) pathway is usually associated with defense response of plants to feeding by herbivores by regulating the production of secondary metabolites. JA has been shown to induce plant volatiles that can repel herbivores and attract their natural enemies, deter feeding and inhibit digestion of plants by herbivores (Chen *et al.* 2005; De Moraes *et al.* 2001; Turlings *et al.* 1990). The JA pathway has been associated with plant pathogen disease responses. For example, in rice, JA is associated with the expression of PR-1 and determines the level of resistance against the fungus *Magnaporthe grisea* (Hobert) Barr (Schweizer *et al.* 1997). Methyl jasmonates have been shown to elicit various phenylpropanoid defense responses in parsley (*Petroselinum crispum* (Mill.) Nyman ex A.W. Hill) cell cultures. This primes the cell cultures for enhanced induction of the early oxidative burst and various phenylpropanoid defense responses (Krauss *et al.* 1994; Krauss *et al.* 1992).

Liguleless1 is shown to be cell autonomous. However, *liguleless2* is also associated with cell to cell signaling (Harper and Freeling 1996). Thus activation of *liguleless1* can activate *liguleless2,* which in turn can activate downstream signaling of the SA pathway. The SA pathway can transcriptionally activate methyl jasmonates and can prime the plants and protect them from further pathogen infections, thus resulting in multiple disease resistance.

Based on our findings, we hypothesize that *liguleless1* and *liguleless2* act together and result in the activation of the SA defense pathway, thus resisting initial pathogen attack. The SA pathway

activates the jasmonic acid pathway and it primes the plants and prevents other pathogens from achieving successful pathogenesis. *Aux/IAA transcription factor 42* also might be involved in the R-gene mediated resistance. These might be the molecular mechanisms that are responsible for some of the broad resistance observed in the maize inbred NY22613.

1.5.7 Future directions

The QTL analysis, colocalization analysis and the gene ontology analysis conducted for three different maize leaf diseases in our study suggest that there is a strong genetic component that might involve multiple molecular mechanisms that result in the multiple disease resistance observed in the maize inbred NY22613. An expression analysis will give us more insights into the genes involved in disease resistance in maize inbred NY22613. *Auxin/Indole-3-Acetic Acid* (*Aux/IAA)-transcription factor 42* and *liguleless1* are interesting candidate genes that might be involved in multiple disease resistance. A loss of function mutation in these genes through genome editing methods such as CRISPR-cas9 (Clustered Regularly Interspaced Short Palindromic Repeats - Cas9), could shed light on the function of these genes and their role in disease resistance. By understanding the molecular pathways involved in disease resistance, breeders will be able to develop crop varieties that have more durable resistance to plant pathogens.

References

Antoniw, J. F., & White, R. F. (1980). The effects of aspirin and polyacrylic acid on soluble leaf proteins and resistance to virus infection in five cultivars of tobacco. Journal of Phytopathology, 98(4), 331-341.

Austin MJ, Muskett P, Kahn K, Feys BJ, Jones JD, Parker JE. Regulatory role of *SGT1* in early R gene-mediated plant defenses. Science 2002; 295:2077–2080.

Azevedo C, Sadanandom A, Kitagawa K, Freialdenhoven A, Shirasu K, Schulze-Lefert P. The *RAR1 interactor SGT1*, an essential component of R gene-triggered disease resistance. Science 2002; 295:2073–2076.

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. 67:1–48. doi:10.18637/jss.v067.i01. [accessed 2019 Jan 13]. http://www.jstatsoft.org/v67/i01/.

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. 11:1–23. doi:10.1371/journal.pgen.1005045.

Bernardo RN. 2010. Breeding for quantitative traits in plants. Stemma Press.

Chung CL, Jamann T, Longfellow J, 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. doi:10.1007/s00122-010-1303-z.

Chung CL, Jamann T, Longfellow J, 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. doi:10.1007/s00122-010-1303-z.

Conrath, U., Pieterse, C. M., & Mauch-Mani, B. (2002). Priming in plant–pathogen interactions. Trends in plant science, 7(5), 210-216.

Chen, H., Wilkerson, C. G., Kuchar, J. A., Phinney, B. S., & Howe, G. A. (2005). Jasmonate-inducible plant enzymes degrade essential amino acids in the herbivore midgut. Proceedings of the National Academy of Sciences, 102(52), 19237-19242.

DeChant C, Wiesner-Hanks T, Chen S, Stewart EL, Yosinski J, Gore MA, Nelson RJ, Lipson H. 2017. Automated Identification of Northern Leaf Blight-Infected Maize Plants from Field Imagery Using Deep Learning. Phytopathology 107:1426–1432. doi:10.1094/PHYTO-11-16-0417-R. [accessed 2019 Jan 14]. https://apsjournals.apsnet.org/doi/10.1094/PHYTO-11-16-0417-R.

De Moraes, C. M., Mescher, M. C., & Tumlinson, J. H. (2001). Caterpillar-induced nocturnal plant volatiles repel conspecific females. Nature, 410(6828), 577.

Dietrich, R. A., Richberg, M. H., Schmidt, R., Dean, C., and Dangl, J. L. 1997. A novel zinc finger protein is encoded by the Arabidopsis *LSD1* gene and functions as a negative regulator of plant cell death. Cell 88: 685– 694.

Durner, J., Shah, J., & Klessig, D. F. (1997). Salicylic acid and disease resistance in plants. Trends in plant science, 2(7), 266-274.

Ehrlich, M., Ehrlich, K. C., & Cary, J. W. (1992). A broad bean cDNA clone encoding a DNA-binding protein resembling mammalian CREB in its sequence specificity and DNA methylation sensitivity. Gene, 117(2), 169-178. doi:10.1016/0378-1119(92)90726-6

Feltkamp, D., Masterson, R., Starke, J., & Rosahl, S. (1994). Analysis of the involvement of *ocs*-like bZip-binding elements in the differential strength of the bidirectional mas1'2' promoter. Plant Physiology, 105(1), 259-268. doi:10.1104/pp.105.1.259

Foley, R. C., Grossman, C., Ellis, J. G., Llewellyn, D. J., Dennis, E. S., Peacock, W. J., & Singh, K. B. (1993). Isolation of a maize bZIP protein subfamily: Candidates for the *ocs*-element transcription factor. The Plant Journal, 3(5), 669-679. doi:10.1111/j.1365-313X.1993.00669.x

Gebhardt C, Valkonen JPT. 2001. Organization of genes controlling disease resistance in the potato genome. Annu. Rev. Phytopathol. 39:79–102. doi:10.1146/annurev.phyto.39.1.79. [accessed 2019 Jan 13]. http://www.ncbi.nlm.nih.gov/pubmed/11701860.

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. Tinker NA, editor. PLoS One 9:e90346. doi:10.1371/journal.pone.0090346. [accessed 2019 Jan 13]. https://dx.plos.org/10.1371/journal.pone.0090346.

Gray, J., Close, P. S., Briggs, S. P., and Johal, G. S. 1997. A novel suppressor of cell death in plants encoded by the Lls1 gene of maize. Cell 89: 25–31.

Harper, L., & Freeling, M. (1996). Interactions of *liguleless1* and *liguleless2* function during ligule induction in maize. Genetics, 144(4), 1871-1882.

Jamann TM, Luo X, Morales L, Kolkman JM, Chung CL, Nelson RJ. 2016. A remorin gene is implicated in quantitative disease resistance in maize. Theor. Appl. Genet. 129:591–602. doi:10.1007/s00122-015-2650-6.

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. doi:10.1534/genetics.114.167486. [accessed 2018 Nov 30]. http://www.ncbi.nlm.nih.gov/pubmed/25009146.

Kawata, T., Imada, T., Shiraishi, H., Okada, K., Shimura, Y., & Iwabuchi, M. (1992). a cDNA clone encoding *hbp-1b* homolog in *Arabidopsis thaliana*. Nucleic Acids Research, 20(5), 1141-1141.

Kauss, H. *et al.* (1994) Pretreatment of parsley *(Petroselinum crispum L.)* suspension cultures with methyl jasmonate enhances elicitation of activated oxygen species. Plant Physiol. 105, 89–94

Kauss, H. *et al.* (1992) Methyl jasmonate conditions parsley suspension cells for increased elicitation of phenylpropanoid defense responses. Biochem. Biophys. Res. Commun. 189, 304–308

Kim, Y., Buckley, K., Costa, M. A., & An, G. (1994). A 20 nucleotide upstream element is essential for the nopaline synthase (nos) promoter activity. Plant molecular biology, 24(1), 105-117.

Lam, E., Katagiri, F., & Chua, N. (1989). Two tobacco DNA-binding proteins with homology to the nuclear factor CREB. Nature, 340(6236), 727-730. doi:10.1038/340727a0

Li Y, Chen L, Li C, Bradbury PJ, Shi Y, Song Y, Zhang D, Zhang Z, Buckler ES, Li Y, *et al.* 2018. Increased experimental conditions and marker densities identified more genetic loci associated with southern and northern leaf blight resistance in maize. Sci. Rep. 8:6848. doi:10.1038/s41598-018-25304-z. [accessed 2018 Dec 31]. http://www.nature.com/articles/s41598-018-25304-z.

Liu, X., & Lam, E. (1994). Two binding sites for the plant transcription factor *ASF-1* can respond to auxin treatments in transgenic tobacco. Journal of Biological Chemistry, 269(1), 668-675.

Ludwig, Y., Zhang, Y., & Hochholdinger, F. (2013). The maize (Zea mays L.) AUXIN/INDOLE-3-ACETIC ACID gene family: phylogeny, synteny, and unique root-type and tissue-specific expression patterns during development. PloS one, 8(11), e78859.

Luo, J., Zhou, J. J., & Zhang, J. Z. (2018). Aux/IAA Gene Family in Plants: Molecular Structure, Regulation, and Function. International journal of molecular sciences, 19(1), 259. doi:10.3390/ijms19010259

MacQueen J. 1967. Some Methods for Classification and Analysis of Multivariate Observations. [accessed 2019 Jan 13]. https://www.semanticscholar.org/paper/Some-Methods-for-Classification-and-Analysis-of-MacQueen/a718b85520bea702533ca9a5954c33576fd162b0.

Manosalva PM, Davidson RM, Liu B, Zhu X, Hulbert SH, Leung H, Leach JE. 2009. A Germin-Like Protein Gene Family Functions as a Complex Quantitative Trait Locus Conferring Broad-Spectrum Disease Resistance in Rice. PLANT Physiol. 149:286–296. doi:10.1104/pp.108.128348. [accessed 2018 Dec 19]. http://www.ncbi.nlm.nih.gov/pubmed/19011003.

Matsuta, C., Van den Bulcke, M., Bauw, G., Van Montagu, M., and Caplan, A. G. 1991. Differential effects of elicitors on the viability of rice suspension cells. Plant Physiol. 97: 619–629.

Miao Zhong He, Zhong He, Miao, Z., Liu, X., Liu Xiangjun, X., Lam, E., & Lam, E. (1994). TGA3 is a distinct member of the TGA family of bZIP transcription factors in *Arabidopsis thaliana*. Plant Molecular Biology, 25(1), 1-11. doi:10.1007/BF00024193

Morris, S. W., Vernooij, B., Titatarn, S., Starrett, M., Tho- mas, S., Wiltse, C. C., Frederiksen, R. A., Bhandhufalck, A., Hulbert, S., and Uknes, S. 1998. Induced resistance responses in maize. Mol. Plant-Microbe Interact. 11: 642–658.

Mohanty SP, Hughes DP, Salathé M. 2016. Using Deep Learning for Image-Based Plant Disease Detection. Front. Plant Sci. 7:1419. doi:10.3389/fpls.2016.01419. [accessed 2019 Jan 14]. http://journal.frontiersin.org/article/10.3389/fpls.2016.01419/full.

Oerke E-C. 2006. Crop losses to pests. J. Agric. Sci. 144:31. doi:10.1017/S0021859605005708.

Overvoorde, P. J., Okushima, Y., Alonso, J. M., Chan, A., Chang, C., Ecker, J. R., ... & Smith, A. (2005). Functional genomic analysis of the *AUXIN/INDOLE-3-ACETIC ACID* gene family members in *Arabidopsis thaliana*. The Plant Cell, 17(12), 3282-3300.

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

Poland JA, Bradbury PJ, Buckler ES, Nelson RJ. 2011. Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proc. Natl. Acad. Sci. 108:6893–6898. doi:10.1073/pnas.1010894108.

Qin, X. F., Holuigue, L., Horvath, D. M., & Chua, N. H. (1994). Immediate early transcription activation by salicylic acid via the cauliflower mosaic virus as-1 element. The Plant Cell, 6(6), 863-874

Ross, A. F. 1961. Systemic acquired resistance induced by localized virus infections in plants. Virology 14: 340–- 358.

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Saltz JB, Hessel FC, Kelly MW. 2017. Trait Correlations in the Genomics Era. Trends Ecol. Evol. 32:279–290. doi:10.1016/j.tree.2016.12.008. [accessed 2019 Jan 13]. http://www.ncbi.nlm.nih.gov/pubmed/28139251.

Schindler, U., Beckmann, H., & Cashmore, A. R. (1992). *TGA1* and G-box binding factors: Two distinct classes of Arabidopsis leucine zipper proteins compete for the G-box-like element TGACGTGG. The Plant Cell, 4(10), 1309-1319. doi:10.1105/tpc.4.10.1309

Schwarz, S., Grande, A. V., Bujdoso, N., Saedler, H., & Huijser, P. (2008). The microRNA regulated SBP-box genes *SPL9* and *SPL15* control shoot maturation in Arabidopsis. Plant Molecular Biology, 67(1), 183-195. doi:10.1007/s11103-008-9310-z

Schweizer, P. *et al.* (1997) Gene expression patterns and levels of jasmonic acid in rice treated with the resistance inducer 2,6-dichloroisonicotinic acid. Plant Physiol. 115, 61–70

Singh, R. P., Hodson, D. P., Huerta-Espino, J., Jin, Y., Bhavani, S., Njau, P., ... & Govindan, V. (2011). The emergence of Ug99 races of the stem rust fungus is a threat to world wheat production. *Annual review of phytopathology*, *49*, 465-481.

Sucher J, Boni R, Yang P, Rogowsky P, Büchner H, Kastner C, Kumlehn J, Krattinger SG, Keller B. 2017. The durable wheat disease resistance gene *Lr34* confers common rust and northern corn leaf blight resistance in maize. Plant Biotechnol. J. doi:10.1111/pbi.12647.

Tabata, T., Nakayama, T., Mikami, K., & Iwabuchi, M. (1991). *HBP-1a* and *HBP-1b*: Leucine zipper-type transcription factors of wheat. The EMBO Journal, 10(6), 1459-1467. doi:10.1002/j.1460-2075.1991.tb07666.x

Turlings, T. C., Tumlinson, J. H., & Lewis, W. J. (1990). Exploitation of herbivore-induced plant odors by host-seeking parasitic wasps. Science, 250(4985), 1251-1253.

Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. [accessed 2019 Jan 13]. https://statweb.stanford.edu/~gwalther/gap.

Ulmasov, T., Hagen, G., & Guilfoyle, T. (1994). The *ocs* element in the soybean *GH2/4* promoter is activated by both active and inactive auxin and salicylic acid analogues. Plant molecular biology, 26(4), 1055-1064.

Vallélian-Bindschedler, L., Métraux, J-P., and Schweizer, P. 1998. Salicylic acid accumulation in barley is pathogen specific but not required for defense-gene activation. Mol. Plant-Microbe Interact. 11: 702–705.

Walsh, J., Waters, C. A., & Freeling, M. (1998). The maize gene *liguleless2* encodes a basic leucine zipper protein involved in the establishment of the leaf blade-sheath boundary. Genes & development, 12(2), 208–218.

Wang, K., Uppalapati, S. R., Zhu, X., Dinesh-Kumar, S. P., & Mysore, K. S. (2010). *SGT1* positively regulates the process of plant cell death during both compatible and incompatible plant–pathogen interactions. Molecular plant pathology, 11(5), 597-611.

Wang GL, Mackill DJ, Bonman JM, McCouch SR, Champoux MC, Nelson RJ. 1994. RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. Genetics 136:1421–34. [accessed 2019 Jan 13]. http://www.ncbi.nlm.nih.gov/pubmed/7912216.

Wang, H., Jones, B., Li, Z., Frasse, P., Delalande, C., Regad, F., ... & Bouzayen, M. (2005). The tomato *Aux/IAA transcription factor IAA9* is involved in fruit development and leaf morphogenesis. The Plant Cell, 17(10), 2676-2692.

Wang, J., Czech, B., & Weigel, D. (2009). miR156-regulated *SPL* transcription factors define an endogenous flowering pathway in arabidopsis thaliana. Cell, 138(4), 738-749. doi:10.1016/j.cell.2009.06.014

Wang, J., Schwab, R., Czech, B., Mica, E., & Weigel, D. (2008). Dual effects of miR156-targeted *SPL* genes and CYP78A5/KLUH on plastochron length and organ size in *Arabidopsis thaliana*. The Plant Cell, 20(5), 1231-1243. doi:10.1105/tpc.108.058180

White, R. F. 1979. Acetylsalicylic acid (aspirin) induces resistance to tobacco mosaic virus in tobacco. Virology 99: 410–412.

Wiesner-Hanks T, Nelson R. 2016. Multiple Disease Resistance in Plants. Annu. Rev. Phytopathol. 54:229–252. doi:10.1146/annurev-phyto-080615-100037. [accessed 2018 Sep 20]. http://www.annualreviews.org/doi/10.1146/annurev-phyto-080615-100037.

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. Natl. Acad. Sci. 108:7339–7344. doi:10.1073/pnas.1011739108.

Wu, G., & Poethig, R. S. (2006). Temporal regulation of shoot development in *Arabidopsis thaliana* by *miR156* and its target *SPL3*. Development, 133(18), 3539-3547. doi:10.1242/dev.02521

Xiao W, Zhao J, Fan S, Li L, Dai J, Xu M. 2007. Mapping of genome-wide resistance gene analogs (RGAs) in maize (Zea mays L.). Theor. Appl. Genet. 115:501–508. doi:10.1007/s00122-007-0583-4. [accessed 2019 Jan 13]. http://www.ncbi.nlm.nih.gov/pubmed/17581735.

