USING NEXT-GENERATION SEQUENCING TECHNOLOGIES TO DEVELOP NEW MOLECULAR MARKERS FOR THE LEAF RUST RESISTANCE GENE LR16

by

NICOLE REZAC HARRISON

B.S., Kansas State University, 2010

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Plant Pathology
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2014

Approved by:

Co-Major Professor
John P. Fellers

Approved by:

Co-Major Professor
Allan K. Fritz

Abstract

Leaf rust is caused by *Puccinia triticina* and is one of the most widespread diseases of wheat worldwide. Breeding for resistance is one of the most effective methods of control. *Lr16* is a leaf rust resistance gene that provides partial resistance at the seedling stage. One objective of this study was to use RNA-seq and *in silico* subtraction to develop new resistance gene analog (RGA) markers linked to *Lr16*. RNA was isolated from the susceptible wheat cultivar Thatcher (Tc) and the resistant Thatcher isolines Tc*Lr10*, Tc*Lr16*, and Tc*Lr21*. Using *in silico* subtraction, Tc isoline ESTs that did not align to the Tc reference were assembled into contigs and analyzed using BLAST. Primers were designed from 137 resistance gene analog sequences not found in Tc. A population of 260 F_2 lines derived from a cross between the rust-susceptible cultivar Chinese Spring (CS) and a Thatcher isoline containing *Lr16* (Tc*Lr16*) was developed for mapping these markers. Two RGA markers *XRGA266585* and *XRGA22128* were identified that mapped 1.1 cM and 23.8 cM from *Lr16,* respectively. Three SSR markers *Xwmc764, Xwmc661,* and *Xbarc35* mapped between these two RGA markers at distances of 4.1 cM, 10.7 cM, and 16.1 cM from *Lr16*, respectively. Another objective of this study was to use genotyping-by-sequencing (GBS) to develop single nucleotide polymorphism (SNP) markers closely linked to *Lr16*. DNA from 22 resistant and 22 susceptible F_2 plants from a cross between CS and Tc*Lr16* was used for GBS analysis. A total of 39 Kompetitive Allele Specific PCR (KASP) markers were designed from SNPs identified using the UNEAK and Tassel pipelines. The KASP marker *XSNP16_TP1456* mapped 0.7 cM proximal to *Lr16* in a TcxTc*Lr16* population consisting of 129 F_2 plants. These results indicate that both techniques are viable methods to develop new molecular markers. RNA-seq and *in silico* subtraction were successfully used to develop two new RGA markers linked to *Lr16*, one of which was more

closely linked than known SSR markers. GBS was also successfully used on an F_2 population to develop a KASP marker that is the most closely linked marker to *Lr16* to date.

Table of Contents

List of Figures

Acknowledgements

This thesis would not have been possible without the guidance of my committee members, help from friends, and the support of my family and husband. I would like to thank my advisor John Fellers and my committee members Barbara Valent and Allan Fritz for all of their support and guidance throughout my project. I am also grateful for my fellow lab members and the assistance they have given me during my time working with them. Other people I would like to thank include Alina Akhunova and Paul St. Amand for their help with the KASP assays, Jesse Poland and Shuangye Wu for their help with the genotyping-by-sequencing, and Craig Webb for his assistance in numerous ways. Last but not least, I want to express my gratitude to my husband and family for their unending love and support through the good times and bad.

Chapter 1 - Review of Literature

Wheat

Wheat is the one of the world's most important food crops and is the staple food for about 35% of the population (IDRC, 2013). On average, 2.16 billion bushels of wheat are harvested annually (USDA NASS, 2013). Over 25,000 varieties of wheat are adapted to a wide range of temperate environments and contribute to its success (Shewry, 2009). Currently, about 95% of the wheat grown worldwide is hexaploid bread wheat. The majority of the remaining 5% is tetraploid durum wheat that is commonly used to make pasta (Shewry, 2009). China is the greatest producer of wheat in the world producing 120,580,320 tonnes of wheat in 2012 (FAOSTAT, 2014). China typically plants 23.7 millihectares (mha) of wheat and produces 109.3 million tones annually, and these numbers are continuously increasing (Huerta-Espino et al., 2011). In the United States, 56,156,000 acres of wheat were planted in 2013, which yielded 2,129,695,000 bushels (USDA NASS, 2013).

Wheat is an annual plant that can be classified as either spring or winter, depending on its flowering habit (Simmons et al., 2013). Winter wheat contributed to 43,090,000 of the acres planted in the United States in 2013 (USDA NASS, 2013). Wheat can further be broken down into six major classes in the United States: hard red winter (HRW), hard red spring (HRS), soft red winter (SRW), hard white, soft white, and durum (Vocke and Liefert, 2013; Kansas Wheat, 2013). Varieties are categorized into the different classes based on their hardness, color of the kernels, and vernalization requirement (Kansas Wheat, 2013). In the United States, hard red winter wheat is the predominant market class grown accounting for 29.4 million acres (Minnesota Association of Wheat Growers, 1994; USDA NASS, 2013). HRW wheat is known

for its good milling and baking characteristics and is used to produce mainly bread and rolls. HRW wheat is also used to produce some sweet goods and all-purpose flour (Minnesota Association of Wheat Growers, 1994). The other winter wheats, soft red winter and white winter, accounted for 9.96 million acres and 3.38 million acres of wheat planted, respectively. Kansas alone produced 319 million bushels of winter wheat over 9.5 million acres (USDA NASS, 2013). Hard red spring wheat has the highest production in the United States, accounting for 20%. HRS wheat is known for its high protein levels and is primarily used to make specialty breads and to blend with lower protein wheat (Vocke and Liefert, 2013).