Xiang, C., Miao, Z., & Lam, E. (1995). Isolation of TGA6, a new member of the TGA family of bZIP transcription factors in Arabidopsis thaliana. Plant Physiol, 109, 721.

Xiang, C., Miao, Z. H., & Lam, E. (1996). Coordinated activation of *as-1*-type elements and a tobacco glutathione S-transferase gene by auxins, salicylic acid, methyl-jasmonate and hydrogen peroxide. Plant molecular biology, 32(3), 415-426.

Yang, J. (2007). Phenotypic and genotypic investigation of multiple disease resistance in Maize.

Yang Q, Balint-Kurti P, Xu M. 2017. Quantitative Disease Resistance: Dissection and Adoption in Maize. doi:10.1016/j.molp.2017.02.004. [accessed 2018 Nov 7]. http://dx.doi.org/10.1016/j.molp.2017.02.004.

Ye, J., Zhong, T., Zhang, D., Ma, C., Wang, L., Yao, L., ... & Xu, M. (2018). The Auxin-Regulated Protein ZmAuxRP1 Coordinates the Balance between Root Growth and Stalk Rot Disease Resistance in Maize. Molecular Plant.

Zhang, B., Foley, R. C., & Singh, K. B. (1993). Isolation and characterization of two related Arabidopsis *ocs*-element bZIP binding proteins. Plant Journal, 4(4), 711-716. doi:10.1046/j.1365-313X.1993.04040711.x

Zhang, B., & Singh, K. B. (1994). *ocs* element promoter sequences are activated by auxin and salicylic acid in Arabidopsis. Proceedings of the National Academy of Sciences, 91(7), 2507-2511.

Zhang, B., Chen, W., Foley, R. C., Büttner, M., & Singh, K. B. (1995). Interactions between distinct types of DNA binding proteins enhance binding to *ocs* element promoter sequences. The Plant Cell, 7(12), 2241-2252.

CHAPTER 2

EXPRESSION ANALYSIS OF NORTHERN LEAF BLIGHT RESISTANT MAIZE INBRED NY22613 AND SUSCEPTIBLE MAIZE INBRED Oh7B.

2.1 Background

Northern leaf blight (NLB) is a maize (*Zea mays* L. ssp. *mays*) disease caused by an ascomycete named *Setosphaeria turcica* (Luttrell) Leonard and Suggs (anamorph *Exserohilum turcicum*). The pathogen was previously called *Helminthosporium turcicum*. NLB is known to cause economic loss all over the world and has the potential to cause yield losses from 18% to 75% in controlled yield trials (Perkins and Pedersen 1987; Raymundo and Hooker 1981; Shankara and Gowda 2011; Sibiya *et al.* 2012). This disease has resulted in an estimated yield loss of 1.3 billion bushels in 2012 in the US and Canada. The yield loss increased to 2.1 billion bushels in 2015 (Mueller *et al.*2016). NLB also causes indirect economic loss to the dairy industry since it is also known to affect silage quality and digestibility (Wang *et al.* 2010).

In line with the classic evolutionary theory called the red queen hypothesis, proposed by Leigh Van Valen in 1973, *Setosphaeria turcica* has evolved with maize over the years (Van Valen, 1973; Ferguson and Carson 2007). It is important to note that the pathogen has other hosts than maize. For example, *S. turcica* is known to infect sorghum (*Sorghum bicolor* (L.) Moench ssp. *bicolor),* Johnsongrass (*Sorghum halepen*se (L.) Pers.) and Brassica species. However, the pathogen seems to prefer maize (Martin 2011). *S. turcica* has caused epidemics in a number of places in the United

States since the early 1940s (Jenkins *et al.* 1954) and subsequently in northwestern North Carolina (Leonard *et al.* 1986), in the Gulf Coast of Texas in 1985 (Smith *et al.* 1988), throughout Texas in 1992 (Krausz *et al.* 1993), and frequently on sweet maize in Florida (Pataky, 1991). In New York, the pathogen has been a chronic problem to maize (Keller *et al.* 1990).

Breeders have used both qualitative and quantitative resistance to breed NLB resistant cultivars. In maize, qualitative resistance offered by the *Ht* genes has been used by breeders to develop maize hybrids that are resistant to NLB (Ferguson and Carson 2007). Breeders prefer to use *Ht* genes to tackle NLB due to the virulence of the *Ht* genes towards the pathogen. Based on virulence or avirulence to the *Ht* resistance genes, *Setosphaeria turcica* has been classified into multiple races (see Table 2.1).

S. turcica populations have been studied in tropical regions like Kenya, Mexico, India and southern China and temperate regions like Europe, northern China, and the eastern United States (Borchardt, 1998a; Borchardt, 1998b; Borchardt, 1998c; Ferguson and Carson, 2004). Populations in the tropical regions exhibited higher genotypic diversity, suggesting that *S. turcica* undergoes sexual recombination under those conditions (Borchardt 1998a). Fluctuating temperatures in the northeastern United States create the potential for increased sexual recombination. The potential generation of new pathogen races and the increasing occurrence of NLB in the northeastern United States justify the need to better understand the genetic basis of resistance to NLB.

Table 2.1 *NLB pathogen races and their reaction to Ht genes*. (Svec and Dolezal 2019; Ferguson and Carson 2007;Bergquist 1974; Gianasi,1996; Jordan et al. 1993; Keller *et al*. 1990)

Pathogen Race	Host reaction to each race			
	***Ht1* gene**	***Ht2* gene**	***Ht3* gene**	***HtN* gene**
0	R	R	R	R
1	S	R	R	R
2	R	S	R	R
12	S	S	R	R
23	R	S	S	R
23N	R	S	S	S
123N	S	S	S	S

As discussed in Chapter 1 of this dissertation, a QTL mapping study was conducted with a BC_3S_3 mapping population generated with the resistant maize inbred NY22613 and the susceptible inbred Oh7B. The QTLs for NLB resistance detected in that study are listed in Table 2.2. Table 2.3 lists all the genes that have been currently identified in the QTL regions listed in Table 2.2.

Table 2.2 *QTL regions identified for northern leaf blight with the BC_3S_3 Oh7B X NY22613 mapping population.*

Disease	Locus	Physical map	Chromosome Location
NLB	1	37720906	1:37717388..37725221

NLB	1	19979634	1:19978997..19979751
NLB	2	2484151	2:2476120..2492538
NLB	2	49057160	2:49016939.. 49057266
NLB	2	55560345	2:55560318..55560440
NLB	3	223740910	3:223740841..223745642
NLB	3	171900654	3:171900438..171900685
NLB	4	185566213	4:185529728..185649204
NLB	4	79449676	4:79162068..79451618
NLB	5	215567451	5:215567100..215620081
NLB	6	148650827	6:148633043..148654966
NLB	6	98010324	6:98007087..98010327
NLB	7	100081042	7:100035505..100170373
NLB	8	172365897	8:172365300..172365909
NLB	9	8974761	9:8778978..8974937
NLB	10	36380036	10:36259906..36380814

A differential expression study was conducted with the susceptible inbred Oh7B and the resistant inbred NY22613 under both field and greenhouse conditions. Plants have shown varying resistance to pathogens in greenhouse and field conditions in wheat (*Triticum aestivum* L.) , soybean (*Glycine max* (L.) Merr.), tobacco (*Nicotiana tabacum* L.) and potato (*Solanum tuberosum* L.) (Boshoff 2000; Pretorius 2007; Twizeyimana *et al.* 2007; Daub 1989; Liljeroth, 2010). This variation could be due to differing environmental conditions. We aimed to identify genes that are differentially expressed between the susceptible and resistant inbreds in greenhouse and field conditions in

response to NLB inoculation. Differential expression analysis of genes between the infected and uninfected plants will give us insights into the pathways involved in disease resistance to NLB.

Table 2.3 *Genes identified in the QTL regions for NLB*

Gene stable ID	Chr	Gene description	Gene start (bp)	Gene end (bp)
Zm00001d001887	2	Adenylate kinase 1 chloroplastic	2472648	2482331
Zm00001d001888	2		2482827	2482899
Zm00001d001889	2		2483126	2483198
Zm00001d001890	2		2483438	2483510
Zm00001d001891	2		2483735	2483807
Zm00001d001892	2		2484047	2484119
ENSRNA049444234	2	tRNA-Ala for anticodon CGC	2484047	2484119
Zm00001d001893	2		2487785	2487857
Zm00001d001894	2	hydroxyproline-rich glycoprotein family protein	2491277	2493876
Zm00001d003584	2	12-oxo-phytodienoic acid reductase5	49052848	49056045
Zm00001d052276	4	RNA-processing Lsm domain	185636179	185645200
Zm00001d050294	4	VQ motif-containing protein	79210012	79211052
Zm00001d026840	4		79277218	79277703
Zm00001d050295	4		79407612	79412327

		U-box domain-containing protein		
Zm00001d050296	4	37	79412293	79413137
Zm00001d018153	5		215596858	215600746
		Protein CHROMATIN		
Zm00001d018151	5	REMODELING 25	215569988	215588200
Zm00001d018154	5	Beta-adaptin-like protein C	215609881	215612833
		Protein SUPPRESSOR OF GENE		
Zm00001d018152	5	SILENCING 3	215594980	215596255
		Cell division control protein 6		
Zm00001d012320	8	homolog B	172365657	172369913
Zm00001d024008	10	Auxin-responsive protein IAA27	36379483	36384610

2.2 Materials and methods

Susceptible maize inbred Oh7B and resistant maize inbred NY22613 were planted at Robert B. Musgrave research farm in Poplar Ridge, NY in summer 2018. Each inbred was planted in a 20-plant row with one row inoculated and one non-inoculated, and three replications for each inbred-inoculation treatment combination. Plants in the inoculated treatment rows were inoculated with *S. turcica* race 1 at the mid-whorl stage with 4000 spores/ml liquid inoculum. Leaf samples were collected from maize plants at V6 to V8 stage immediately before inoculation (day - 0) and after inoculation on day - 7, day - 10 and day - 14. Leaf samples were collected with gloves to avoid RNAase degrading the quality of RNA samples. Three plants from each 20-plant row were randomly chosen and leaf samples were collected from leaf 4 to leaf 7 in the leaf blade region.

This is important since differential expression has been observed in developmental genes in distinct leaf zones in maize (Li *et al.* 2010). Leaf samples from each row were pooled. After each sample collection, the samples were frozen in liquid nitrogen.

In the greenhouse, seeds were planted in thermoformed trays. When the plants were 15.24 cm to 25.40 cm tall they were transplanted into 17.78 X 17.15 cm plant pots. Plants were grown with a proprietary mix of growth media called Cornell Mix^{TM}. Plants were monitored and watered two times a day based on need. They were grown under 12 hour photoperiod with high pressure sodium lamps. Plants were inoculated with *S. turcica* race 1 at the mid-whorl stage with 4000 spores/ml liquid inoculum. Samples were collected from maize plants at V6 to V8 stage right before inoculation (day - 0) and after inoculation on day - 7, day - 10 and day – 14 as done in the field. Samples were handled similarly to the samples collected in the field and frozen in liquid nitrogen.

Table 2.4 *Dates of sample collection for mRNA extraction from field and greenhouse grown plants.*

	Sample Collection Date	**mRNA Extraction**
Greenhouse	September 5 2018	September 5 2018
	September 12 2018	September 12 2018
	September 15 2018	September 19 2018
	September 19 2018	September 20 2018
Field	July 9 2018	July 10 2018

July 16 2018	July 17 2018
July 19 2018	July 20 2018
July 23 2018	July 23 2018

The frozen samples were stored in a - 80°C freezer until mRNA library preparation. Library preparation was conducted using the TruSeq™ protocol described in the appendix. Sequencing libraries were obtained for all the samples using Next Seq 500 sequencing platform from Illumina™ in the Biotechnology Resource Center at Cornell University. Samples from the field and greenhouse were run separately in two lanes in Next Seq 500™ for sequencing. The sequence data was then transferred to a remote high performance computer at the Biology High Performance Computing facility at Cornell University. The dates of the samples collected are listed in table 2.4. The number and time of samples collected are listed in field and greenhouses are listed in Table 5.4 in the appendix.

Fastqc™ software was used to generate quality control of the sequence reads. Spliced Transcripts Alignment to a Reference (STAR) software was used to align RNA-seq reads to the B73 maize genome (version 4.0) (Dobin *et al.* 2012). Th eedge R package was used in the R-studio statistical software for performing differential expression analysis. The qqman package was used to generate Manhattan plots in R-studio software. Protein analysis through evolutionary relationships (PANTHER) platform was used to perform the gene ontology analysis (Thomas *et al.* 2003; Mi *et al.* 2013).

2.3 Results

Analyzing RNA Seq data can be done in multiple platforms such as DESeq (Andres and Huber 2010) and TopHat (Trapnell et al. 2013). We used Spliced Transcripts Alignment to a Reference (STAR) software to align RNA-seq reads to the B73 maize genome (version 4.0) (Dobin *et al.* 2012) and the edge R software in R Studio to conduct the differential expression study since this approach is considered to be more efficient in analyzing RNA Seq data than the aforementioned platforms (Sun, Personal communication, 2019; Robinson *et al.* 2010; McCarthy *et al.* 2012). The differential expression analysis in edgeR uses a range of statistical methodology based on the negative binomial distributions. This includes empirical Bayes estimation, exact tests, generalized linear models and quasi- likelihood tests. The p-values were obtained for differential expression analysis using exact test and the alpha value was set to 0.001. The counts were normalized before differential expression analysis. The negative binomial distribution test was performed to obtain p-values in the RNA Seq experiments.

The scree plots show that a significant number of genes were upregulated and downregulated in Oh7B vs. NY22613 both in greenhouse and field conditions. In the greenhouse, the differential expression study between Oh7B and NY22613, revealed that 7082 genes were downregulated and 6842 genes were upregulated. For 8477 genes, expression differences were not significant. In the field, 5134 genes were downregulated, 5250 genes were upregulated, and 12704 genes showed no significant difference in expression between Oh7B and NY22613. There was a more marked difference in the upregulation and downregulation of genes in the greenhouse than in the field collected samples.

The multi-dimensional scaling (MDS) plots (Figure 2.3 and 2.4) show clustering of the reads collected during different time points in one dimension and two genotypes clustered in the other dimension. Each of the values were created after normalization of reads based on the count value. The MDS plots were generated for the expression data collected from the field and greenhouses separately. Both inoculated and uninoculated samples from the resistant (NY22613) and susceptible inbred (Oh7B) were included in the analysis. Multidimensional scaling plots can be used to evaluate the variance between biological replicates, and identify sample outliers and mislabeled samples. The samples collected from maize inbred NY22613 and Oh7B are clearly clustered (Figures 2.3 and 2.4) based on the genotypes and the timepoints of data collection for both greenhouse and field conditions.

The differential expression (DE) analysis between NY22613 and Oh7B, in both field and greenhouse conditions revealed more than 5000 differentially expressed genes between the two inbreds. In order to identify the genes that are related to disease resistance, a series of differential expression studies between the inoculated and uninoculated samples was conducted, including both the resistant and the susceptible inbreds and the field and greenhouse growing environments. The alpha value for false discovery rates were set to 0.001. The differential expression studies are grouped as follows:

1. Resistant inoculated Greenhouse vs. Resistant Uninoculated Greenhouse
2. Susceptible inoculated Greenhouse vs. Susceptible Uninoculated Greenhouse
3. Resistant inoculated Field vs. Resistant Uninoculated Field
4. Susceptible inoculated Field vs. Susceptible Uninoculated Field

5. Resistant inoculated Greenhouse vs. Resistant inoculated Field
6. Susceptible inoculated Greenhouse vs. Susceptible inoculated Field

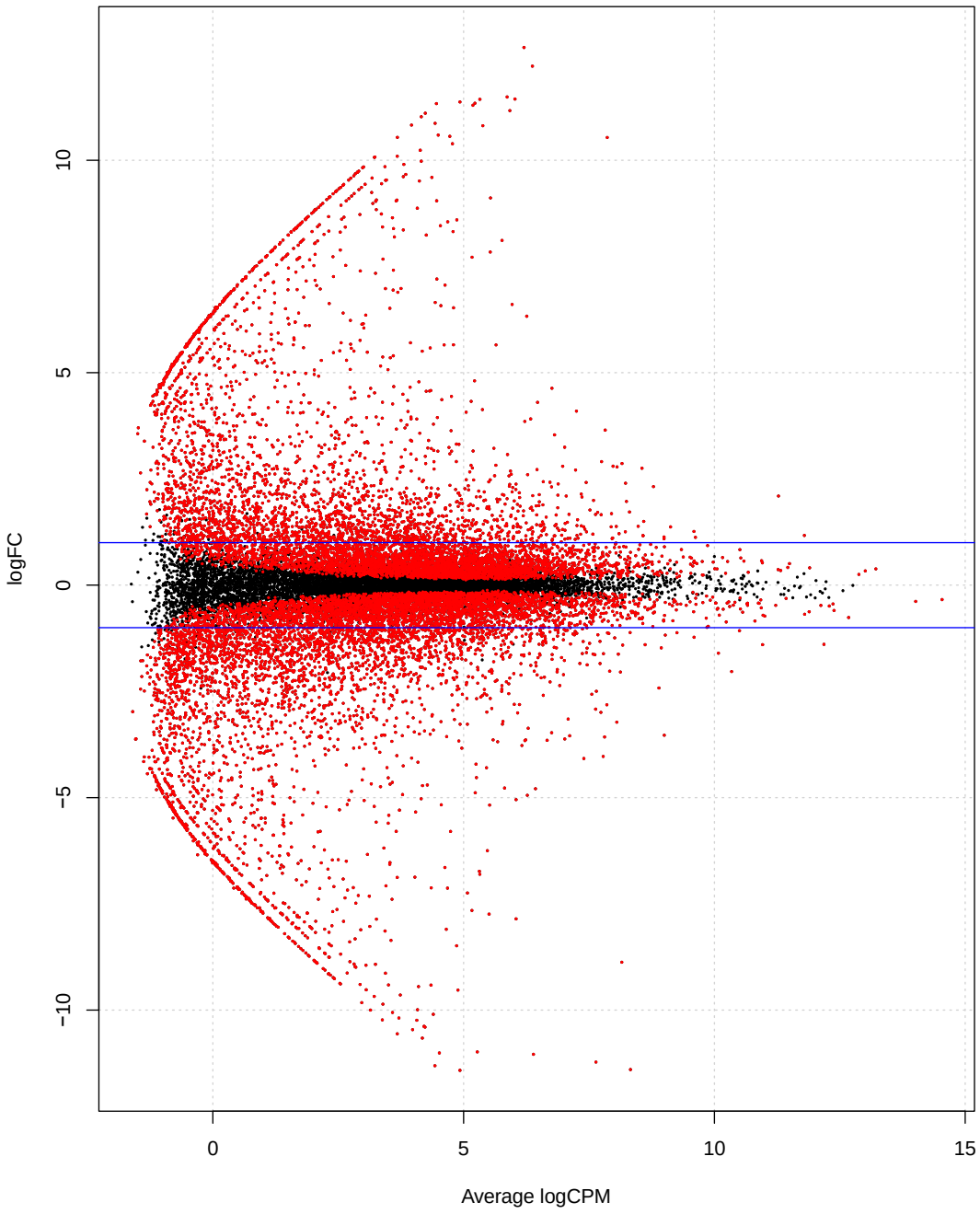

Figure 2.1 *Scree plot showing average LogCPM plotted against LogFC in the differential expression study conducted between Oh7B and NY22613 for both inoculated and uninoculated samples during all the dates in greenhouse. Significantly upregulated and downregulated genes are in red and genes in black showed no significant change. LogFC = log Fold Changes, LogCPM = log counts per million.*