Wheat was domesticated in the Fertile Crescent between 7700 and 7500 B.C. (Nesbitt and Samuel, 1998). Emmer is considered the first domesticated wheat and has a uniform flowering time, a nonbrittle rachis, smaller kernels than wild emmer, and lacks grain dormancy. Modern durum wheat, which is also tetraploid, gradually took the place of domesticated emmer due to the acquisition of free threshing (Haudry et al., 2007). Emmer wheat was formed from the hybridization of *T. urartu*, which contributed the A genome, and a close relative *Aegilops speltoides* contributed the B genome (Brenchley et al., 2012). Modern bread wheat is an allohexaploid derived from hybridization of three different *Aegilops* and *Triticum* species (Salentijn et al., 2009). The tetraploid species *T. turgidum ssp. dicoccoides* (AABB) was one the progenitors of bread wheat and is more commonly referred to as cultivated emmer wheat. Some studies suggest that the B genome may also have originated from an introgression of several parental species also related to *Aegilops speltoides* (Huang et al., 2002). *Aegilops tauschii* contributed the DD genome of bread wheat after hybridization with *T. dicoccoides* only 8,000 years ago (Brenchley et al., 2012). Sequence analysis from two accessions in the D genome of

Ae. tauschii to D genome gene orthologs in *T. aestivum* confirmed their sequences are very closely related (Huang et al., 2002).

Common wheat (*Triticum aestivum* L.), also known as bread wheat, is just one of nineteen species in the *Triticum* genera. *Triticum* is a member of the family *Poaceae*, along with rice (genus *Oryza*) and maize (genus *Zea*; Huang et al., 2002). The wheat genome is 17 Gb, which is almost five times as large as the human genome, and 40 times larger than the rice (*Oryza sativa*) genome (Brydon, 2012; Choulet et al., 2010). The wheat genome contains an estimated 95,000 genes (Brydon, 2012). The A genome is estimated to contain 28,000 genes, while 38,000 genes are estimated to be in the B genome, and 36,000 genes are the in *Ae. tauschii* D genome (Brenchley et al., 2012). It had previously been estimated that about 80% of the genome is actually repeat sequences (Smith and Flavell, 1975). More recently, Li et al. (2004) found that the D genome contains close to 92% repeated sequences. Paux et al. (2006) also revealed that chromosome 3B of wheat has a repeat content of 86%. Due to this large genome size and high repetitive sequence content, sequencing the wheat genome has proven to be difficult (Choulet et al., 2010).

Leaf Rust

Leaf rust is caused by the fungus *Puccinia triticina* Eriks. and is the most widespread worldwide. The other two rusts, stem rust of wheat (*P. graminis* f. sp. *tritici*) and stripe rust (*P. striiformis* f. sp. *tritici*), are also significant but are more localized (Steffenson et al., 2007). *Triticum* species, including durum, emmer, and *triticale* spp., are the primary hosts for leaf rust. However, leaf rust can also be found on some *Aegilops* spp., *Hordeum maritimum*, and *Secale cereale* (Neu et al., 2003). Leaf rust has been recognized as a significant disease because it can cause yield losses ranging from 1 to 20% nationwide in wheat (Kolmer, 2013). In China, leaf

rust typically occurs on 15 mha of the 23.7 mha of wheat grown every year with infection levels in fields of 10-30% and reaching as high as 60% in some instances (Huerta-Espino et al., 2011). In Kansas, losses have been as high as 13.9% in 2007 (Appel et al., 2012). The losses in Kansas from 2009 to 2013 have been lower than the 20-year average of 2.3%. In 2013, the lowest losses due to leaf rust since 1980 were recorded with only 0.01% loss were attributed to drought and freeze damage (Appel et al., 2013). Genetic resistance and the use of fungicides are effective methods to control the disease (Kolmer et al., 2011).

Leaf rust is an obligate biotrophic, heteroecious fungus that requires two different hosts to complete its life cycle. These two hosts are a telial/uredinial host and an alternative pycnial/aecial host. Wheat is usually the primary host, and *Thalictrum speciosissimum* or *Isopyrum fumaroides* is the alternate host on which the sexual stage of the life cycle is completed (Bolton et al., 2008). However, the alternate host is not found in North America, so adaptations can only arise from genetic mutations due to the lack of sexual reproduction. Even so, worldwide populations of leaf rust are highly diverse, and 50 to 60 new races of leaf rust are identified every year in the United States (Kolmer, 2013). Leaf rust is a macrocyclic fungus with five distinct spore types. Teliospores, basidiospores, and urediniospores infect the cereal hosts, and pycniospores and aeciospores infect *Thalictrum*. Due to a lack of an alternate host, only uredinial infections occur in North America. Urediniospores are capable of continuously cycling on the host and can survive on any wheat leaves that remain green throughout the winter (Kolmer, 2013). Uredospores on overseasoning mycelium may also act as a source of inoculum (Roelfs, 1989). The "Puccinia Pathway" allows spores to overwinter on winter wheat cultivars in the warmer, Southern states, such as Texas. Spores are wind dispersed as far north as Canada as temperatures become warmer. These traits, in combination with the near-contiguous area of

wheat production in the central Great Plains, give the fungi the ability to induce serious epidemics (Eversmeyer and Kramer, 2000).

Erumpent, red-brown pustules on the upper leaf surface that are circular or ovoid in shape are indicative of leaf rust (Bockus et al., 2010). In severe cases, pustules may appear on the leaf sheath and glumes as well. These urediniospores are usually scattered across the leaf surface versus the more striped symptoms associated with stripe rust (*P. striiformis*; Clark et al., 2009). As the disease progresses, black teliospores may also be produced, but this is not as common with leaf rust as with stem rust (Singh and Saari, 1992). The urediniospores germinate upon contact with free water (Bockus et al., 2010). Warmer temperatures between 15 and 25°C and adequate moisture favor leaf rust infection (Bockus, 2010; Roelfs, 1985).

Leaf Rust Resistance

Wheat varieties resistant to leaf rust are one of the most profitable and environmentally friendly methods of control (Krattinger et al., 2009). Breeding for resistance is one of the priorities in many breeding programs. The use of resistant varieties puts pressure on the pathogen population to adapt and overcome the resistance (Kolmer, 2013). Development of early maturing wheat varieties has also helped reduced the impact of cereal rusts by allowing less time for epidemics to develop (Eversmeyer and Kramer, 2000). New sources of effective leaf rust resistance must be continuously added into breeding programs due to the ability of rust to mutate and multiply rapidly. The use of slow-rusting resistance genes and combinations of resistance genes can help prolong the effectiveness of the resistance cultivars (Singh et al., 2005). Combining seedling resistance genes and adult plant resistance genes into single cultivars may also aid in maintaining resistance (Dyck and Kerber, 1985).