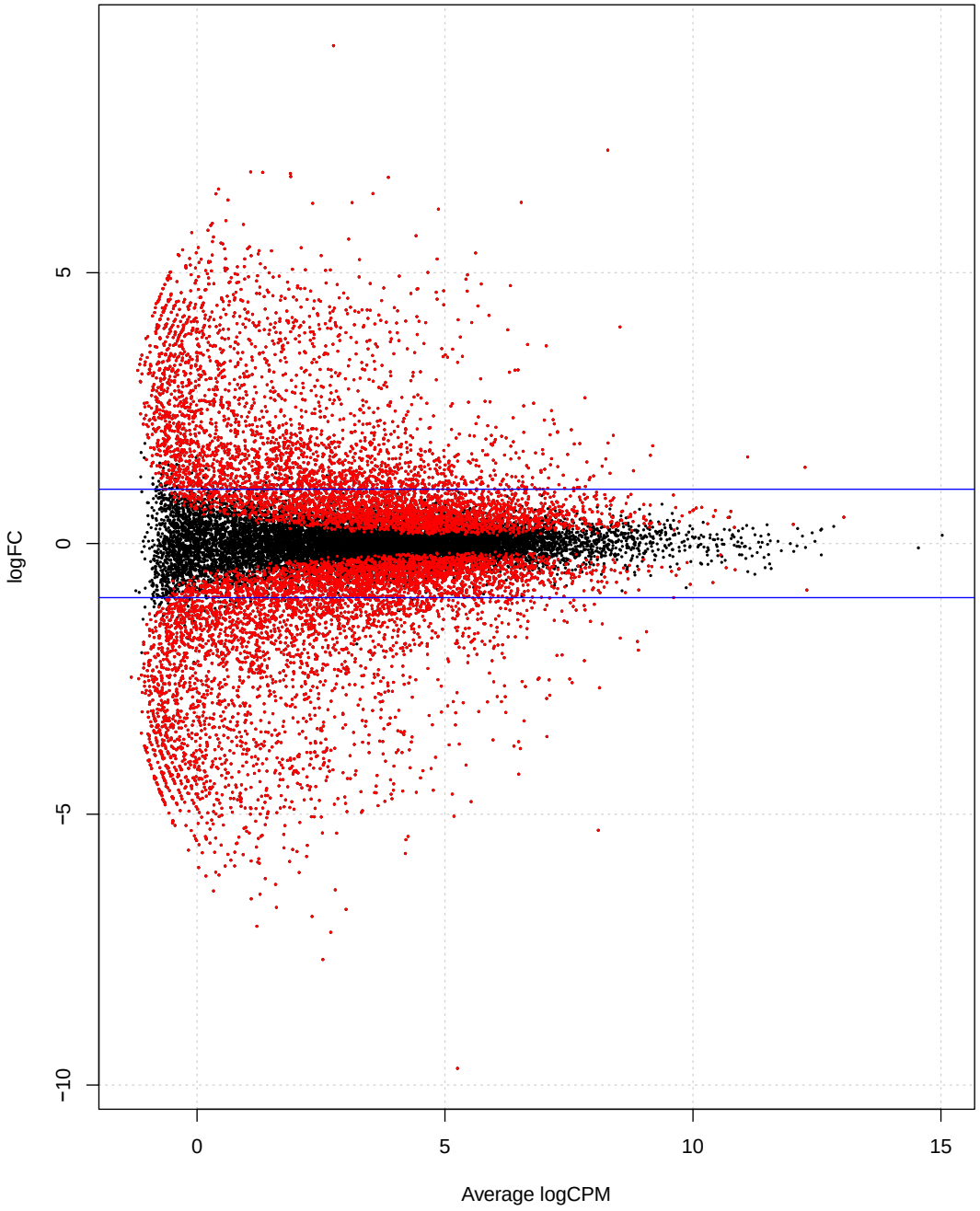

Figure 2.2 *Scree plot showing average LogCPM plotted against LogFC in the differential expression study conducted between Oh7B and NY22613 for both inoculated and uninoculated samples during all the dates in field. Significantly upregulated and downregulated genes are in red and genes in black showed no significant change. LogFC = log Fold Changes, LogCPM = log counts per million.*

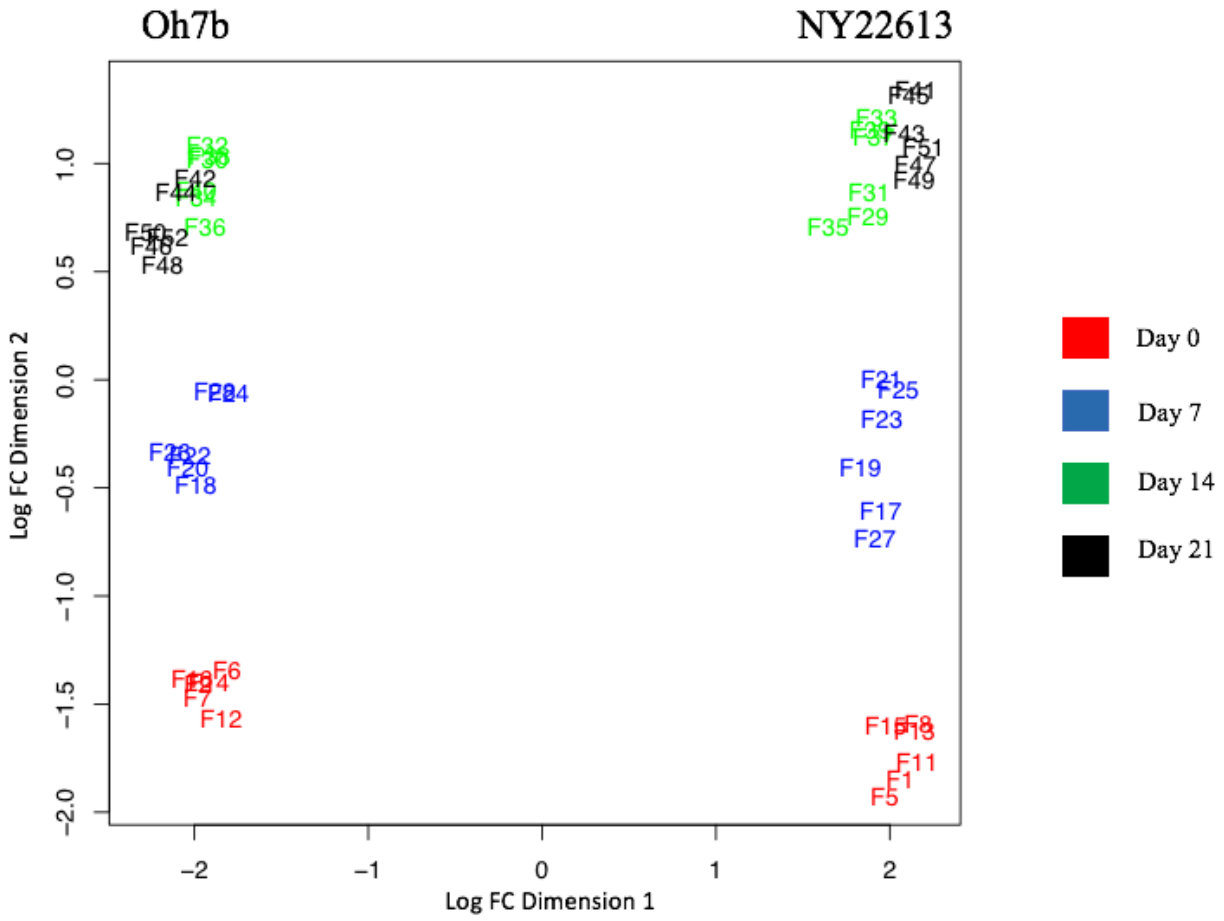

Figure 2.3 *Multidimensional scaling plot of the libraries from the field samples of maize inbreds Oh7B and NY22613 of both inoculated and uninoculated during all the four sample collection dates.*

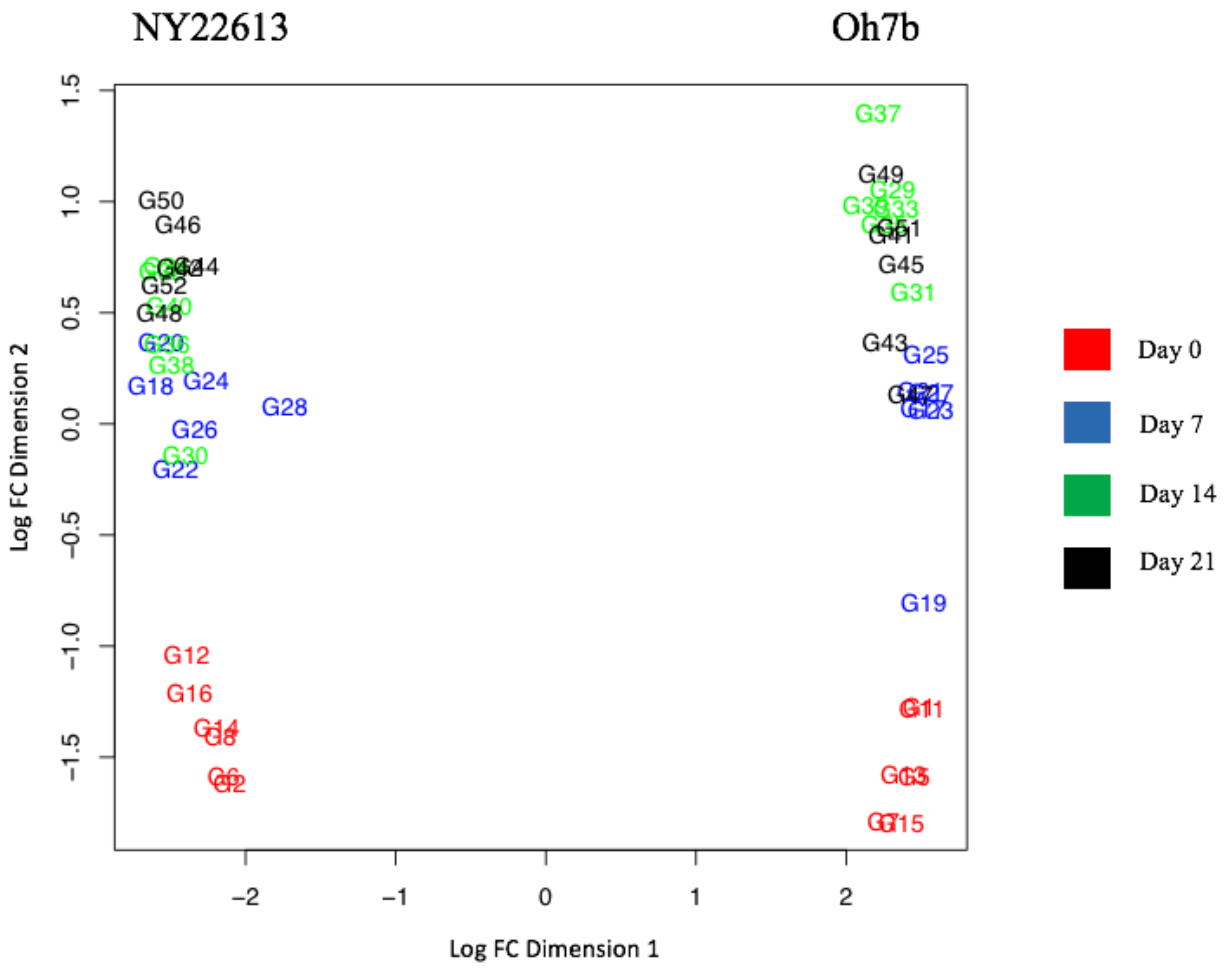

Figure 2.4 *Multidimensional scaling plot of the libraries from the greenhouse samples of maize inbreds Oh7B and NY22613 of both inoculated and uninoculated during all the four sample collection dates.*

The DE analysis between the inoculated and uninoculated NY22613 samples in the greenhouse revealed that there were 1747 genes that are differentially expressed (Table 2.5) under all the time points. However, the corresponding analysis in the field revealed only 255 genes that were differentially expressed under all the time points. The resistant inbred had many more genes differentially expressed in the greenhouse than in the field. The differential expression analysis

between the inoculated and uninoculated Oh7B samples in the greenhouse showed only 85 genes that were differentially expressed. In field samples, inoculated vs. uninoculated Oh7B samples revealed 299 differentially expressed genes. NY22613 has many more genes differentially expressed when inoculated with *S.turcica* than did Oh7B in the greenhouse but in the field Oh7B had slightly more genes differentially expressed than NY22613.

Table 2.5 *Differentially expressed genes between the inoculated and uninoculated NY22613 and Oh7B samples in greenhouse and field conditions under all the time points.*

Number of differentially expressed genes for inoculated vs. Uninoculated samples			
		Based on P-value (P<0.001)	Based on q-value (q<0.001)
Greenhouse	NY22613	1747	561
	Oh7B	85	1
Field	NY22613	255	44
	Oh7B	299	111

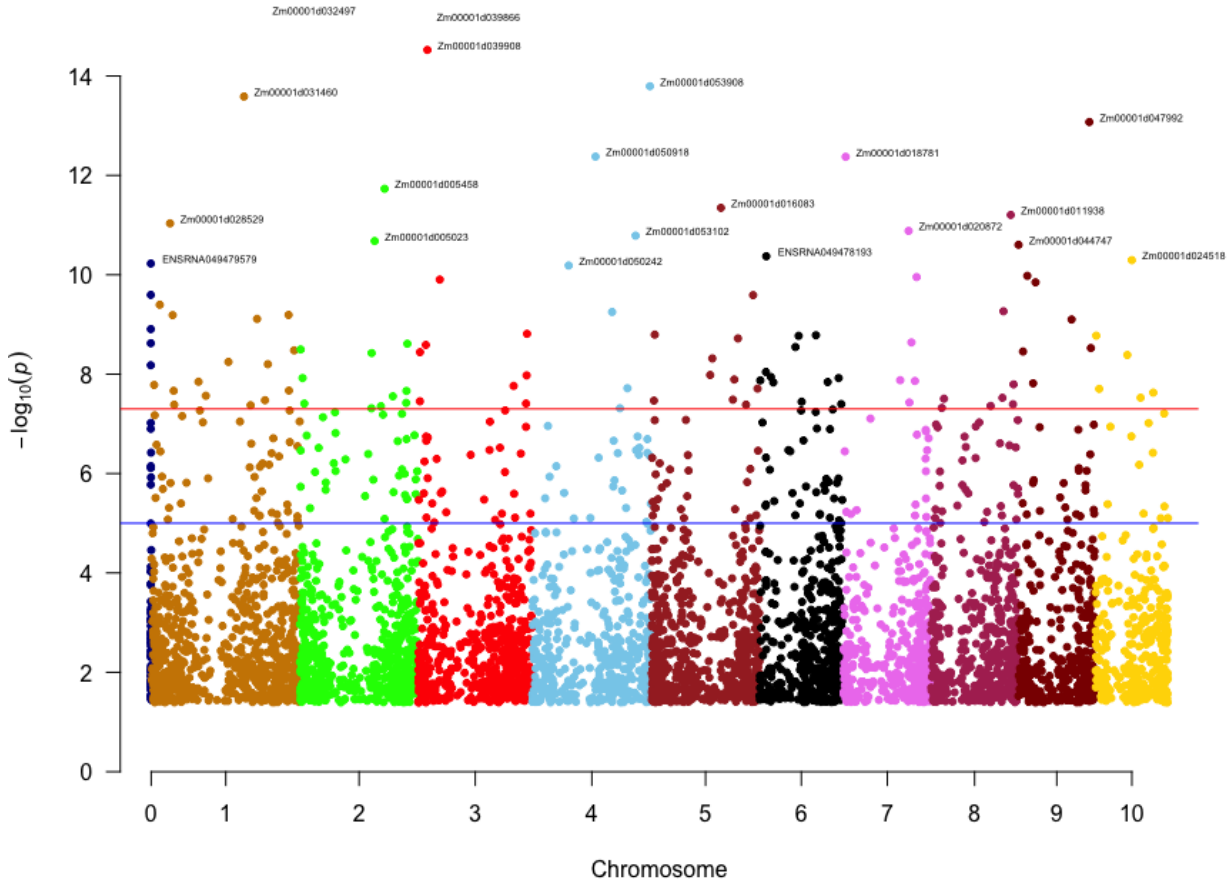

Figure 2.5 *Differentially expressed genes observed between the inoculated and uninoculated NY22613 samples in greenhouse conditions. DE genes with P<0.0000000001 are labelled. (See Figures 5.1 to 5.5 in appendix for Manhattan plots of other DE studies).*

Instead of using p values to select the differentially expressed genes, if q values (q<0.001) or false discovery rates are used, NY22613 DE analysis between the inoculated and uninoculated samples in the greenhouse revealed 561 genes differentially expressed. The DE analysis between the uninoculated and inoculated field samples for NY22613 revealed 44 differentially expressed

genes. Inoculated and uninoculated samples of Oh7B from the greenhouse showed only one differentially expressed gene. The DE analysis between the inoculated and uninoculated samples in the field revealed 111 genes differentially expressed. For further analysis, genes that were obtained using p-values were considered. This is because an alpha value of 0.001 was considered to be stringent enough for the selection of differentially expressed genes. In conclusion, if we use either p-values or q -values, when inoculated with *S. turcica* the resistant inbred reveals more genes differentially expressed than the susceptible inbred in greenhouse conditions and the susceptible inbred has more differentially expressed genes than the resistant inbred in field condition.