Leaf rust is one of many pathogens that infect plants. Both monocotyledonous and dicotyledonous plants respond in a similar way to infection by pathogens by inducing resistance. There are two types of resistance, effector-triggered immunity (ETI) and pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI; Tsuda and Katagiri et al., 2010). PAMP-triggered immunity is the first line of defense against pathogens in plants and involves immune responses in the plant triggered via host receptor-mediated recognition of evolutionarily conserved patterns of pathogens (Zipfel, 2009). Many pathogens have the ability to overcome PTI by translocating effector proteins inside host cells to suppress defense responses and enhance virulence (Göhre and Robatzek, 2008). Some of these effectors are termed avirulence (*Avr*) proteins (Lokossou et al., 2009). ETI is activated when receptors encoded by a resistance (*R*) gene in the plant recognize ligands produced by the corresponding *Avr* protein (Lokossou et al., 2009; Dinesh-Kumar et al., 2000). Typically, the ETI response involves a hypersensitive response (HR), which is localized cell death at the infection site to contain the spread of the pathogen (Jones and Dangl, 2006). ETI represents a gene-for-gene relationship, where cultivars carrying a single dominant *R* gene demonstrate an incompatible reaction with the product of the corresponding dominant avirulence gene in the pathogen (Flor, 1971). Gene-for-gene specificity can be used to hypothesize which resistance genes are present in a cultivar by testing the material with a collection of *P. triticina* strains with known virulence/avirulence (Kolmer, 1996).

Traditional plant breeding relies on the analysis of visual traits and the use of effective pedigrees to incorporate resistance. As a result, many wheat cultivars are developed for production without knowing which resistance genes they contain (Collard and Mackill, 2008). Knowing which resistance genes are present can increase efficiency of breeding programs and help diversify leaf rust resistance through the use of targeted, effective leaf rust resistance genes

(Kolmer and Liu, 2002). This will also allow for gene pyramiding and better incorporation of slow-rusting genes (Vanzetti et al., 2011). Currently, differential rust cultures are used to identify the resistance genes present in cultivars. Multiple leaf rust isolates are selected based either on their importance in leaf rust populations or because they are known to have low infection types to specific leaf rust resistance genes. Low infection types indicate that a gene-for-gene reaction is occurring between the *R* gene of the cultivar and the *Avr* gene of the pathogen (Oelke and Kolmer, 2004). The cultivars with unknown resistance genes are inoculated with the selected leaf rust isolates and screened alongside Thatcher and Thatcher near-isogenic lines with single resistance genes. Seedling and adult plants are then evaluated based on their infection type responses (Oelke and Kolmer, 2004).

The number of isogenic lines and genes evaluated has changed over time. Dyck and Samborski originally backcrossed lines containing single genes into a Thatcher background (1968). Their initial lines contained the genes *Lr1, Lr2a, Lr3, Lr3ka, Lr14a, Lr14b,* and *Lr14ab*. Shortly after they added lines with the genes *Lr15, Lr23*, and three unidentified genes (Samborski and Dyck, 1976). A Thatcher background was chosen because resistance genes are strongly expressed (Long and Kolmer, 1989). At one point in the United States and Canada, the Thatcher differential set consisted of twelve near-isogenic lines (NILs) containing the genes *Lr1, Lr2a, Lr2c, Lr3, Lr9, Lr16, Lr24, Lr26, Lr3ka, Lr11, Lr17,* and *Lr30* (Huerta-Espino et al., 2011). Currently, 20 near-isogenic lines containing single resistance genes are used to screen for new races of leaf rust (USDA-ARS, 2008). The resistance genes included are grouped into five host sets, each with four different near-isogenic lines (NILs). Host set one contains lines with *Lr1*, *Lr2a*, *Lr2c*, and *Lr3*. Host set two contains lines *Lr9, Lr16, Lr24*, and *Lr26*. Host set three contains *Lr3ka, Lr11, Lr17*, and *Lr30*. Host set four contains *LrB, Lr10, Lr14a*, and *Lr18*. Host

set 5 contains *Lr21, Lr28, Lr41*, and *Lr42* (USDA-ARS, 2008). The isolines are inoculated with leaf rust and then evaluated based on their phenotypic reaction and infection type (McIntosh et al., 1995). Leaf rust races are assigned letters based on their reaction to each gene (Long and Kolmer, 1989).

Currently, 91 leaf rust resistance genes have been given gene designations (McIntosh et al., 2014). Most resistance genes give a moderate level of resistance, are effective from the seedling through the adult plant stage, and condition race-specific resistance (Dyck and Kerber, 1985). Race-specific resistance is more easily overcome than that of genes conferring long-term, non-race specific resistance (Kolmer, 2013).

Durable leaf rust resistance does not follow the classic gene-for-gene model. *Lr34* and *Lr46* were two of the first slow-rusting genes identified for leaf rust in wheat (Rubiales and Niks, 1995; Martinez et al., 2001). Neither specifies race-specific resistance and instead conditions broad-spectrum, non-hypersensitive response resistance. Infected cultivars with these genes will have fewer or smaller uredinia with either no or varying amounts of chlorosis and necrosis in adult plants (Martinez et al., 2001). Cultivars containing *Lr34* also develop a necrotic leaf tip, referred to as leaf tip necrosis, as a resistance response (Krattinger et al., 2009). More recently, the gene *Lr67/Yr46* and *Lr68* were also designated as a slow-rusting genes (Herrera-Foessel et al., 2011; Herrera-Foessel et al., 2012).

Molecular Markers

Molecular markers can aid in the process of incorporating desired resistance genes into new wheat varieties. Ultimately genes would be cloned and sequenced and perfect PCR markers would be developed for the gene. However, most genes have not been isolated and closely linked markers are the most useful (Feuillet et al., 2003). Restriction fragment length

polymorphism (RFLP) markers were originally used for linkage mapping (Tanksley et al., 1989). One major advantage of RFLPs is that they are codominant, so heterozygotes can be distinguished. RFLPs utilize restriction enzymes to digest DNA, and good quality DNA is required. The digested fragments are separated by gel electrophoresis and blotted onto a filter. This process is very time consuming and difficult to automate. The final step involves hybridization of probes to the DNA (Karp et al., 1996). Due to <10% of all RFLP loci in wheat being polymorphic, gene mapping progressed relatively slowly (Röder et al., 1998). Recombinant inbred substitution lines were often used to increase the level of polymorphism and mapping focused on particular chromosomes (Cadalen et al., 1996). After the development of the polymerase chain reaction (PCR), PCR-based methods became more predominant (Karp et al., 1996). Amplified fragment length polymorphism (AFLP) combines the use of restriction enzymes with the use of PCR amplification to identify restriction fragments. Stringent reaction conditions for primer annealing make the AFLP technique reliable. Restriction fragments are analyzed on high-resolution acrylamide gels (Vos et al., 1995). Random amplified polymorphic DNA (RAPD) markers are also PCR based. RAPDs are dominant markers and use randomly constructed oligonucleotides as primers. Amplified products are analyzed under ultraviolet light on agarose gels containing ethidium bromide (Lynch and Milligan, 1994). Like RFLPs, the low levels of polymorphism in wheat have hindered the success of AFLPs and RAPDs for use in marker-assisted selection (Song et al., 2005).

More recent methods have focused on the use of microsatellites or simple sequence repeats (SSRs; Röder et al., 1998). The initial development of microsatellites is very time-consuming and expensive but has many advantages (Song et al., 2005). Microsatellites are mostly genome specific, have high levels of polymorphism, and are codominant (Röder et al.,

1995). Moreover, Morgante el al. (2002) demonstrated the association of microsatellites with transcribed regions and discovered the frequency of microsatellites was significantly higher in ESTs than in genomic DNA. Röder et al., (1998) first reported a procedure optimized for development of microsatellites in wheat that used the methylation-sensitive restriction enzyme *Pst*I to isolate microsatellite clones from the hypomethylated regions of the wheat genome and reduce the number of microsatellite clones derived from repeated DNA. A total of 279 loci were mapped using the ITMI population, and 80% of the primer sets developed were genome-specific. Their lab was located in Gatersleben, Germany, and the microsatellites they mapped were thus designated *Xgwm* for Gatersleben wheat microsatellite. Because the initial development of SSRs is costly and time consuming, an International Wheat Microsatellites Mapping Network (IWMMN) was created to map additional microsatellite primers on the bread wheat International Triticeae Mapping Initiative (ITMI) map (Gupta et al., 2002). The majority of the 66 new loci mapped by the IWMMN were the WMC (Wheat Microsatellite Consortium) primers (Gupta et al., 2002). The Wheat Microsatellite Consortium was an international collaboration originated and directed by Dr. Peter Isaac (IDnagenetics, Norwich, UK) in order to develop new wheat microsatellites (Isaac, 2004). The Beltsville Agricultural Research Center also developed and mapped microsatellite markers that were given the prefix BARC (Song et al., 2002). Mapping of the *Xbarc* loci indicated that a good number of them were in close proximity to genes and that SSRs may be useful for investigating gene-rich regions of the wheat genome (Song et al., 2005). Somers et al. (2004) mapped the GWM, WMC, and BARC primers on four different bread wheat populations to develop a microsatellite consensus map.

Lr16

Lr16 is typically found in hard red spring wheat cultivars grown in the northern Great Plains, consisting of Minnesota, North Dakota, and South Dakota, as well as in Canada. *Lr16* is not highly effective when present alone but has provided moderate to high levels of resistance in hard red spring wheat cultivars for over 30 years when used in combinations with *Lr13*, *Lr23*, and *Lr34* (McIntosh et al., 1995; Kolmer, 2013). In Canada, the frequency of virulence on Thatcher lines containing *Lr16* has decreased greatly from 58.8% in 2001 to 0.6% in 2009 and reached as low as 0.3% in 2006 (Huerta-Espino et al., 2011). *Lr16* has been demonstrated to be more effective at higher temperatures than other genes and especially shows a greater necrotic response (Dyck and Johnson, 1983). The phenotypic response in wheat cultivars containing *Lr16* consists of small to moderate-sized uredinia surrounded by a chlorotic and/or necrotic halo (Long and Kolmer, 1989; McIntosh et al., 1995). One unique feature of *Lr16* is that resistance may be initially overcome by virulent leaf rust phenotypes. However, after lines containing *Lr16* are removed from cultivation for a period of time and then redeployed, the effectiveness of *Lr16* is often restored. This suggests a possible fitness cost for virulent *P. triticina* phenotypes. This same phenomenon has also been noted for *Lr9* and *Lr24* (Kolmer et al., 2009).

Lr16 originated in common wheat, and one source traces back to the Australian cultivar, "Warden" (Quartz/Ward's White//Red Birdeaux; McIntosh et al., 1995). Warden was used in a cross with Hybrid English to produce the variety Exchange, and Exchange was used in the cross for Selkirk. Hence, Exchange and Selkirk have also been found to have *Lr16* (Anderson, 1961). Etoile de Choise and Columbus are the final two original sources of *Lr16* germplasm (Bartos et al., 1969; Samborski and Dyck, 1982). *Lr16* was originally assigned to chromosome 4A using a cross between Tc*Lr16* and the Rescue monosomic series (Dyck and Kerber, 1971). It was later discovered that Rescue and Chinese spring differ by a 2B-4A translocation, and *Lr16* was