Collectively, the DE analyses conducted in this study revealed that the DE genes are present throughout the maize genome in both susceptible and resistant inbreds when inoculated with *S. turica* in both field and greenhouse conditions. For example, Figure 2.5 depicts a Manhattan plot that shows that the DE genes observed between the inoculated and uninoculated NY22613 samples from the greenhouse were present throughout the maize genome. Even though not all the genes are associated with disease resistance, this shows the dynamic nature of the maize genome during plant development and reaction to pathogen attack. The multi-dimensional plots in Figures 2.3 and 2.4 show that the samples collected in the field and greenhouse were clustered based on their genotype and timepoints of sample collection. In this study, the differential expression between different timepoints are not explored to focus on the few genes that are differentially expressed between the susceptible and resistant inbred during both inoculated and uninoculated conditions in field and greenhouse. However, differential expression between different time points can be explored in the future to understand the difference in the dynamic expression of the resistance between the resistance and susceptible genotypes.

2.4 Discussion

Table 2.6 *Differentially expressed gene in susceptible inbred Oh7B identified both in field and greenhouse conditions when inoculated with S. turcica.*

Gene Name	Chromosome	Location	Function
Zm00001d024772	Chr10	87301613:87302477	No previously known function in maize.

The *Zm00001d024772* gene was the only gene that was differentially expressed between the uninoculated and inoculated susceptible (Oh7B) samples both in field and greenhouse conditions. *Zm00001d024772* is a protein coding gene and the function of this gene has not been discovered in maize. However, homologs of this gene are observed in multiple species as shown in Table 2.7. Given that this gene is evolutionarily conserved across multiple species, *Zm00001d024772* may be an important candidate gene for further functional analysis.

Table 2.7 *Comparative genomic analysis of Zm00001d024772 using Conekt platform (Proost and Mutwil 2018)*

Sequence ID	Species	Description
AT1G48330	*Arabidopsis thaliana*	No description available
AT3G17580	*Arabidopsis thaliana*	No description available
GSVIVT01017708001	*Vitis vinifera*	No description available

LOC_Os01g18450	*Oryza sativa*	No description available
Solyc05g053180.3.1	*Solanum lycopersicum*	ABC-2 and Plant PDR ABC-type transporter family protein
Solyc12g098880.2.1	*Solanum lycopersicum*	TRAM, LAG1 and CLN8 (TLC) lipid-sensing domain containing protein
Zm00001d008196	*Zea mays*	No description available
Zm00001d010912	*Zea mays*	No description available
Zm00001d024772	*Zea mays*	No description available

Table 2.8 *Differentially expressed genes in the resistant inbred NY22613 identified both in field and greenhouse conditions when inoculated with S. turcica.*

Gene Name	**Chromosome**	**Location**	**Function**
Zm00001d027691	Chr1	11207688:11225217	No previously known function in maize.
Zm00001d011152	Chr8	140853342:140860116	No previously known function in maize.
Zm00001d008951	Chr8	26872145:26875807	No previously known function in maize.
Zm00001d033623	Chr1	269047817:269052874	*lox3 (lipoxygenase3)* gene – No previously known function in

			maize. Major role in immunity in rice.
Zm00001d021770	Chr7	162753384:162762778	No previously known function in maize.
Zm00001d034421	Chr1	292486760:292492680	No previously known function in maize.

Six genes were differentially expressed in response to NLB inoculation in the resistant inbred NY22613 in both field and greenhouse conditions (Table 2.8). *Zm00001d027691* has been identified in shoot apex and immature ears of maize. The function is currently unknown in maize and there are no identified homologs of this gene in other plant species. *Zm00001d011152* and *Zm00001d008951* are expressed in the whole plant. The functions of these genes are unknown in maize. Homologs of these genes are currently not available in other plant species.

Zm00001d033623 or *lipoxygenase 3 (lox3)* is a promising candidate gene for disease resistance in NY22613. Lipoxygenases (LOXs) are a class of enzymes that are non-heme-iron-containing dioxygenases. They are involved in the catalysis by peroxidation of many polyunsaturated fatty acids and lipids that results in the production of biologically active oxylipins. Oxylipins are involved in signal transduction during growth and development, senescence, and several biotic and abiotic stress responses (Feussner and Wasternack, 2002; Porta and Rocha-Sosa, 2002; Duan *et al.* 2005; Liavonchanka and Feussner, 2006). Other byproducts of LOX activity are also involved in defense responses including direct inhibition of pathogen and accumulation of phytoalexins (Alami *et al.* 1999; Lin and Ishii, 2009).

In arabidopsis, chloroplast lipoxygenase (*AT1G72520*), a gene that is homologous to *lox3*, is required for wound-induced jasmonic acid (JA) accumulation. Arabidopsis plants with this gene were found to be resistant to *Staphylococcus aureus* Rosenbach and to accumulate salicylic acid upon infection (Chauvin *et al.* 2013; Zhao *et al.* 2014; Umate 2011; Chauvin 2014; Oliw 2002). Lipoxygenases are associated with disease resistance in rice against the blast disease caused by *Magnaporthe grisea* (T.T.Hebert) M.E.Barr in rice (Marla and Singh 2012). LOX genes are also known to aid in disease resistance in various other plants such as potato (*Solanum tuberosum* L.), eastern cottonwood (*Populus deltoides* W. Bartram ex Marshall), tomato (*Solanum lycopersicum* L.) and grape (*Vitis vinifera* L.) (Kolomiets *et al.* 2000; Cheng 2006; Song 2015; Podolyan 2010).

Zm00001d021770 is a protein coding gene that is expressed throughout the maize plant. The function of this gene in maize is currently unknown. However, homologs of this gene are expressed in tomato (*Solyc04g071880.3.1*) and arabidopsis (*AT1G78920*) and are involved in pyrophosphate-energized membrane proton pumps (Bombarely *et al.* 2011). *Zm00001d034421* is expressed throughout the plant. The function of this gene is currently unknown in maize. However, a homolog of this gene, *AT3G60830,* is known to code for actin-related protein 7. This protein is required for normal embryogenesis, plant architecture and floral organ abscission in arabidopsis (Kandasamy *et al.* 2003).

The commonly differentially expressed genes between group 5 and group 6 is listed in table 5.6 in appendix.

2.4.1 Genes that are differentially expressed between greenhouse and field conditions in both NY22613 and Oh7B when inoculated with *S.turcica*

The resistant inbred had 2215 genes differentially expressed and the susceptible inbred had 2302 genes differentially expressed. Of these, 623 genes were commonly differentially expressed between greenhouse and field conditions in both resistant and susceptible inbred. Gene ontology analysis was conducted for the genes that were commonly differntially expressed between the resistant and susceptible inbreds in the afore mentioned GE analyses using the PANTHER platform (Mi et al. 2013).

Genes that are commonly expressed between the resistant and susceptible genotype when inoculated with *S.turcica* in both greenhouse and field conditions might be involved in pathogen recognition and activation of plant defense mechanisms that might result in disease resistance. The molecular function and cellular component classification of these 623 genes are shown in Figure 2.6. Catalytic activity, binding and transporter activity are the major categories of genes represented in the analysis under molecular function. Most of the genes were involved in catalytic activity under molecular functions.

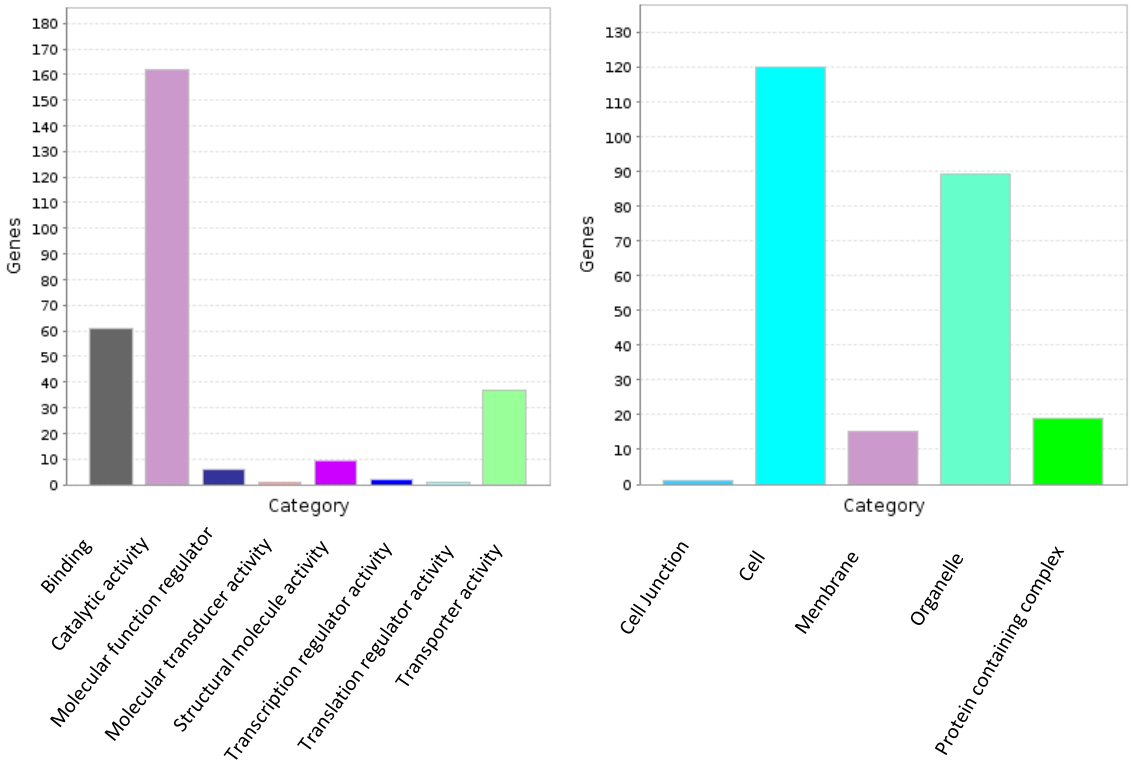

Figure 2.6 a and b *The molecular function (a) and cellular component (b) classification of the 623 genes that are commonly differentially expressed in NY22613 and Oh7B in the differential expression analysis between greenhouse and field conditions when inoculated with S. turcica.*

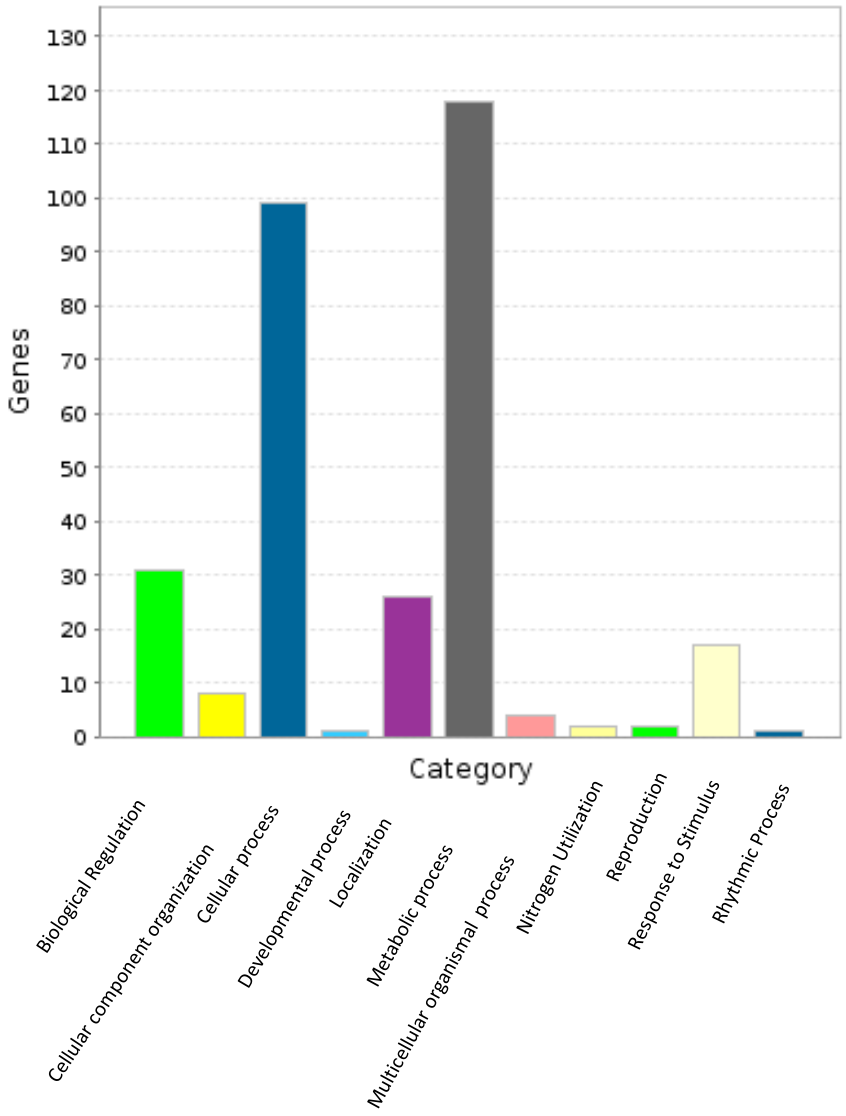

Figure 2.7 *Biological process classification of the 623 genes that are commonly differentially expressed in NY22613 and Oh7B in the differential expression analysis between greenhouse and field conditions when inoculated with S. turcica.*

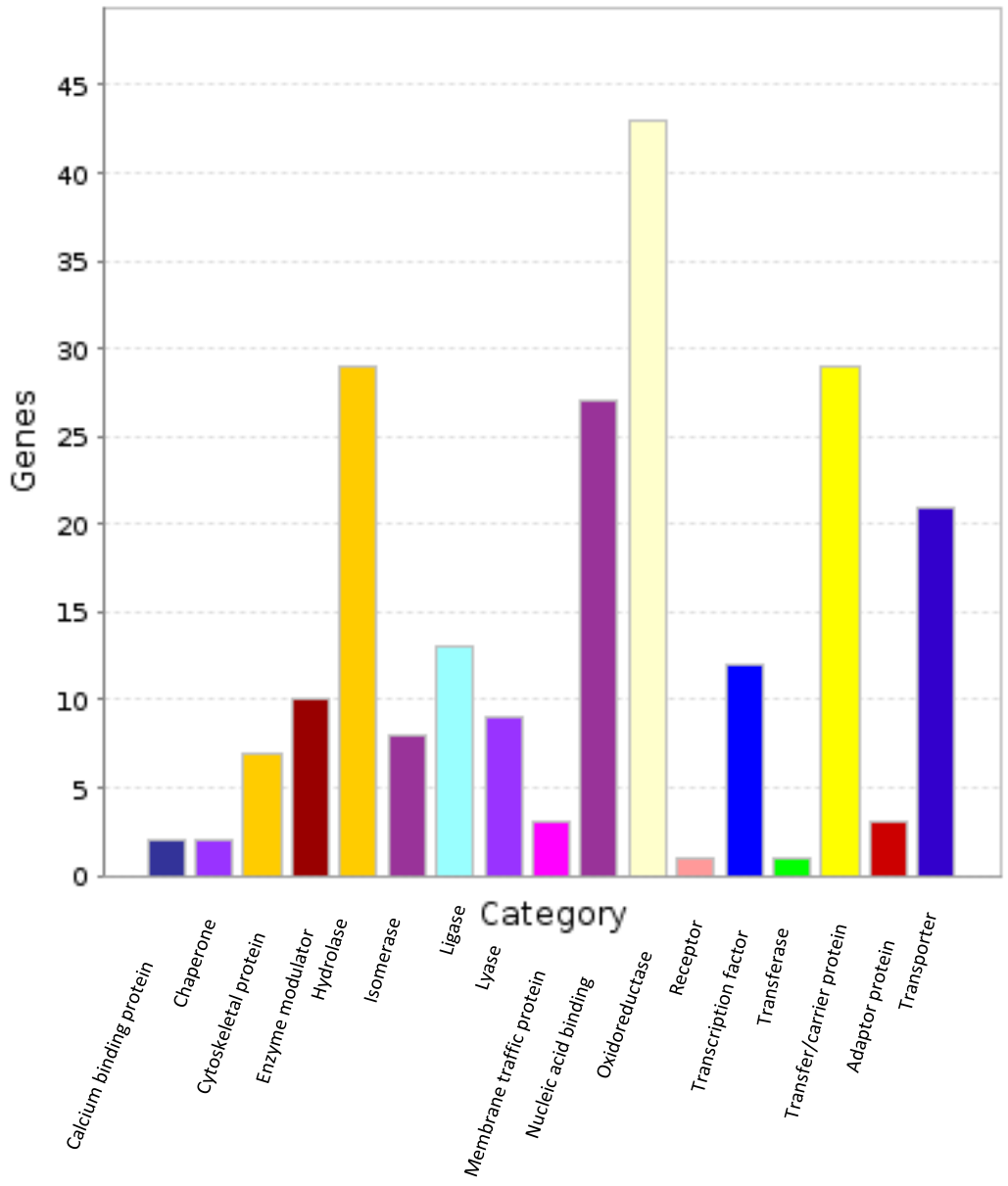

Figure 2.8 *Protein classification of the 623 genes that are commonly differentially expressed in NY22613 and Oh7B in the differential expression analysis between greenhouse and field conditions when inoculated with S. turcica, but not differentially expressed in Oh7B under the same conditions.*

Cell and organelle are the major categories of genes under cellular components. Under biological processes (Figure 2.7), most genes are classified under metabolic process, cellular process, biological regulation, localization and response to stimulus. Surprisingly, very few of these genes are classified under developmental process. Figure 2.6 and 2.7 show only the genes that have been already annotated.