reassigned to the chromosome 2BS (McIntosh et al., 1995). In three different populations of recombinant inbred or doubled haploid lines, seven microsatellite markers have been mapped within 10 cM of *Lr16*. The closest marker was *Xwmc764* and mapped 1, 3, and 9 cM from *Lr16*, depending on the population. *Xgwm210*, *Xgwm614*, *Xwmc382*, *Xwmc661*, *Xwmc489*, *Xbarc45,* and *Xbarc35* were the other markers that mapped within 10 cM in all three tested populations (McCartney et al., 2005). The current consensus map also places *Xwmc764* nearest to *Lr16*. In order of their respective mapping distances from *Lr16*, *Xbarc45* and *Xbarc35*, *Xwmc661*, *Xgwm210* and *Xwmc489*, and *Xwmc382* are the other SSRs within 10 cM of *Lr16*. In order of nearest to *Lr16* to farther from *Lr16*, the other markers closely linked and within 10cM of *Lr16* are *Xbarc45* and *Xbarc35*, *Xwmc661*, *Xgwm210* and *Xwmc489*, and *Xwmc382* (Somers et al., 2004). Crossa et al., (2007) more recently discovered a DArT marker closely linked to *Xwmc661* called *Xwpt0100*. The stem rust resistance gene *Sr36* and leaf rust resistance gene *Lr13* are both proximally located on chromosome 2BS as well but are genetically independent (McIntosh et al., 1995). However, *Lr16* is always associated with *Sr23* (McIntosh and Luig, 1973). Altieri et al. (2008) developed a genetic linkage map using AFLP and SSR markers to characterize the resistance genes in the Argentinean cultivar Buck Manantial. They were able to confirm the presence of *Lr16* in the cultivar and map an unknown adult plant resistance gene named BMP1 within 1.3 cM of *Lr16*. The order of the markers used in this study was also highly similar to that of the consensus map by Somers et al. (2004). *Lr16* was also placed in *cis* configuration with *Sm1*, which confers resistance to orange wheat blossom midge (*Sitodiplosis mosellana* Gehin). The two genes were found to have different alleles at their *Xbarc35* locus and to have two recombinants between the two genes (Thomas et al., 2005).

Cloned Resistance Genes

To date, eight classes of plant disease resistance genes have been identified (Gururani et al., 2012). The majority of resistance genes belong to the classes containing both a nucleotide-binding site (NBS) and leucine rich repeat (LRR) domain. NBS-LRRs can further be divided into two groups based on their N-terminus. The first group includes *RPS2* of *Arabidopsis thaliana* and encodes a coiled-coil (CC) domain (Bent et al., 1994). Members of a second group encode a toll-interleukin-1-receptor (TIR) at their N-terminus. The *N* gene for *Tobacco mosaic virus* resistance and *L6* for flax rust resistance were two of genes first noted in this class (Lawrence et al., 1995). Whitham et al. (1994) first noted the sequence similarity of the N-terminus of the *N* gene of tobacco to the Drosophila Toll protein. When *RRS1* from *Arabidopsis thaliana* was cloned, these same conserved TIR-NBS-LRR domains were identified. However, a WRKY domain characteristic of transcription factors was also identified in its C-terminal, giving *RRS1* a novel structure. *RRS1* is also unique because it does not condition a hypersensitive response and because the resistant allele is recessive (Deslandes et al., 2002). The tomato *Ve2* gene for Verticillium wilt also has a novel structure with extracellular LRRs (eLRRs) and a PEST motif typically observed in rapidly degraded proteins (Fradin et al., 2009). The genes for resistance to *Cladisporium fulvum* from tomato originally comprised a group of resistance genes encoding eLRRs attached to C-terminal membrane anchor (Dixon et al., 1996; Thomas et al., 1997). Resistance genes at the *Vf* locus from apple have also been shown to have homologies to the *Cf* resistance genes (Vinatzer et al., 2001). *Xa21* from rice shares some characteristics with the *Cf* class of resistance genes, such as the eLRR and transmembrane domains. *Xa21* also encodes a serine-threonine kinase-like domain, and it is classified as a receptor kinase-like protein (Song et al., 1995). *Pto* also encodes a serine-threonine kinase but no LRRs and confers resistance to *Pseudomonas syringae* in tomato. *Pto* was the first serine-threonine protein kinase

to be cloned and established a phosphorylation cascade as part of recognition and response during plant defense (Martin et al., 1993). *Rpg1* from barley, which encodes a receptor kinase-like protein with two kinase-kinase domains, and the maize *Hm1* gene, which encodes the enzyme HC toxin reductase, also belong to this class of resistance genes encoding neither LRR nor NBS domains (Brueggeman et al., 2002; Johal and Briggs, 1992). The final class of resistance genes contains a transmembrane domain fused to a putative CC domain. The *Arabidopsis RPW8* gene for powdery mildew resistance is one such gene (Xiao et al., 2001). Wheat and barley have both been extensively studied in regard to mapping of resistance genes and agronomic traits. However, cloning of both has taken longer than numerous other species due to their complex and larger genomes (Seah et al., 1998). Traditionally, transposon tagging and map-based cloning strategies have been used to clone new R genes (Seah et al., 1998). In wheat, transposon tagging is not a feasible method due to the lack of an active transposon system (Huang et al., 2003). The predominant method is to use a map-based cloning approach using large mapping populations, BAC libraries, and mutants. Map-based cloning requires an accurate, fine genetic map with flanking markers linked closely to the gene of interest (Somers et al., 2004).

Hm1 was the first plant, disease resistance gene to be cloned and characterized (Richter and Ronald, 2000). The *Hm1* gene in maize provides resistance to the fungus *Cochliobolus carbonum* and determines sensitivity to HC-toxin (Johal and Briggs, 1992). Transposon-induced mutagenesis was used to generate and clone several *Hm1* alleles. This led to the discovery that the dominant *Hm1* allele reduces the plant's sensitivity by encoding an NADPH-dependent HC-toxin reductase that inactivates HC-toxin (Johal and Briggs, 1992). *Hm1* is just one member of a class of resistance genes that encode detoxifying enzymes (Richter and Ronald, 2000).

Tomato offers many advantages for cloning of *R* genes based upon linkage mapping because it has been extensively studied through linkage analysis and other cytogenetic studies (Martin et al., 1992). A map-based cloning approach was used to isolate *Pto* using a previously generated high-resolution RFLP map and the use of a yeast artificial chromosome clone that spanned the *Pto* region. *Agrobacterium tumefaciens* was used to transform a susceptible tomato cultivar with plasmids containing two cDNA clones representing a gene family primarily clustered at the *Pto* locus, and these plants were resistant (Martin et al., 1993). *Pto,* from tomato *Solanum lycopersicum*, is a serine/threonine kinase and was the first gene cloned that resembled a signal transduction component (Richter and Ronald, 2000). *Pto* interacts with the avirulence gene (*avrPto*) in *P. syringae* pv. *tomato* to induce resistance against bacterial speck in tomato cultivars (Martin et al., 1993). Kinase-phosphatase activities had previously been implicated in signaling between plants and pathogens, and the discovery of a serine/threonine protein kinase resistance gene further established the involvement of a phosphorylation cascade as part of plant defense (Martin et al., 1993).

Cre3 was the first extracellular NBS-LRR gene cloned from monocotyledonous plants (Lagudah et al., 1997). Lagudah et al. (1997) cloned the *Cre3* gene in wheat using map-based cloning. C*re3* is inherited as a single dominant locus and confers resistance to the cereal cyst nematode (*Heterodera avenae)* pathotype Ha13, which causes severe yield losses in wheat and other temperate cereals. An RFLP marker had been previously linked to *Cre3*, and this was used as the starting point for the map-based cloning (Eastwood et al., 1994). Some resistance gene loci have been shown to contain both a pseudogene and complete, functional gene sequence. Genomic analysis suggested that *Cre3* is one such locus (Lagudah et al., 1997).

Pm3 was the first locus identified for resistance to powdery mildew (*Blumeria graminis* f. sp. *tritici*) in wheat and the first powdery mildew resistance gene to be cloned. *Pm3b* encodes for a CC-NBS-LRR (Yahiaoui et al., 2004). The four stripe rust (*Puccinia striiformis*) resistance genes *Yr5, Yr10, Yr18,* and *Yr36* have been cloned in wheat (Chen et al., 2013; Laroche et al., 2000; Krattinger et al., 2009; Fu et al., 2009). *Yr10* was the first stripe rust resistance gene cloned to confer seedling resistance and encode a CC-NBS-LRR (Laroche et al., 2000). *Yr36* and *Yr18* were cloned around the same time and are both adult-plant resistance genes (Fu et al., 2009; Krattinger et al., 2009). *Yr36* is not present in modern pasta or bread wheat varieties but is present in wild emmer wheat. Introgression of *Yr36* into modern wheat could greatly improve resistance to stripe rust because *Yr36* confers durable, broad-spectrum resistance at relatively high temperatures (Fu et al., 2009). The race non-specific stripe rust resistance gene *Yr18* and powdery mildew resistance gene *Pm38* cosegregate with *Lr34* for leaf rust resistance, which has been cloned (Krattinger et al., 2009). *Lr34* encodes an ATP binding cassette (ABC) transporter protein and confers broad-spectrum, durable resistance in adult plants. *Lr34* was the first gene associated with durable, adult plant resistance to be cloned (Krattinger et al., 2009). Only three other leaf rust resistance genes, *Lr1*, *Lr10*, and *Lr21*, have been cloned so far. *Lr1*, *Lr10*, and *Lr21* all confer race-specific resistance at the seedling stage and encode proteins with nucleotide binding site-leucine rich repeat regions (Cloutier et al., 2007; Feuillet et al., 2003; Huang et al., 2003).

Lr21 was the first leaf rust resistance gene cloned. Resistance to *Lr21* has since been overcome (Keller et al., 2012). *Lr21* was also introgressed into wheat from an *Ae. tauschii* accession. The cosmid library was developed from the *Ae. tauschii* donor accession TA1649, and genetic analysis was done using wheat (Huang et al., 2003). The previously developed

RFLP marker *XKSUD14* was the closest marker to *Lr21* and was used as a probe to screen a cosmid library, which identified one clone that contained *Lr21*. Ironically, *XKSUD14* was contained in *Lr21* (Huang et al., 2003).

Subgenome chromosome walking using BAC clones from the *T. monococcum* DV92 library was used to clone *Lr10* because the genome of *T. monococcum* is closely related to the A genome of hexaploid wheat (Feuillet et al., 2003). Previously a serine/threonine protein kinase named LRK10 was linked to the *Lr10* locus (Feuillet et al., 1997). *Lr10* was mapped to chromosome 1DS. Two resistance gene analogs (RGAs), both encoding CC-NBS-LRRs, were identified in a region completely linked to *Lr10* in *T. aestivum.* Mutational analysis and stable transformation were used to demonstrate which of the two RGAs was actually *Lr10* (Feuillet et al., 2003). Comparison of the *Lr10* sequence showed significant similarity to RPM1 (Grant et al., 1995) of *A. thaliana* and low similarity to *Lr21* (Huang et al., 2003).

Lr1 was the first leaf rust resistance gene described by Mains et al. (1926). When Cloutier et al. (2007) cloned *Lr1*, virus induced gene silencing (VIGS) was used in addition to particle bombardment to confirm the candidate gene. This method proved valuable for identification of *R* genes. Sequence analysis of *Lr1* revealed that it had little sequence similarity to *Lr10* or *Lr21* despite encoding for a CC-NBS-LRR as well. *Lr1* does have a transmembrane domain predicted in the signal peptide that neither *Lr10* nor *Lr21* have. *Lr1* also has multiple coiled coil motifs (Cloutier et al., 2007).