Under protein classification, (Figure 2.8) most genes are classified under oxidoreductases, hydrolases, transfer/carrier protein, nucleic acid binding and transporters. Very few receptor, transferase or adaptor proteins were commonly expressed. Oxidoreductases are enzymes that have been shown to be involved in plant disease resistance and hypersensitive response in arabidopsis (Somssich, 1996; Govrin and Levine 2000).

2.4.2 Gene ontology analysis of proteins uniquely expressed in the resistant inbred NY22613 in the differential expression analysis between greenhouse and field conditions when inoculated with *S.turcica*

The resistant inbred had 1592 genes uniquely expressed in the DE study between inoculated inbred plants in greenhouse and field conditions (i.e. not differentially expressed in Oh7B under these conditions) (Appendix Table 5.6). Gene ontology analysis was conducted for the genes from NY22613 that are not commonly differentially expressed. The molecular function, cellular component and biological process classifications for the uniquely expressed genes in NY22613 had a similar pattern to the commonly expressed genes (data not shown). However, protein

classification had a markedly different pattern in the uniquely expressed genes than the commonly expressed genes (Figure 2.9).

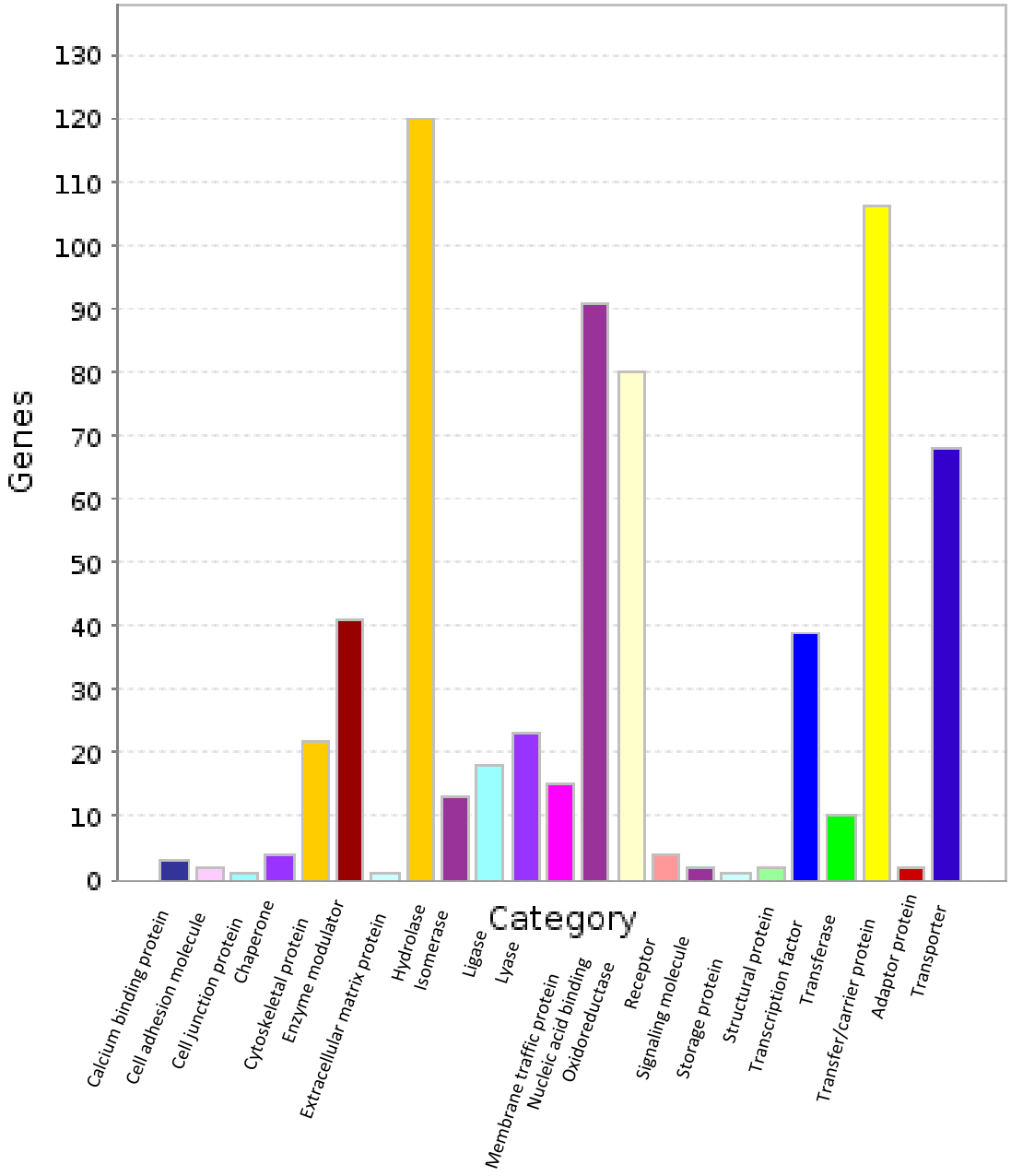

Figure 2.9 *Protein classification of the 1592 genes that are uniquely differentially expressed in NY22613 in the differential expression analysis between greenhouse and field conditions when inoculated with S. turcica.*

Cell adhesion molecule, cell junction protein, extra cellular matrix protein, signaling molecule, storage protein and structural protein are additional classes of proteins observed in the DE genes that are uniquely expressed in NY22613 between greenhouse and field conditions when inoculated with *S. turcica.*

Table 2.9 *Genes in the unique protein classes that were differentially expressed only in NY22613 between greenhouse and field conditions when inoculated with S. turcica.*

Gene ID	**Protein Classification**	**Function**
Zm00001d012494	Cell adhesion molecule	RING-type E3 ubiquitin transferase
Zm00001d050216	Cell adhesion molecule	Afadin/alpha-actinin-binding protein
Zm00001d043667	cell junction protein	Myosin family protein with Dil domain
Zm00001d042058	extra cellular matrix protein	SART-1 family protein DOT2
Zm00001d028966	signaling molecule	ELMO/CED-12 family protein
Zm00001d012494	signaling molecule	RING-type E3 ubiquitin transferase
Zm00001d012494	storage protein	RING-type E3 ubiquitin transferase
Zm00001d000443	structural protein	Uncharacterized protein
		Putative WEB family protein
Zm00001d002485	structural protein	Chloroplastic

Among these genes, *Zm00001d012494* was classified under cell adhesion molecule, signaling molecule and storage protein. Thus multifunctional proteins could play a role in the difference in disease resistance observed between the field and greenhouse conditions.

2.4.3 NLB specific expression and multiple disease resistance in NY22613

The *liguleless1 (Zm00001d002005)* gene is differentially expressed in the susceptible Oh7B between the greenhouse and field samples when inoculated with *S. turcica* (Table 5.5 in the Appendix). This gene was not commonly differentially expressed in the resistant inbred, between the greenhouse and field samples when inoculated with *S. turcica.* We conducted a QTL mapping study for NLB, GLS and SW using the Oh7B X NY22613 mapping population. In that study 49 QTLs were involved in disease resistance. Four QTLs were in regions where QTLs for resistance for more than one disease were colocalized (see Table 1.6). The *liguleless1* gene is present in the QTL region that is found in chromosome 2 between the markers S2_4211875 and S2_4257983. Interestingly, QTL for disease resistance to GLS and SW colocalized in this region; but no NLB QTL were found in this region in our study. With the expression study, *liguleless1* was differentially expressed between field and greenhouse conditions when inoculated with *S. turcica.* In the QTL analysis, *liguleless1* was associated with disease resistance to GLS and SW. Thus, it is an interesting candidate gene that might be associated with multiple disease resistance.

Identifying causative genes using expression studies can be challenging because of various confounding genes that might be detected during the analysis that might be actually associated with other functions. Further, functional characterization of the genes of interest using genome

editing methods such as CRISPR-cas9 (Clustered Regularly Interspaced Short Palindromic Repeats - Cas9), can help in understanding the function of these genes and their possible role in disease resistance.

References

Adipala, E., Lipps, P. E., and Madden, L. V. 1993. Occurrence of *Exserohilum turcicum* on maize in Uganda. Plant Dis. 77:202-205.

Alami, I., Jouy, N., & Clerivet, A. (1999). The lipoxygenase pathway is involved in elicitor-induced phytoalexin accumulation in plane tree *(Platanus acerifolia)* cell-suspension cultures. Journal of Phytopathology, 147(9), 515-519.

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. Genome biology, 11(10), R106. Bergquist, R. R., and Masias, O. R. 1974. Physiologic specialization in *Trichometasphaeria turcica f. sp. zeae* and *T. turcica f. sp. sorghi* in Hawaii. Phytopathology 64:645-649.

Borchardt, D. S., Welz, H. G., and Geiger, H. H. 1998. Genetic structure of *Setosphaeria turcica* populations in temperate and tropical climates. Phytopathology 88:322-329.

Borchardt, D. S., Welz, H. G., and Geiger, H. H. 1998. Molecular marker analysis of European *Setosphaeria turcica* populations. Eur. J. Plant Pathol. 104:611-617.

Borchardt, D. S., Welz, H. G., and Geiger, H. H. 1998. Spatial and temporal variation of genetic marker patterns in *Setosphaeria turcica* populations from Kenya. J. Phytopathol. 146:451-457.

Boshoff WHP (2000) Control of foliar rusts of wheat in South Africa with special emphasis on *Puccinia striiformis f. sp. tritici*. PhD Thesis, University of the Free State, South Africa.

Chauvin, A., Caldelari, D., Wolfender, J. L., & Farmer, E. E. (2013). Four 13-lipoxygenases contribute to rapid jasmonate synthesis in wounded *Arabidopsis thaliana* leaves: a role for lipoxygenase 6 in responses to long-distance wound signals. New Phytologist, 197(2), 566-575.

Chauvin, A. (2014). A 13-LOX hierarchy for defense in *Arabidopsis thaliana* (Doctoral dissertation, University of Geneva).

Cheng, Q., Zhang, B., Zhuge, Q., Zeng, Y., Wang, M., & Huang, M. (2006). Expression profiles of two novel lipoxygenase genes in Populus deltoides. Plant science, 170(6), 1027-1035.

Daub, M. E., & Jenns, A. E. (1989). Field and greenhouse analysis of variation for disease resistance in tobacco somaclones. Phytopathology, 79(5), 600-605.

Drechsler, C. 1923. Some graminicolous species of Helminthosporium: I. J. Agric. Res. 24:641-739.

Duan, H., Huang, M. Y., Palacio, K., & Schuler, M. A. (2005). Variations in CYP74B2 (hydroperoxide lyase) gene expression differentially affect hexenal signaling in the Columbia and Landsberg erecta ecotypes of Arabidopsis. Plant physiology, 139(3), 1529-1544.

Ferguson, L. M., & Carson, M. L. (2007). Temporal variation in *Setosphaeria turcica* between 1974 and 1994 and origin of races 1, 23, and 23N in the United States. Phytopathology, 97(11), 1501-1511.

Feussner, I., & Wasternack, C. (2002). The lipoxygenase pathway. Annual review of plant biology, 53(1), 275-297.

Gianasi, L., de Castro, H. A., and da Silva, H. P. 1996. Raçes fisiológocas de *Exserohilum turcicum* identificadas em regiões produtoras de milho no Brasil, Safra 93/94. Summa Phytopathol. 22:214-217.

Govrin, E. M., & Levine, A. (2000). The hypersensitive response facilitates plant infection by the necrotrophic pathogen *Botrytis cinerea*. Current biology, 10(13), 751-757.

Hooker, A. L., and Perkins, J. M. 1980. Helminthosporium leaf blights of corn–The state of the art. Pages 68-87 in: Proc. Annu. Corn Sorghum Res. Conf., 35th. H. D. Loden and D. Wilkinsonk, eds. ASTA, Washington, D.C.

Jenkins, M. T., Robert, A. L., and Findley, W. R. 1954. Recurrent selection as a method for concentrating genes for resistance to *Helminthosporium turcicum* leaf blight in corn. Agron. J. 46:89-94.

Jordan, E. G., Perkins, J. M., Schall, R. A., and Pedersen, W. l. 1983. Occurrence of race 2 of *Exserohilum turcicum* on corn in the central and eastern United States. Plant Dis. 67:1163-1165.

Kandasamy, M. K., McKinney, E. C., & Meagher, R. B. (2003). Cell cycle-dependent association of Arabidopsis actin-related proteins AtARP4 and AtARP7 with the nucleus. The Plant Journal, 33(5), 939-948.

Keller, N. P., and Bergstrom, G. C. 1990. Predominance in New York of isolates of *Exserohilum turcicum* virulent on maize with gene Ht1. Plant Dis. 74:530.

Kolomiets, M. V., Chen, H., Gladon, R. J., Braun, E. J., & Hannapel, D. J. (2000). A leaf lipoxygenase of potato induced specifically by pathogen infection. Plant physiology, 124(3), 1121-1130.

Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S. L., ... & Reidel, E. J. (2010). The developmental dynamics of the maize leaf transcriptome. Nature genetics, 42(12), 1060.

Liavonchanka, A., & Feussner, I. (2006). Lipoxygenases: occurrence, functions and catalysis. Journal of plant physiology, 163(3), 348-357.

Liljeroth, E., Bengtsson, T., Wiik, L., & Andreasson, E. (2010). Induced resistance in potato to *Phytphthora infestans*—effects of BABA in greenhouse and field tests with different potato varieties. European Journal of Plant Pathology, 127(2), 171-183.

Lin, T. C., & Ishii, H. (2009). Accumulation of H_2O_2 in xylem fluids of cucumber stems during ASM-induced systemic acquired resistance (SAR) involves increased LOX activity and transient accumulation of shikimic acid. European journal of plant pathology, 125(1), 119.

Lipps, P. E., and Hite, R. E. 1982. *Exserohilum turcicum* virulent on corn with the Ht resistance gene in Ohio. Plant Dis. 66:397-398.

Marla, S. S., & Singh, V. K. (2012). LOX genes in blast fungus (*Magnaporthe grisea*) resistance in rice. Functional & integrative genomics, 12(2), 265-275.

McCarthy, J. D, Chen, Yunshun, Smyth, K. G (2012). "Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation." Nucleic Acids Research, 40(10), 4288-4297.

Mi, H., Muruganujan, A., Casagrande, J. T., & Thomas, P. D. (2013). Large-scale gene function analysis with the PANTHER classification system. Nature protocols, 8(8), 1551.

Mueller, D. S., Wise, K. A., Sisson, A. J., Allen, T. W., Bergstrom, G. C., Bosley, D. B., Bradley, C. A., Broders, K. D., Byamukama, E., Chilvers, M. I., Collins, A., Faske, T. R., Friskop, A. J., Heiniger, R. W., Hollier, C. A., Hooker, D. C., Isakeit, T., Jackson-Ziems, T. A., Jardine, D. J., Kelly, H. M., Kinzer, K., Koenning, S. R., Malvick, D. K., McMullen, M., Meyer, R. F., Paul, P. A., Robertson, A. E., Roth, G. W., Smith, D. L., Tande, C. A., Tenuta, A. U., Vincelli, P., and Warner, F. 2016. Corn yield loss estimates due to diseases in the United States and Ontario, Canada from 2012 to 2015. Plant Health Prog. 17:211-222.

Oliw, E. H. (2002). Plant and fungal lipoxygenases. Prostaglandins & other lipid mediators, 68, 313-323.

Pataky, J. K. 1991. *Exserohilum turcicum* on sweet corn in Florida. (Abstr.) Phytopathology 81(suppl.):813.

Perkins, J., and Pedersen, W. 1987. Disease development and yield losses associated with northern leaf blight on corn. Plant Dis. 71:940-94

Pretorius, Z. A., PienaarA, L., & Prins, R. (2007). Greenhouse and field assessment of adult plant resistance in wheat to *Puccinia striiformis f. sp. tritici.* Australasian plant pathology, 36(6), 552-559.

Proost, S., & Mutwil, M. (2018). CoNekT: an open-source framework for comparative genomic and transcriptomic network analyses. Nucleic acids research, 46(W1), W133-W140.

Porta, H., & Rocha-Sosa, M. (2002). Plant lipoxygenases. Physiological and molecular features. Plant physiology, 130(1), 15-21.

Podolyan, A., White, J., Jordan, B., & Winefield, C. (2010). Identification of the lipoxygenase gene family from *Vitis vinifera* and biochemical characterisation of two 13-lipoxygenases expressed in grape berries of Sauvignon Blanc. Functional Plant Biology, 37(8), 767-784.

Raymundo, A., and Hooker, A. 1981. Measuring the relationship between northern corn leaf blight and yield losses. Plant Dis. 65:325-327.

Renfro, B. L., and Ullstrup, A. J. 1976. A comparison of maize diseases in temperate and in tropical environments. Proc. Natl. Acad. Sci. USA 22:491-498.

Robinson MD, McCarthy DJ, Smyth GK (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." Bioinformatics, 26(1), 139-140.

Smith, D. R., and Kinsey, J. G. 1980. Further physiologic specialization in *Helminthosporium turcicum.* Plant Dis. 64:779-781.

Shankara, K., and Gowda, K. 2011. Assessment of yield loss due to turcicum leaf blight of maize caused by *Exserohilum turcicum*. Mysore J. Agric. Sci. 45:691-692.Thakur, R. P., Leonard, K. J., and Jones, R. K. 1989. Characterization of a new race of *Exserohilum turcicum* virulent on corn with resistance gene HtN. Plant Dis. 73:151-155.

Somssich, I. E., Wernert, P., Kiedrowski, S., & Hahlbrock, K. (1996). *Arabidopsis thaliana* defense-related protein ELI3 is an aromatic alcohol: NADP+ oxidoreductase. Proceedings of the National Academy of Sciences, 93(24), 14199-14203.

Smith, D. R., and White, D. G. 1988. Diseases of corn. In: Corn and Corn Improvement, 3rd ed. Am. Soc. Agron. Madison, WI.

Song, Y., Chen, D., Lu, K., Sun, Z., & Zeng, R. (2015). Enhanced tomato disease resistance primed by arbuscular mycorrhizal fungus. Frontiers in plant science, 6, 786.