Two of the most recent genes cloned from wheat are the stem rust resistance genes *Sr33* and *Sr35*. Both genes encode CC-NBS-LRRs and are of particular interest because they confer resistance to the *Ug99* races of stem rust (Periyannan et al., 2013; Saintenac et al., 2013). *Sr33* was introgressed into bread wheat from *Aegilops tauschii* and provides a moderate level of

resistance (Rouse et al., 2011). *Sr35* was cloned from *Triticum monococcum* and provides high levels of resistance to *Ug99* and related races of stem rust (Singh et al., 2011). Based on previous studies, the same level of resistance should be provided when transferred to hexaploid wheat by crossing or recombination (Saintenac et al., 2103). Combining the two genes at a single locus would provide the optimal level of resistance to stem rust because they collectively provide resistance to all races of stem rust (Periyannan et al., 2013).

Resistance Gene Analogs

Resistance gene analogs (RGAs) are DNA fragments with homology to conserved motifs among known *R* genes such as kinases, nucleotide binding site (NBS) domains, and leucine rich repeats (LRRs). Many resistance genes to date have been identified as members of the nucleotide binding site-leucine rich repeats (NBS-LRR) class of resistance genes (Quirin et al., 2012). NBS-LRRs are present in both monocots and dicots are involved in gene-for-gene disease resistance in fungal, viral, bacterial, and nematode systems (Seah et al., 1998). Typically these genes induce a hypersensitive response, which is a form of programmed cell death (Padmanabhan et al., 2009). The genome of *Arabidopsis* has 163 NBS-containing resistance genes, and the genome of rice has over 600 (Linden et al., 2004). The NBS sites have regions of high sequence similarity, especially the P-loop and Kinase-2a, which can be exploited when searching for other resistance genes (Seah et al., 1998).

Several experiments have been done, especially in *Solanum* species, to exploit the conserved sequences among resistance genes in order to map and clone important resistance genes (Leister et al., 1996; Quirin et al., 2012). Previous research has also indicated that *R* gene clusters in the *Solanaceae* family are conserved across species (Grube et al., 2000). Leister et al. (1996) first identified a PCR-based approach that allowed their laboratory to isolate potato

resistance genes using primers designed from conserved sequence motifs in *Arabidopsis thaliana* and tobacco resistance genes. The *N* gene from tobacco and *RPS2* gene from *A. thaliana* had been previously shown to have a common NBS motif (Mindrinos et al., 1994). PCR products from three primer combinations were analyzed for fragment polymorphism. A two-nucleotide difference in the antisense primer resulted in one main product consisting of highly repetitive sequence and another main product of DNA fragments that were highly informative and able to be mapped. They identified a total of 28 positions where genes were located on the genetic map of potato, and fifteen of these showed sequence homology to known plant R genes. One of the gene families identified was shown to be physically and genetically linked to the *Gro1* nematode resistance gene (Leister et al., 1996). A similar approach was taken in soybean, and oligonucleotide primers were designed from *N* and *RPS2* as well as *L6* from flax. Kanazin et al. (1996) were able to map classes of RGA loci to several previously mapped linkage groups containing known resistance genes, including *Rps1* and *Rps2*. Two RGAs mapped within a 3.8-cM cluster of resistance genes that includes the loci for *Rps2*, *Rmd*, and *Rj2*. A QTL associated with soybean cyst nematode resistance was also thought to be located near this region. These results indicated that mapping RGA sequences could be useful for identifying candidate resistance genes.

NBS profiling has been used to target plant disease resistance genes and provide molecular markers linked to *R* genes and RGAs without having to clone the amplified RGA fragments (van der Linden et al., 2004). NBS profiling uses degenerate primers that are homologous to conserved sequences in the NBS domain of NBS-LRR gene family members (Mantovani et al., 2006). Along with the NBS primer, NBS profiling also uses an adapter to amplify genomic DNA that has been digested with a restriction enzyme (van der Linden et al.,

2004). Van der Linden et al. (2004) found that depending on the NBS-profiling marker used, 50-90% of the amplified fragments in potato were significantly similar to known *R*–gene and RGA sequences. NBS profiling has also successfully been applied to multiple plants including apple (Calenge et al., 2005), lettuce (van der Linden et al., 2004; Syed et al., 2006), barley, tomato (van der Linden et al., 2004), pepper (Zhang et al., 2008; Kochieva and Ryzhova, 2009), cauliflower (Gu et al., 2008), and durum wheat (Mantovani et al., 2006).

Quirin et al. (2012) used seven partially degenerate primers and one nondegenerate primer to generate 97 RGAs from a disease-resistant, wild potato *Solanum bulbocastum* genotype. These 97 RGAs were combined into one meta-analysis consisting of over 800 sequences that included all of the known RGA sequences and cloned NBS-LRR gene sequences from all *Solanum* species. This was the first time a cross-species, genome-wide survey of the Solanum *R* gene space had been published. This research showed that *Solanum* species retained 80% NBS DNA sequence identity, even after ancient lineages underwent speciation events (Quirin et al., 2012).

In wheat, RGAs have been used in several mapping projects. Using markers designed from the previously cloned *Cre3* locus in wheat, Seah et al. (1998) were able to isolate a family of resistance gene analogs in wheat and barley using a PCR-based cloning approach. Primers were derived from regions that were conserved between *Cre3* and other known *R* genes. RGAs were isolated from the wheat cultivar Chinese Spring and the barley cultivars Chebec and Harrington. The RGAs isolated were shown to have 55-99% sequence identity and 72-99% sequence similarity to the NBS-LRR amino acid sequence at the *Cre3* locus. The isolated RGAs also contained conserved motifs present in known plant *R* genes (Seah et al., 1998). Primers designed from conserved sequences of cloned plant resistance genes have also been used to

develop molecular markers for the wheat stripe rust genes *Yr5* and *Yr9* (Yan et al., 2003; Shi et al., 2001). Dilbirligi and Gill (2003) used RNA fingerprinting and data mining approaches to identify expressed resistance gene sequences in wheat. Out of 220 expressed *R*-gene candidates identified, 125 sequences resembled structures of known *R*-genes. McFadden et al. (2006) obtained similar results when a total of 129 NBS-containing sequence groups were identified from wheat and barley ESTs. Two divergent NBS-LRR-gene sequences were used to retrieve ESTs containing NBS-region motifs and develop a set of RGA markers that covered a significant portion of the wheat genome. They detected RGA loci on all 21 chromosomes of a set nullitetetrasomic wheat lines and mapped RGAs to 18 chromosomes using a Cranbrook x Halberd doubled haploid mapping family. One RGA probe showed significant sequence homology to *Lr21*. This same probe mapped to a linkage group corresponding to the known location for *Lr21* on the distal region of chromosome 1DS (McFadden et al., 2006).