Sun, Qi (2019) Director, Biology High Performance Computing (BioHPC), Cornell University, Ithaca, NY.

Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., ... & Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. Genome research, 13(9), 2129-2141.

Turner, M. T., and Johnson, E. R. 1980. Race of *Helminthosporium turcicum* not controlled by *Ht* genetic resistance in corn in the American corn belt. Plant Dis. 64:216-217.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., ... & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols, 7(3), 562.

Twizeyimana, M., Ojiambo, P. S., Ikotun, T., Paul, C., Hartman, G. L., & Bandyopadhyay, R. (2007). Comparison of field, greenhouse, and detached-leaf evaluations of soybean germplasm for resistance to *Phakopsora pachyrhizi*. Plant Disease, 91(9), 1161-1169.

Umate, P. (2011). Genome-wide analysis of lipoxygenase gene family in Arabidopsis and rice. Plant signaling & behavior, 6(3), 335-338.

Van Valen, L. A New Evolutionary Law. Evol. Theory 1, 1–30 (1973).

Wang, P., Souma, K., Kobayashi, Y., Iwabuchi, K., Sato, C., and Masuko, T. 2010. Influences of northern leaf blight on corn silage fermentation quality, nutritive value and feed intake by sheep. Anim. Sci. J. 81:487-493.

Welz, H. G., Wagner, R., and Geiger, H. H. 1993. Virulence variation in *Setosphaeria turcica* populations collected from maize in China, Mexico, Uganda, and Zambia. (Abstr.) Phytopathology 83:1356.

Welz, H. G. 1998. Genetics and Epidemiology of the Pathosystem *Zea mays / Setosphaeria turcica*. Shaker Verlag, Germany.

Welz, H. G., and Geiger, H. H. 2000. Genes for resistance to northern corn leaf blight in diverse maize populations. Plant Breed. 119: 1-14.

White, D. G. (ed.) 1999. Compendium of Corn Diseases, 3rd ed. The American Phytopathological Society, St. Paul, MN.

Windes, J. M., and Pedersen, W. L. 1991. An isolate of *Exserohilum turcicum* virulent on maize inbreds with resistance gene *HtN*. Plant Dis. 75:430.

Yang, X. Y., Jiang, W. J., & Yu, H. J. (2012). The expression profiling of the lipoxygenase (LOX) family genes during fruit development, abiotic stress and hormonal treatments in cucumber *(Cucumis sativus L.*). International journal of molecular sciences, 13(2), 2481-2500.

Zhao, Y., Zhou, J., & Xing, D. (2014). Phytochrome B-mediated activation of lipoxygenase modulates an excess red light-induced defense response in Arabidopsis. Journal of experimental botany, 65(17), 4907-4918.

CHAPTER 3

SCIENCE COMMUNICATION OF QUANTITATIVE GENETICS FOR PUBLIC UNDERSTANDING OF DIRECT TO CONSUMER - GENOME TESTING

3.1 Introduction

In science, distinct disciplines exist in order to study and specialize in a field. During this pursuit of specialization, it is easy to disregard the application of ideas developed in a certain area of research to a different context. The ideas developed in crop genetics can easily be translated in the field of animal and human genetics and *vice versa*. For example, several ideas that have been developed in plant genetics (like the discovery of genetic theory in pea plants by Mendel), laid the foundation of the field of genetics of human and animal diseases. Several data analysis methods (like mixed models) that were developed in animal breeding have been widely used in plant breeding (Kampourakis 2013; Searle 1991). Thus ideas emanate between different fields of research in biology. Despite this phenomenon in academia, due to a number of reasons, ideas in genetics are sharply separated from the public (Bensaude-Vincent 2001). History has shown that if ideas are not transparent, lack of trust between scientists and the public can result in unforeseen societal and political consequences (Garvin 2001).

We are currently at a crucial point in history, where low cost genome sequencing has enabled the masses to have their own, and in some cases their family's genomic data. This is because low cost genome sequencing technologies enabled direct to consumer genetic testing

(DTC-GT) companies to estimate ancestry and predict disease using genomic data (Fujimura and Rajagopalan 2011). Most DTC-GT companies achieve this through genome wide profiling and conducting genome wide association studies (GWAS) on individuals (Bloss *et al.* 2011). GWAS has been used to study trait genetics in plants, animals and humans (Hayes 2013). Genomic selection and prediction studies have been more widely used on plant and animal genomic data more than on human genomic data. This is mainly due to the large scale availability of plant and animal genomic data from industry and academic breeding programs (Daetwyler et al. 2013; Brachi *et al.* 2011). However, the rise of DTC-GT companies has the potential to increase the number of genome wide association and genomic prediction studies using human genomic data. Among the criteria for collecting human genome data ethically are clear communication with publics about the scientific methods being used to collect and use human genomic data (Fraker and Mazza 2011). This situation precipitates the need for governments and educators of our generation think seriously about communicating various aspects of genomic data.

However, public understanding of genetics is a complex communication challenge that has been studied by several authors. In the US, it has been found to range from believing that genes absolutely determine a person's health to believing that genes might make them susceptible to certain health conditions (Parrot *et al.* 2012). The lack of efficient communication of the results of genetic testing has been reported in various studies (Powell 2012; Hock 2011; Marteau 2010). Communicating the risks and benefits of genomic technologies to the public is an area that is largely unaddressed. This chapter explores the potential of science communication of quantitative genetics to the public through the story of quantitative genetics in agriculture and the participation of plant and animal breeders/geneticists in the science communication of quantitative genetics and

policy related to DTC-GT. To understand the issues that publics should know about quantitative genetics, it will be helpful to understand several key concepts: low cost genome sequencing, genome wide association studies (GWAS), and genomic prediction.

3.2 Low Cost Genome Sequencing

Low cost genome sequencing has led to an abundance in data collected in biology. The sequence reserve archives of the United States National Institutes of Health (NIH) currently harbor ~32,000 microbial genomes, ~5,000 plant and animal genomes, and ~250,000 individual human genomes.

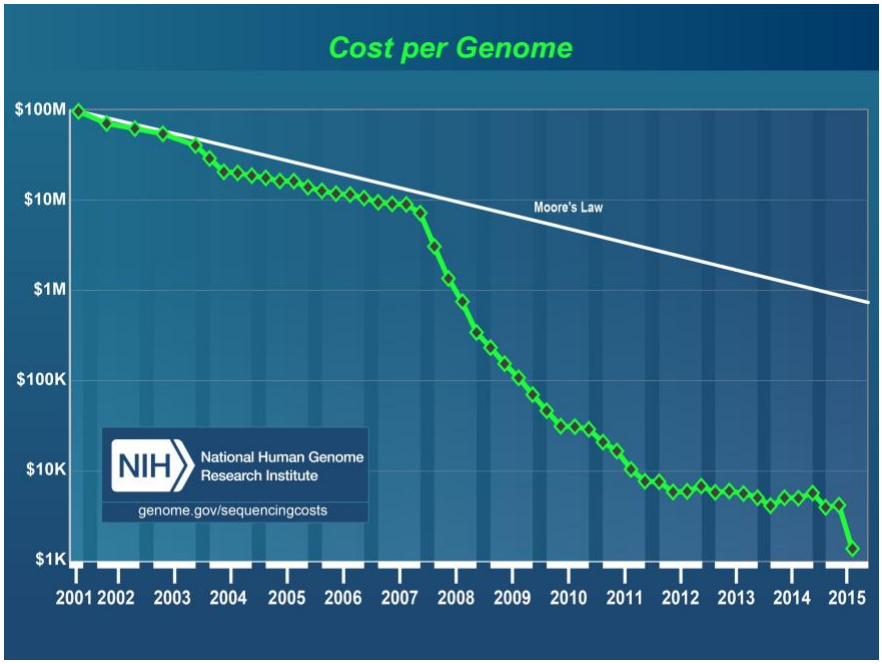

Figure 3.1 *Cost for sequencing per genome has decreased sharply from the year 2001 to the year 2015. The cost has gone down much faster than predicted by Moore's law. Note that estimating the accurate cost for genome sequence is hard due to the complex nature of the process.*

The development in computation and the recent 5G technology enables transfer and analysis of large genomic data sets within a much shorter time. The time when people can view their genomes through their smart phone applications is not far and it is therefore important to educate the public on genetics and genomics in order for them to understand their genome. By 2025, the rate of data generated through genomics will exceed the data generated in the fields of astronomy, YouTube and Twitter and approach one zetta base of sequence data/year (Stephens *et al.* 2015). While the data generated through Twitter and YouTube is readily available to anyone with access to internet, genomic data has been largely unavailable to the public. This trend is changing rapidly due to the decreasing cost of genome sequencing and increasing computing power.

The first human genome was sequenced in 2001 and it cost US $2.7 billion. Since then remarkable developments in DNA sequencing technology have resulted in significant reduction in cost and the time required to sequence a human genome (Figure 3.1). After the first genome was sequenced, scientists predicted the cost of DNA sequencing could go down to $1000 (Wolinsky 2007). It is important to note that sequencing is not genotyping. DTC-GT companies like 23andMe and AncestryDNA look into less than one percent of an individual's complete genome sequence. As shown in the Figure 3.1, development in DNA sequencing technology have been much faster than the development in transistor technology. By 2015, the cost of sequencing whole human genomes was less than $1000. In 2018, Veritas Genetics offered to sequence an individual's genome for less than $200 (Molteni, 2018). This relatively low cost of genome sequencing led to the rapid development of low cost direct to consumer genome technologies (DTC-GT) and personal genomics. Along with this rise in DTC-GT technology arose complex ethical, legal and social

issues that are largely unaddressed by companies offering DTC-GT and personal genomics services (Molteni, 2018).

By 2018, more than one million people had their whole genomes sequenced and 17 million people had their DNA analyzed with direct-to-consumer tests (Molteni, 2018). Expecting the public to understand their genomic data using quantitative genetics can occur only after rigorous outreach efforts by multiple academic and governmental institutions. Public universities and regulatory agencies have an opportunity to play a role in helping the public understand their genomic data clearly so that they can make informed life decisions through their genomic data. One way to communicate genomics to the public is through communication of quantitative genetics. So far, only the genetics community has really delved into quantitative genetics. The eugenic origins of quantitative genetics in humans has prevented it from being communicated to the masses (Kevles 1995). While the history of quantitative genetics in humans is ugly, it is inspiring in agriculture. Quantitative genetics has enabled plant breeders to save millions of lives in the developing world by breeding high yielding crops. It has improved the milk and egg production by several times. It has helped us increase crop production and is one of the major tools employed by scientists to produce enough food for our future. These stories in agriculture can be used to explain quantitative genetics to a wider audience of the public.

3.3 Genome Wide Association Studies

GWAS is a research tool to study the genetic basis of traits and was first developed in humans to study myocardial infarction (Ohnishi *et al.* 2002). Due to low cost genome sequencing and other

developments in genomics, it is currently used consistently for gene discovery, especially to reveal risk alleles for complex traits (non-Mendelian traits usually controlled by multiple genes) in humans, plants and animals (Hirschhorn 2005). However, GWAS is only able to explain a small fraction of the heritability of complex traits through the variants (high density molecular markers that are associated with the trait). This missing heritability of complex traits while using GWAS is a major drawback in studying complex traits such as disease resistance in humans (Manolio 2009; Schork et al. 2013). This is because the genetic architecture of such traits might be controlled by many genetic variants with individually small effects, thus limiting the power of GWAS (Glazier et al. 2002). For example, genome wide association studies of obesity explain less than two percent of the variance observed in heritable body mass index. Even though GWAS only explains a small fraction of the heritability at this point in time, it is still one of the most powerful tools available to study complex traits. Well-designed GWAS studies have the potential to identify causative alleles or the nature of inheritance of several important quantitative traits such as yield in agriculture and heart disease in humans.

However, the clinical validity of GWAS for complex traits has been questioned by scientists and medical professionals. This is because GWAS is meant to be used as a research tool by scientists rather than a diagnostic tool. Credible genetic associations, positive and negative predictive values of single nucleotide polymorphisms (SNPs) and the uncertainty of risk due to the interaction between genes are some of the major concerns with genome wide association studies (Fraker and Mazza 2012).

In order to have standardized practice, researchers who conduct GWAS are encouraged to have large and diverse study samples from populations that provide useful information about the trait being studied; molecular markers that are polymorphic for the trait being studied, inexpensive and reliable; they are also asked to employ experimental designs and statistical methods that reduce bias and increase statistical power (Hirschhorn 2005). Despite these recommendations, practical difficulties prevent researchers from having a standard GWAS protocol, especially in humans. GWAS in humans for certain diseases has been shown to be not diverse enough to include individuals from many ethnic groups and can result in spurious diagnosis in minority communities (Medina-Gomez *et al.* 2015; Shields *et al.* 2005).

GWAS has been used to predict educational achievement of students and has been suggested as a basis for designing educational systems (Selzam et al. 2016). However, GWAS on cognitive abilities and intelligent quotient (IQ) tests do not account for social and environmental conditions accurately and have been called "meaningless" (Richardson 2017). If this is not communicated to the public properly, such meaningless studies can impact society. For example, the use of IQ to predict education success has been debunked over and over again (McCall 1977; Berliner *et al.* 2014). However, the challenges in communicating this to the public has led to people still using IQ tests to predict intelligence (Stanovich 2009).

In another example, a GWA study identified 44 risk loci that are linked to major depressive disorder (MDD) (Wray et al. 2018). Even though GWAS for diseases like MDD help in the identification of major molecular pathways that are associated with the trait, using them as a diagnostic tool is not recommended. This is because complex traits like MDD are controlled both

by several genes and by a number of environmental factors, as well as the interaction between the genes and the environment. Another major drawback in predicting MDD using genotyping is that traits like MDD vary over time (Dunn et al. 2015).

In 2017, multiple GWA studies were used to study male sexual orientation and some studies have shown significant association of markers to sexual orientation (Svrakic 2017; Sanders 2017). However, other studies have shown no significant association of markers to sexual orientation (Ramagopalan *et al.* 2010). Just like MDD, sexual orientation is also a trait that is affected by a number of environmental and social factors. These GWA studies on quantitative traits can be used to genotype individuals and used to classify humans based on their genetics. Such studies, while helping us understand more about the human condition and biology, should also be communicated to the public clearly. If not, the social ramifications due to genomic data can be quite damaging to societies. This is because classifying people based on their genetics can put minorities at risk, especially in parts of the world that are intolerant to diversity.

Several other quantitative traits that are explained by GWAS suffer from similar drawbacks. Even in the scientific community, the drawbacks of GWAS are not completely understood by all researchers (Begum *et al.* 2012). Academia suffers from problems like p-hacking where researchers modify their data or statistical tests to obtain a significant result (Head *et al.* 2015) Expecting the public to understand the limitations of GWAS and asking them to make decisions based on them is risky business, to say the least.

Even though genotyping and GWAS struggle with explaining quantitative disorders in humans, they can scan an individual for certain important Mendelian traits and even help people identify long lost family (Meagher and Berg, 2018). However, the questions related to oversight, responsibility to disseminate information, privacy, personal and social value are still unanswered. In order to address these issues, the public should be made aware of the potential of DTC-GT and genomics. The US National Academies of Sciences, Engineering and Medicine recommended enhancing credible knowledge synthesis and dissemination of information to providers of DTC-GT and to consumers (Fraker and Mazza 2011). Communication of genetics, genotyping and genomic technologies to the public is a complex process. Communicating how GWAS works by addressing the benefits and drawbacks of the technology using quantitative genetics is a clear and direct approach that can start to address the knowledge gap between the academia and the public.

3.4 GWAS and quantitative traits

Quantitative genetics deals with inheritance of traits that are continuous in nature like human height and yield in plants. Traits that are qualitative in nature are discontinuous, such as eye color in humans or flower color in peas. These traits are studied by a field of genetics called Mendelian genetics (Lynch and Walsh 1996). Mendelian traits have virtually 100 percent probability of being expressed when the relevant gene is present, like Huntington's disease (*HTT* gene), or a significant increase in the probability of being expressed like the breast cancer genes *BRCA1* and *BRCA2* (Fraker and Mazza 2011).

Studies have shown that the public are interested in knowing more about quantitative traits like cancer rather than Mendelian diseases like thalassemia, which are less common among the general population. However, since quantitative genetics is a sophisticated field and requires advanced skill sets, even many biologists are not well versed in it (Lynch and Walsh 1996).

The major option available for the public to currently understand their genomic data is through genetic counselors. However, access to genetic counselors is limited based on health insurance coverage and state regulation. In the US, only 45 universities offer accredited genetic counseling training (ACGC, 2019). According to the National Society of Genetic Counselors (NSGC), only 22 states provide genetic counseling licenses. Genetic information interpretation is a specialized skill set. Even medical professionals such as nurses and primary care physicians have been shown to struggle with interpreting genetic data (Anderson 1996; Daack-Hirsch et al. 2012; Prows et al. 2005; Powell *et al.* 2012; Nippert et al. 2011). The Accreditation Council for Genetic Counseling (ACGC) states that the number of genetic counselors needs to increase fast and they need to be licensed by state in order to provide services.

In the US, health insurance providers cover genetic testing variably and genetic counseling is not usually covered. The Access to Genetic Counselor Services Act of 2018 proposes the coverage of genetic counseling by Medicare, the national health insurance program in the United States, that is under the Social Security Administration. This act has the potential to increase the availability of genetic counselling to people for clinical genetic testing, but not DTC-GT. The genome sequencing companies like Veritas Genetics offer free counseling services for 30 minutes. DTC-GT companies offer introductory genetics training through their websites (Molteni, 2018).

However, the efficacy of these training strategies in providing accurate information about quantitative trait inheritance is yet to be studied. Thus access to accurate information about quantitatively inherited diseases for DTC-GT is sparse and could result in misinterpretation of genetic testing. The social and ethical implications of this knowledge gap between the consumers and the DTC-GT providers can be only speculated. Communication programs that help in teaching quantitative genetics can be a part of the solution to this problem.

Another reason for communication of quantitative genetics to the public is its complexity relative to qualitative genetics. Genotyping results for traits that are inherited in a Mendelian fashion are straight forward. In humans, diseases that are inherited in Mendelian fashion are not common (usually one in one thousand in small populations). Despite this, 3,678 Mendelian genes are identified to be definitively linked to disorders in humans. Quantitatively inherited diseases like Alzheimer's, cancer, cardiovascular disease and diabetes are more complex and are common in the general population. However, only 505 genes are definitively associated with complex or quantitative traits in humans (OMIM, 2019). This demonstrates how the inheritance of quantitative traits needs more communication programs than the inheritance of Mendelian traits.

Further, several GWAS studies have been done in plants. In plants, GWAS has been used to study traits like yield, disease resistance, height, flowering time and drought tolerance (Wang et al. 2019; Chen et al. 2015; Wang and Qin 2017; Peiffer et al. 2014; Buckler et al. 2009). Communicating how GWAS works with plant traits like yield can be less controversial than using human quantitative traits like intelligence. Geneticists could use the plant traits to generate communication

strategies and educators could use plant genetics to design course materials for schools that could teach about GWAS and how it works.

3.5 Genomic Prediction

For the most part of the 20th century, the availability of morphological, biochemical or genetic markers was the major constraint in plant and animal selection. The discovery of single nucleotide polymorphisms (SNPs) and their abundance totally changed the field of plant breeding into a data driven field. Using the genome of an organism (usually genome-wide SNP data) to predict its performance in the future can be referred to as genomic prediction. In plant breeding, genomic prediction is a strategy in which individuals of a breeding population are selected in an early stage (sometimes even at seed stage) based on their genomic data. Individuals are selected based on genomic estimated breeding values (GEBV) that are developed in a training population (Meuwissen et al. 2001). Plant and animal breeders have used genomic prediction to select for quantitative traits. Several quantitative genetics concepts like heritability, best linear unbiased predictors and GEBVs were in fact developed in the field of animal and plant breeding (Hayes 2013; Bernardo 2010).

The development of genomic selection has revolutionized animal and plant breeding (Hayes 2013). Genomic selection assumes that all the molecular markers (SNPs) might be linked to a gene affecting the trait. This strategy focuses on estimating the effect of molecular markers rather than testing for their individual significance (Heffner et al. 2009) This process has been widely successful in plant and animal breeding programs in selecting for quantitative traits like yield, milk

production etc. (Goddard et al. 2007). The spurious availability of human genomic data and the ethics of looking at certain traits were the major hurdles for such studies in humans (Abraham et al. 2016).

Before the availability of single nucleotide polymorphism (SNP) molecular markers, usage of other molecular markers (e.g., restriction fragment length polymorphisms or simple sequence repeats) and biochemical markers was practiced in animal and plant breeding through practices like marker assisted selection (MAS). Molecular markers in plant breeding have been used for selecting simply inherited traits that are hard to select for using phenotypic selection since they were first available (Bernardo 2014). This is similar to the genetic testing conducted in sperm banks to identify donors with major diseases that are inherited in Mendelian fashion. The ethics related to genetic testing is discussed by several authors (Benward 2015; Aziza-Shuster 1994). The ethics of use of genomic tools in human genomic data should be studied more and communicated to the public.

Genomic prediction and genomic selection strategies for multiple crop and animal breeding scenarios have been discussed by several authors (Crossa et al. 2013; Windhausen et al. 2012; Morota et al. 2014; Hayes et al. 2010; Edwards et al. 2019). Without proper regulation or ethical guidelines, one can imagine how the quantitative genetic tools developed in plant and animal breeding (e.g., genomic prediction and selection models) could be easily used for selecting human embryos based on their genetics for quantitative traits like intelligence. In fact, some companies are actually attempting to conduct selection in human embryos for polygenic and single gene disorders. A company named “Genomic Prediction” has raised $4.5 million. They offer tests that

cost between $150 to $1000 for selecting individual embryos (Genomeweb 2019). Even though the name of the company might not have anything to do with the selection strategy used in such practices, the practice of selecting embryos for quantitative traits using a genomic prediction strategy developed through plant and animal breeding raises questions on oversight. Science communication can help the society to come together and make informed decisions on these issues.

Genomic prediction strategies willfully ignore the biology between molecular markers and traits. Assigning individuals or embryos a breeding value based on their genetic information and using genomic prediction as an diagnostic tool in humans or embryos is unethical. While discriminating against embryos doesn't have oversight, discriminating humans based on their genomic information is illegal in the US. However, several countries have little or no legislation on oversight of using genomic data.

This necessitates the public to get educated on quantitative genetics or else individuals who don't have the privilege of protecting their genetic information will suffer the consequences of genetic information discrimination. GWAS and genomic predicction have served well as research tools, however their use as diagnostic tools in humans is highly questionable and can be highly damaging to societies (Kay et al. 2010). The geneticists who develop GWAS and genomic prediction tools can be in the plant, animal or human genetics world. They should be able to communicate to the lawmakers on why these tools are not suitable for diagnostic purposes in humans. Further, geneticists can play a major role in making the "black box" of GWAS and genomic prediction more transparent so that they can be used ethically to benefit humanity.

3.6 Regulation of DTC-GT

In the US, the laws that govern the regulation of DTC-GT are the 1976 Medical Device Amendments, the Food, Drug, and Cosmetic Act, the Clinical Laboratory Improvement Amendments, the Health Insurance Portability and Accountability Act, and the Genetic Information Nondiscrimination Act (Fraker and Mazza 2011). The regulatory authority that is involved in regulating DTC-GT is the Food and Drug Administration (FDA). In the US, due to regulatory reasons, DTC-GT companies like AncestryDNA and 23andMe focused on ancestory tests alone. The validity and ethics of these ancestry tests have been questioned by several geneticists, ethicists and anthropologists (Schaper and Schicktanz 2018). In some cases, even twins were shown to have different DTC-GT sets of results (Agro and Denne 2018). In 2017, the FDA allowed DTC-GT companies to conduct screening for ten human diseases after carefully reviewing and confirming that certain genetic markers are more consistently linked with those human diseases (Ramos and Weissman 2018) (See Table 3.1).

Globally, the regulatory scenario for direct to consumer genetic testing is changing fast. For example, in the European Union, DTC-GT is subject to the 1998 In Vitro Diagnostic Devices Directive (IVDD). In the UK, in addition to IVDD, the Human Tissue Act of 2004 also regulates the DTC-GT. The regulatory institutions for DTC-GT in the UK are the Genetic Testing Network and the Human Genetics Commission. In 2003, the Human Genetics Commission advised the use of DTC-GT only in a clinical setting with pre-and post-test counseling. However, by 2007, the commission classified DTC-GT as medium risk and did not require the strict regulatory procedures

that it suggested in 2003. In 2009, the UK encouraged the DTC-GT companies to come up with a universal code of practice (Fraker and Mazza 2019).

In Germany, by April 2009, the parliament passed legislation that banned DTC-GT. Only physicians could administer genetic tests to patients with informed consent (Rafiq et al.. 2015). In most of the developing countries like India, even though DTC-GT is available, the regulations are still being developed or there are no regulations for such tests (Wright and Gregory-Jones 2010).

Regulation of the same technology has been very different in the plant, animal and human genetic worlds. In humans, genetically engineered t-cells have been used to cure leukemia (Porter et al. 2011). Genetically engineered viruses have been used to treat brain cancer (Desjardins et al. 2018). Although these genetically engineered organisms are usually experimental procedures, in some cases genetically engineered viruses have been approved by the FDA for treatment. For example, FDA approved the treatment of inherited retinal blindness through genetically engineered viruses for a company named Luxturna (Darrow 2019).

In plants, regulation of genetically engineered organisms (so-called "GMOs") has been a struggle in several parts of the world. Regulation of genome edited organisms has also been a confusing regulatory challenge (Strauss and Sax 2016). For example, the genome edited crop plants will not be considered as GMOs in the US but in Europe they were declared to be GMOs (Kim and Kim 2016; Wight 2018). The genome edited cows in the US are currently legally considered as a drug and are currently not allowed in the market (Carroll et al. 2016). Genome edited salmon are commercially available for consumers in Canada. For consumers in the US, FDA has regulated

these salmon, but they are still not available in the market due to new rules stipulating establishment of labeling requirements (Ledford 2019).

While regulation has been complicated for plant and animal applications, use of these tools in humans should merit even more stringent oversight. DTC-GT through GWAS should not be used as a diagnostic tool and the regulatory procedures for these technologies should be clearly outlined for both research and diagnostic purposes by FDA and other respective regulatory organizations. From the regulatory stories of GMO and gene edited organisms, it is clear that the same technology can be regulated differently among plants, animals and humans and in different countries. Further without proper communication, the laws governing these technologies will be more complex in the globalized market.

3.7 Quantitative genetics and society

When it comes to genetic studies, the public understanding of genetics varies in a wide range and usually it is considered poor. At this stage, expecting the public to understand quantitative genetics might sound like an ambitious goal. However, the recent rise in DTC-GT technologies leave us no choice but to communicate quantitative genetics to the public. Usage of quantitative genetics in plant and animal breeding has resulted in the abundance of food in the developed world, especially meat, dairy, and corn (Hayes 2013; Hallauer et al. 2010). Despite enjoying the benefits of quantitative genetics every day, the public is not aware of the science. This is because of the complex nature of the science and the number of pre-requisites that are necessary to understand the science along with several other reasons (Lynch and Walsh 1998). For example,

communicating uncertainty through probabilities in climate change to the public has been proven to be challenging (Morgan 2009 ; Fischhoff and Davis 2014). According to the Center For Disease Control and Prevention, explaining probabilities with words like "one in five adults suffer from mental illness" could help people understand probabilities better. In a similar manner, the public can be educated about quantitative genetics with educational programs that use language that targets the general audience instead of just graduate students and scientists.

One of the major challenges to communicate quantitative genetics to the public is the lack of trust that exist between geneticists and the general public. This is due to the eugenic origins of quantitative genetics. Popular scientists who first studied human genetics like Dr. Francis Galton and Dr. Charles Davenport were eugenicists (Mukherjee 2016; Zimmer 2018). Reestablishing the trust between scientists and the public, especially with people of color and minorities, is important for the public to fully appreciate the positive effects of DTC-GT and quantitative genetics.

In this era of lack of online privacy, it is hard to protect your personal health information from hackers. Even companies with massive resources to protect their data are known to be hacked (Morey et al. 2015). The risk of DTC-GT companies being hacked raises regulatory questions related to DNA privacy. Concerns related to identity theft, targeted marketing by pharmaceutical companies, discrimination in insurance or employment and even concerns related to creation of synthetic DNA have been raised by the public (Giovanni et al. 2010). Even though, DTC-GT results have been used by law enforcement to identify culprits who got away with their crimes, it is still unclear if using a family member's DNA to catch a culprit oversteps DNA privacy (Guerrini et al. 2018).

Table 3.1 *List of diseases that are allowed by the FDA to be screened by DTC-GT services.*

Serial No.	Condition	Description
1	Parkinson's disease	Nervous system disorder impacting movement
2	Late-onset Alzheimer's disease	Progressive brain disorder that destroys memory and thinking skills
3	Celiac disease	Disorder resulting in the inability to digest gluten
4	Alpha-1 antitrypsin deficiency	Disorder that raises the risk of lung and liver disease
5	Early-onset primary dystonia	Movement disorder involving involuntary muscle contractions and other uncontrolled movements
6	Factor XI deficiency	Blood clotting disorder
7	Gaucher disease type 1	Organ and tissue disorder
8	Glucose-6-phosphate dehydrogenase deficiency	Red blood cell condition
9	Hereditary hemochromatosis	Iron overload disorder
10	Hereditary thrombophilia	Blood clot disorder

The benefits of DTC-GT cannot be overlooked. Individuals who seek to have their DNA tested should be able to do so. The clinical and personal value depends on the trait and the individual.

Table 3.1 shows the list of diseases that the FDA has approved for the DTC-GT companies to diagnose. A good science communication program can enable the consumers to understand and appreciate their genome better.

Science miscommunication of genetics might result in political and civil unrest in some parts of the world. In certain parts of the world, the consequences of a person being identified to belong to a certain tribe, caste, sexuality or intelligence based on their genomic data might result in discrimination (Niemiec and Howard 2016).

3.8 Conclusion

The rise of DTC-GT companies has led to the reduction in price to have one's genome sequenced or tested. The public should be able to make informed decisions based on their genomic data. Science communication of quantitative genetics is essential for the public to understand their genomic data. However, communicating quantitative genetics to a diverse audience can be challenging due to the eugenic origins of the field in humans. The positive stories of quantitative genetics in agriculture can be used to communicate quantitative genetics to a diverse audience. Public learning of science comes from news, entertainment, advertising and social media. Communicating science to the public clearly without misinformation in this era is a complex process that is still being studied (Peterson 2001; Scheufele *et al.* 2019). Strategies to address science miscommunication can be developed specifically for quantitative genetics. Studying and targeting major misconceptions of the public when it comes to quantitively inherited diseases like

cancer will aid in developing communication strategies that help in narrowing down and addressing major public misconceptions.

Research grants for genetics projects can include a budget section for science communication projects (Motta 2019). Even though science communication will not happen over-night, scientists should strive for that ultimate goal. The US National Academies of Sciences, Engineering and Medicine have advised scientists to aid in credible knowledge synthesis and dissemination of information to providers and consumers (Fraker and Mazza 2011). This is possible only through a multi-faceted approach in which scientists in multiple fields such as quantitative geneticists, communication professionals, social scientists, genetic counsellors and primary care physicians need to follow an interdisciplinary approach.

GWAS has been used in plants, animals and human populations to study quantitative traits. Using GWAS to diagnose diseases, especially quantitatively inherited diseases, is considered unethical (Kaye et al. 2010). Clear regulation and oversight of use of GWAS to diagnose diseases should be outlined by regulatory agencies.

GWAS and genomic prediction studies employ a black box approach to select for a certain trait or to explain the variance of a certain trait. If these methods are used in diagnosis of traits, how they work should be more transparent. Science communication of quantitative genetics can play a major role in making this black box more transparent to the consumers. Geneticists are not usually trained to be effective science communicators. Programs that offer training to be effective science communicators for faculty can assist in the effective dissemination of quantitative genetics

information to the public. For example, the Alan Alda center for Communicating Science is known to train scientists in the art of communication (Kaldy, 2019).

The rapid development of DTC-GT can be a powerful resource to the public and scientific community in understanding and studying quantitative traits. Involving the public in the post GWAS analysis for quantitative traits through a virtual web-based platform is being experimented by scientists in Queensland University, Australia (Cuellar-Partida et al. 2019). Such efforts increase the transparency in the scientific process. Academics should develop robust communication strategies that have the potential to build trust between geneticists and the public. Only then the fruits of genomics can serve diverse communities from all walks of life.

References

Abraham, G., Havulinna, A. S., Bhalala, O. G., Byars, S. G., De Livera, A. M., Yetukuri, L., . . . Inouye, M. (2016). Genomic prediction of coronary heart disease. *European Heart Journal*, *37*(43), 3267-3278. doi:10.1093/eurheartj/ehw450

ACGC, (2019). *States Issuing Licenses for Genetic Counselors.* https://www.nsgc.org/p/cm/ld/fid=19

Anderson, G. W. (1996). The evolution and status of genetics education in nursing in the united states 1983-1995. *Journal of Nursing Scholarship, 28*(2), 101-106.

Aziza-Shuster, E. (1994). Ethics and society: a child at all costs: posthumous reproduction and the meaning of parenthood. *Human Reproduction*, *9*(11), 2182-2185.

Begum, F., Ghosh, D., Tseng, G. C., & Feingold, E. (2012). Comprehensive literature review and statistical considerations for GWAS meta-analysis. *Nucleic acids research*, *40*(9), 3777-3784.

Benward, J. (2015). Mandatory counseling for gamete donation recipients: ethical dilemmas. *Fertility and sterility*, *104*(3), 507-512.

Bensaude-Vincent, B. (2001). A genealogy of the increasing gap between science and the public. Public Understanding of Science, 10(1), 99-113.

Bernardo, R. (2002). *Breeding for quantitative traits in plants*(No. 576.5 B523). Stemma Press.

Berliner, D. C., & Glass, G. V. (Eds.). (2014). *50 myths and lies that threaten America's public schools: The real crisis in education*. Teachers College Press.

Bloss, C. S., Schork, N. J., & Topol, E. J. (2011). Effect of direct-to-consumer genomewide profiling to assess disease risk. *New England Journal of Medicine*, *364*(6), 524-534.

Brachi, B., Morris, G. P., & Borevitz, J. O. (2011). Genome-wide association studies in plants: the missing heritability is in the field. *Genome biology*, *12*(10), 232.

Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., . . . McMullen, M. D. (2009). The genetic architecture of maize flowering time. *Science, 325*(5941), 714-718. doi:10.1126/science.1174276

Carroll, D., Van Eenennaam, A. L., Taylor, J. F., Seger, J., & Voytas, D. F. (2016). Regulate genome-edited products, not genome editing itself. *Nature biotechnology, 34*(5), 477.

Charlsie, A., Luke, D., (2018). *Twins get some 'mystifying' results when they put 5 DNA ancestry kits to the test*

https://www.cbc.ca/news/technology/dna-ancestry-kits-twins-marketplace-1.4980976CBC News

Chen, G., Wang, X., Hao, J., Yan, J., & Ding, J. (2015). Genome-wide association implicates candidate genes conferring resistance to maize rough dwarf disease in maize. *Plos One, 10*(11), e0142001-e0142001. doi:10.1371/journal.pone.0142001

Crossa, J., Beyene, Y., Semagn, K., Pérez, P., Hickey, J. M., Chen, C., . . . Babu, R. (2013). Genomic prediction in maize breeding populations with genotyping-by-sequencing. *G3: Genes, Genomes, Genetics, 3*(11), 1903-1926. doi:10.1534/g3.113.008227

Cuellar-Partida, G., Lundberg, M., Kho, P. F., D'Urso, S., Gutierrez-Mondragon, L. F., & Hwang, L. D. (2019). Complex-Traits Genetics Virtual Lab: A community-driven web platform for post-GWAS analyses. bioRxiv, 518027.

Darrow, J. (2019). Luxturna: FDA documents reveal the value of a costly gene therapy. *Drug Discovery Today, 24*(4), 949-954. doi:10.1016/j.drudis.2019.01.019

Daack-Hirsch, S., Driessnack, M., Perkhounkova, Y., Furukawa, R., & Ramirez, A. (2012). A practical first step to integrating genetics into the curriculum. *Journal of Nursing Education, 51*(5), 294-298. doi:10.3928/01484834-20120309-02

Daetwyler, H. D., Calus, M. P., Pong-Wong, R., de los Campos, G., & Hickey, J. M. (2013). Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics*, *193*(2), 347-365.

Desjardins, A., Gromeier, M., Herndon, J. E., Beaubier, N., Bolognesi, D. P., Friedman, H. S., . . . Bigner, D. D. (2018). Recurrent glioblastoma treated with recombinant poliovirus. *The New England Journal of Medicine, 379*(2), 150-161. doi:10.1056/NEJMoa1716435

Dunn, E. C., Brown, R. C., Dai, Y., Rosand, J., Nugent, N. R., Amstadter, A. B., & Smoller, J. W. (2015). Genetic determinants of depression: recent findings and future directions. *Harvard review of psychiatry*, *23*(1), 1–18. doi:10.1097/HRP.0000000000000054

Edwards, S. M., Buntjer, J. B., Jackson, R., Bentley, A. R., Lage, J., Byrne, E., . . . Hickey, J. M. (2019). The effects of training population design on genomic prediction accuracy in wheat.*TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik,* , 1-10. doi:10.1007/s00122-019-03327-y

Fischhoff, B., & Davis, A. L. (2014). Communicating scientific uncertainty. *Proceedings of the National Academy of Sciences*, *111*(Supplement 4), 13664-13671.

Fraker, M., & Mazza, A.-M. (2011). *Direct-to-consumer genetic testing: summary of a workshop*. Washington, D.C.: National Academies Press.

Fujimura, J. H., & Rajagopalan, R. (2011). Different differences: The use of 'genetic ancestry'versus race in biomedical human genetic research. *Social Studies of Science*, *41*(1), 5-30.

Garvin, T. (2001). Analytical paradigms: the epistemological distances between scientists, policy makers, and the public. *Risk Analysis*, *21*(3), 443-456.

Genomeweb (2019) *Genomic Prediction Raises $4.5M.* https://www.genomeweb.com/sequencing/genomic-prediction-raises-45m#.XLOiHS-ZNhE

Giovanni, M. A., Fickie, M. R., Lehmann, L. S., Green, R. C., Meckley, L. M., Veenstra, D., & Murray, M. F. (2010). Health-care referrals from direct-to-consumer genetic testing. *Genetic Testing and Molecular Biomarkers*, *14*(6), 817-819.

Glazier AM, Nadeau JH, Aitman TJ (2002) Finding genes that underlie complex traits. Science 298: 2345–2349.

Guerrini, C. J., Robinson, J. O., Petersen, D., & McGuire, A. L. (2018). Should police have access to genetic genealogy databases? Capturing the Golden State Killer and other criminals using a controversial new forensic technique. *PLoS biology*, *16*(10), e2006906. doi:10.1371/journal.pbio.2006906

Hallauer, A. R., Carena, M. J., & Miranda Filho, J. D. (2010). *Quantitative genetics in maize breeding* (Vol. 6). Springer Science & Business Media.

Hayes, B. (2013). Overview of statistical methods for genome-wide association studies (GWAS). In *Genome-wide association studies and genomic prediction* (pp. 149-169). Humana Press, Totowa, NJ.

Hayes, B. J., Pryce, J., Chamberlain, A. J., Bowman, P. J., & Goddard, M. E. (2010). Genetic architecture of complex traits and accuracy of genomic prediction: Coat colour, milk-fat percentage, and type in holstein cattle as contrasting model traits. *PLoS Genetics, 6*(9), e1001139. doi:10.1371/journal.pgen.1001139

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS biology*, *13*(3), e1002106.

Hirschhorn JN, Daly MJ: Genome-wide association studies for common diseases and complex traits. Nat Rev Genet. 2005, 6 (2): 95-108.

Hock, K. T., Christensen, K. D., Yashar, B. M., Roberts, J. S., Gollust, S. E., & Uhlmann, W. R. (2011). Direct-to-consumer genetic testing: an assessment of genetic counselors' knowledge and beliefs. *Genetics in Medicine*, *13*(4), 325.

Kaldy, J. (2019). Performers Put Practitioner Communication in the Footlights. *Caring for the Ages*, *20*(2), 12.

Kampourakis, K. (2013). Mendel and the path to genetics: Portraying science as a social process. *Science & Education*, *22*(2), 293-324.

Kampourakis, K. (2017). Public Understanding of Genetic Testing and Obstacles to Genetics Literacy. In *Molecular Diagnostics* (pp. 469-477). Academic Press.

Kaye, J., Boddington, P., De Vries, J., Hawkins, N., & Melham, K. (2010). Ethical implications of the use of whole genome methods in medical research. *European Journal of Human Genetics*, *18*(4), 398-403. doi:10.1038/ejhg.2009.191

Kevles, D. J. (1995). *In the name of eugenics: Genetics and the uses of human heredity* (No. 95). Harvard University Press.

Kim, J., & Kim, J. (2016). Bypassing GMO regulations with CRISPR gene editing. *Nature Biotechnology*, *34*(10), 1014-1015. doi:10.1038/nbt.3680

Ledford, H. (2019). Gene-edited animal creators look beyond US market. *Nature*, *566*(7745), 433-434. doi:10.1038/d41586-019-00600-4

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits* (Vol. 1, pp. 535-557). Sunderland, MA: Sinauer.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. (2009) Finding the missing heritability of complex diseases. Nature 461: 747–753.

Marteau, T. M., French, D. P., Griffin, S. J., Prevost, A. T., Sutton, S., Watkinson, C., ... & Hollands, G. J. (2010). Effects of communicating DNA-based disease risk estimates on risk-reducing behaviours. *Cochrane Database of Systematic Reviews*, (10).

McCall, R. B. (1977). Childhood IQ's as predictors of adult educational and occupational status. *Science*, *197*(4302), 482-483.

Meagher, K. M., & Berg, J. S. (2018). Too much of a good thing? Overdiagnosis, or overestimating risk in preventive genomic screening.

Medina-Gomez, C., Felix, J. F., Estrada, K., Peters, M. J., Herrera, L., Kruithof, C. J., ... & Jaddoe, V. W. V. (2015). Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: the Generation R Study. *European journal of epidemiology*, *30*(4), 317-330.

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819-1829.

Molteni,M (2018) *Now you can sequence your whole genome for just $200.* https://www.wired.com/story/whole-genome-sequencing-cost-200-dollars/

Morota, G., Abdollahi-Arpanahi, R., Kranis, A., & Gianola, D. (2014). Genome-enabled prediction of quantitative traits in chickens using genomic annotation. *BMC Genomics, 15*(1), 109-109. doi:10.1186/1471-2164-15-109

Motta, M. (2019). Explaining science funding attitudes in the United States: The case for science interest. *Public Understanding of Science, 28*(2), 161-176.

Morey, T., Forbath, T., & Schoop, A. (2015). Customer data: Designing for transparency and trust. *Harvard Business Review, 93*(5), 96-105.

Morgan, M. G. (2009). *Best practice approaches for characterizing, communicating and incorporating scientific uncertainty in climate decision making*. DIANE publishing.

Mukherjee, S. (2016). *The gene: an intimate history.* (First Scribner hardcover edition.). New York: Scribner.

Nippert, I., Harris, H. J., Julian-Reynier, C., Kristoffersson, U., Leo, P., Anionwu, E., ... & Harris, R. (2011). Confidence of primary care physicians in their ability to carry out basic medical genetic tasks—a European survey in five countries—part 1. *Journal of community genetics, 2*(1), 1-11.

Niemiec, E., & Howard, H. C. (2016). Ethical issues in consumer genome sequencing: Use of consumers' samples and data. *Applied & translational genomics, 8*, 23-30.

OMIM, (2019). *OMIM Gene Map Statistics.* http://omim.org/statistics/geneMap

Goddard, M. E., & Hayes, B. J. (2007). Genomic selection. *Journal of Animal breeding and Genetics*, *124*(6), 323-330.

Ohnishi, Y., Yamada, R., Sato, H., Ozaki, K., Sato, H., Nakamura, Y., . . . Iida, A. (2002). Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. *Nature Genetics, 32*(4), 650-654. doi:10.1038/ng1047

Parrott, R., Kahl, M. L., Ndiaye, K., & Traeder, T. (2012). Health communication, genetic determinism, and perceived control: The roles of beliefs about susceptibility and severity versus disease essentialism. *Journal of health communication*, *17*(7), 762-778.

Peiffer, J. A., Romay, M. C., Gore, M. A., Flint-Garcia, S. A., Zhang, Z., Millard, M. J., . . . Buckler, E. S. (2014). The genetic architecture of maize height. *Genetics, 196*(4), 1337-1356. doi:10.1534/genetics.113.159152

Petersen, A. (2001). Biofantasies: genetics and medicine in the print news media. *Social Science & Medicine, 52*(8), 1255-1268.

Porter, D. L., Levine, B. L., Kalos, M., Bagg, A., & June, C. H. (2011). Chimeric antigen Receptor–Modified T cells in chronic lymphoid leukemia. *The New England Journal of Medicine, 365*(8), 725-733. doi:10.1056/NEJMoa1103849

Powell, K. P., Christianson, C. A., Cogswell, W. A., Dave, G., Verma, A., Eubanks, S., & Henrich, V. C. (2012). Educational needs of primary care physicians regarding direct-to-consumer genetic testing. *Journal of genetic counseling*, *21*(3), 469-478.

Prows, C. A., Glass, M., Nicol, M. J. (., Skirton, H., & Williams, J. (2005). Genomics in nursing education. *Journal of Nursing Scholarship*, *37*(3), 196-202. doi:10.1111/j.1547-5069.2005.00035.x

Sanders, A. R., Beecham, G. W., Guo, S., Dawood, K., Rieger, G., Badner, J. A., ... & Gejman, P. V. (2017). Genome-wide association study of male sexual orientation. *Scientific reports*, *7*(1), 16950.

Searle, S. R. (1991). CR Henderson, the statistician; and his contributions to variance components estimation. *Journal of dairy science*, *74*(11), 4035-4044.

Selzam S, Krapohl E, von Stumm S, O'Reilly PF, Rimfeld K, Kovas Y, Dale PS, Lee JJ, Plomin R (2016) Predicting educational achievement from DNA. *Mol Psychiat* 22: 267–272

Schaper, M., & Schicktanz, S. (2018). Medicine, market and communication: ethical considerations in regard to persuasive communication in direct-to-consumer genetic testing services. *BMC medical ethics*, *19*(1), 56. doi:10.1186/s12910-018-0292-3

Scheufele, D. A., & Krause, N. M. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences*, 201805871.

Schork, A. J., Thompson, W. K., Pham, P., Torkamani, A., Roddey, J. C., Sullivan, P. F., ... & Andreassen, O. A. (2013). All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS genetics*, *9*(4), e1003449.

Svrakic, D. M., & Cloninger, C. R. (2017). Genome-wide association study of male sexual orientation.

Shields, A. E., Fortun, M., Hammonds, E. M., King, P. A., Lerman, C., Rapp, R., & Sullivan, P. F. (2005). The use of race variables in genetic studies of complex traits and the goal of reducing health disparities: a transdisciplinary perspective. *American Psychologist*, *60*(1), 77.

Stanovich, K. E. (2009). *What intelligence tests miss: The psychology of rational thought*. Yale University Press.

Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., ... & Robinson, G. E. (2015). Big data: astronomical or genomical?. *PLoS biology*, *13*(7), e1002195.

Ramagopalan, S. V., Dyment, D. A., Handunnetthi, L., Rice, G. P., & Ebers, G. C. (2010). A genome-wide scan of male sexual orientation. *Journal of Human Genetics*, *55*(2), 131.

Rafiq, M., Ianuale, C., Ricciardi, W., & Boccia, S. (2015). Direct-to-consumer genetic testing: a systematic review of European guidelines, recommendations, and position statements. *Genetic testing and molecular biomarkers*, *19*(10), 535-547.

Ramos, E., & Weissman, S. M. (2018, March). The dawn of consumer-directed testing. In *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* (Vol. 178, No. 1, pp. 89-97).

Richardson, K. (2017). GWAS and cognitive abilities: Why correlations are inevitable and meaningless: GWA studies to identify genetic factors for educational achievements largely ignore underlying social structures and dynamics. *EMBO reports*, *18*(8), 1279-1283.

Strauss, S. H., & Sax, J. K. (2016). Ending event-based regulation of GMO crops. *Nature Biotechnology*, *34*(5), 474-477. doi:10.1038/nbt.3541

Wang, H., & Qin, F. (2017). Genome-wide association study reveals natural variations contributing to drought resistance in crops.*Frontiers in Plant Science*, *8*, 1110. doi:10.3389/fpls.2017.01110

Wang, Y., Wang, Y., Wang, X., & Deng, D. (2019). Integrated meta-QTL and genome-wide association study analyses reveal candidate genes for maize yield. *Journal of Plant Growth Regulation*, , 1-10. doi:10.1007/s00344-019-09977-y

Wight, A. J. (2018). Strict EU ruling on gene-edited crops squeezes science. *Nature*, *563*, 15-16.

Windhausen, V. S., Atlin, G. N., Hickey, J. M., Crossa, J., Jannink, J., Sorrells, M. E., . . . Melchinger, A. E. (2012). Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. *G3 (Bethesda, Md.), 2*(11), 1427-1436. doi:10.1534/g3.112.003699

Wolinsky, H. (2007). The thousand-dollar genome: Genetic brinkmanship or personalized medicine?. *EMBO reports*, *8*(10), 900-903.

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., ... & Bacanu, S. A. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature genetics*, *50*(5), 668.

Wright, C. F., & Gregory-Jones, S. (2010). Size of the direct-to-consumer genomic testing market. *Genetics in Medicine, 12*(9), 594.

Yagci, A. H. (2019;2018;). Policy knowledge, collective action and advocacy coalitions: Regulating GMOs in turkey. *Journal of European Public Policy, 26*(6), 927-945. doi:10.1080/13501763.2018.1509884

Zimmer, C. (2018). *She has her mother's laugh: the powers, perversions, and potential of heredity.*New York, New York: Dutton, an imprint of Penguin Random House LLC.

CHAPTER 4

SUMMARY

Genomic technologies are increasingly used in human, plant and animal biology to understand the genetic basis of complex traits. We aimed to understand the genetic basis of multiple disease resistance demonstrated by maize inbred NY22613. We developed a BC_3S_3 biparental mapping population titled the OHNY mapping population using the resistant inbred NY22613 and the susceptible inbred Oh7B. We used the mapping population to identify disease resistance QTL for northern leaf blight (NLB), Stewart's wilt (SW) and gray leaf spot (GLS). Our QTL mapping analysis identified that 16 QTL were associated with NLB resistance, 17 QTL were associated with GLS resistance and 16 QTL were associated with SW resistance. No QTL were colocalized for all three diseases. In our study, three QTL were shared for NLB and GLS and one QTL was shared for GLS and SW.

In chromosome 2, the colocalized GLS and SW QTL is located between markers S2_4211875 and S2_4257983. This region includes the *liguleless1* gene that encodes a nuclear-localized protein that is needed for induction of ligules and auricles during maize leaf organogenesis. In addition, *liguleless1* works with *liguleless2* and might be associated with the salicylic acid pathway. In chromosome 7, the colocalized GLS and NLB QTL is located between markers S7_100035505 and S7_100170373. In chromosome 9, the colocalized GLS and NLB QTL is located between markers S9_8778978 and S9_8974937. In chromosome 10, the colocalized GLS and NLB QTL is located between markers S10_36259906 and S10_36380814. This region has the *Auxin/Indole-3-Acetic Acid* (*Aux/IAA)-transcription factor 42* gene. The function of this gene is previously

unknown in maize. However, the class of transcription factors has been associated with the salicylic acid pathway and might be involved in the activation of R genes.

Selecting for families that are resistant to multiple diseases for breeding purposes using only phenotypic data can be challenging. We demonstrate a new method titled High dEnsity Marker Phenotype (HEMP) QTL selection, in which we use phenotypic data along with SNP marker data and QTL data to select for disease resistant families. We show that we were not able to select for even a single family that was resistant for NLB, GLS and SW using phenotypic data alone. However, by using the HEMP QTL selection strategy, we were able to select 23 families that might have alleles with the potential to improve multiple disease resistance.

A series of differential expression studies were conducted between the resistant inbred NY22613 and the susceptible inbred Oh7B under both greenhouse and field conditions. The differential expression analysis study revealed that in greenhouses, 7082 genes were downregulated in NY22613 relative to Oh7B, 6842 genes were upregulated, and 8477 genes were showed no differences in expression level between these two inbreds. In field conditions, 5134 genes were downregulated in NY22613 relative to Oh7B, 5250 genes were upregulated and 12704 genes did not differ significantly in expression between the two inbreds.

When inoculated with *S. turcica*, six genes were shown to be differentially expressed in both field and greenhouse conditions in NY22613. *Zm00001d027691* and *Zm00001d034421* located on chromosome 1, *Zm00001d011152* and *Zm00001d008951* located on chromosome 8 and *Zm00001d021770* located on chromosome 7 have no previously known function.

Zm00001d033623, also known as *lipoxygenase3*, has no previously known function in maize, however it is known to play a major role in disease resistance in rice.

From the QTL analysis and the expression analysis, it was found that the *liguleless1* gene is an interesting candidate gene for multiple disease resistance. This gene is present on chromosome 2 between the markers S2_4211875 and S2_4257983. This region has a QTL that is colocalized for GLS and SW disease resistance. In the differential expression study that involved inoculated susceptible inbred Oh7B in both field and greenhouse conditions, *liguleless1* was associated with disease resistance for NLB.

The discussion of science communication of quantitative genetics might seem a bit distant from the first two chapters. However, working on projects that dealt with big data such as RNASeq and QTL analyses made me realize the importance of quantitative genetics to interpret biological data sets. Currently, due to reduced genome sequencing costs, direct-to-consumer genomic technology companies are using human genome sequencing to predict ancestry and disease probabilities. If these predictions are misinterpreted, it could lead the public to make major life changing decisions that might not be justified. This also raises questions on ethics, regulation and science communication. The third chapter details the importance of science communication of quantitative genetics and stresses the importance of participation by plant and animal geneticists in the regulation and communication of quantitative genetics.