RNA-seq

The transcriptome is the set of all RNAs in the cell, including mRNAs, non-coding RNAs, and small RNAs (Wang et al., 2009). RNAs are part of a larger, complex network that ultimately drives biological processes. These networks need to allow for rapid adaptation of gene expression in order to cope with environmental stresses and other external challenges (López-Maury et al., 2008). RNA-seq was recently developed as a new approach for analysis of transcriptomes and will allow for better insight into functional elements of the genome as well as the molecular constituents of cells and tissues. A better understanding of plant transcriptomes will also provide insight into development and disease (Wang et al., 2009). However, the full potential of RNA-seq lies in the ability to study transcriptomes of organisms both with extensively annotated genomes and without a reference genome (Haas et al., 2013).

RNA-seq allows for RNA analysis by using next-generation DNA sequencing (NGS) technologies to sequence cDNA (Ozsolak and Milos, 2011). RNA-seq can be used for transcript profiling, single nucleotide polymorphism (SNP) detection, and analysis of differentially expressed genes (Li et al., 2013). The high resolution and sensitivity of RNA-seq has allowed mapping of the 5' and 3' boundaries of many genes along with the identification of many novel transcribed regions and splicing isoforms of known genes (Wang et al., 2009). RNA-seq has been applied to the two rice (*Oryza sativa*) subspecies, *indica* and *japonica* for discovery of novel transcribed regions, detection of splicing sites, assessment of gene expression, and identification of SNPs (Lu et al., 2010). Researchers identified 15,708 novel transcriptionally active regions (nTARs) and found that ~48% of rice genes show alternative splicing patterns. SNPs were identified through transcriptome comparisons between the two subspecies. The ratio of nonsynonymous vs. synonymous SNPs was nearly 1:1.06. The application of RNA-seq also allowed validation of current rice gene models and extension of thousands of identified gene boundaries by at least 50 bp. This approach was proven to be highly successful for global transcriptome profiling (Lu et al., 2010).

One major advantage of RNA-seq over other sequencing technologies is that RNA-seq does not require cloning and allows direct sequencing of cDNA fragments (Haas and Zody, 2010). Because the technology is not limited to detection of known transcripts, RNA-seq allows the characterization and quantification of new splice isoforms. Transcriptional boundaries of genes can also be defined at single nucleotide resolution (Costa et al., 2010). Another advantage of RNA-seq is that the data produced has proven to be reliable and highly reproducible for both technical and biological replicates (Costa et al., 2010). Studies have also proven that RNA-seq

performs at least as well as microarrays for measuring mRNA expression and identifying differentially expressed genes (Marioni et al., 2008).

Microarrays have previously been used to simultaneously monitor the expression levels in cells for all annotated genes (Dudoit et al., 2002). Microarrays have several disadvantages over RNA-seq though, including a limited dynamic range of detection. The absence of an upper limit for quantification with RNA-seq allows detection of transcripts over a larger dynamic range of expression levels (Costa et al., 2010). Analysis of small regulatory RNAs has been especially limited with the use of microarrays because small RNAs are typically too short to be adequately captured due to the limited resolution (Ulvé et al., 2007). Microarrays are also limited by the array content and measure only the relative abundance of transcripts instead of the total abundance. They can also suffer from background and cross-hybridization issues (AC't Hoen et al., 2008). RNA-seq does not suffer from these limitations and instead has low background signal and high sensitivity, which allows the detection of the expression levels of substantially more transcripts. Quantitative PCR (qPCR) has also shown RNA-seq to be highly accurate and superior to microarrays for quantifying expression levels (Nagalakshmi et al., 2008).

NGS protocols have quickly been adapted to allow the sequencing of double-stranded cDNA instead of genomic DNA (Costa et al., 2010). Illumina, Roche 454, Life Technologies, and Helicos BioSciences are the NGS platforms currently used for RNA-seq (Ozsolak and Milos, 2011). The improvements to current sequencing technologies enable longer reads that can be aligned to a reference genome more readily and be used to define splice sites without prior annotations (Haas and Zody, 2010). Early RNA-seq experiments produced relatively short read of only 25-32 bp in length (Haas and Zody, 2010). The length of reads produced in RNA-seq experiments has since increased to 30-400 bp depending on which DNA-sequencing technology

is used (Wang et al., 2009). Paired-end sequencing is also available for RNA-seq and allows for better alignment and characterization of reads. The ability to obtain sequence information from two points in a transcript makes it possible to search for splicing patterns without prior knowledge of transcript annotations (Ozsolak and Milos, 2011). Direct RNA sequencing (DRS) approaches such as the Helicos single-molecule sequencing platform have their own set of advantages and may overcome some of the problems associated with RNA-seq presented during cDNA synthesis. Smaller concentrations of RNA are also needed for DRS technologies, which would allow analysis of short, degraded, and/or small quantity RNA samples (Ozsolak et al., 2009). However, this technology does not have the high-throughput capacity needed to compete with current NBS sequencing technologies (Wang et al., 2009).

Reads from RNA-seq may be assembled using either a reference genome or by *de novo* assembly. Approaches using a reference genome have quickly become a standard for analysis with model organisms (Haas et al., 2013). This method of assembly is not feasible in organisms without a reference genome, such as wheat, and much of the focus is now shifting to *de novo* assembly (Oono et al., 2013). With d*e novo* assembly, reads are directly assembled into transcripts (Grabherr et al., 2011). Even with a reference genome, a *de novo* assembly approach can be useful for identification of transcripts that are missing or incomplete (Haas and Zody, 2010). However, *de novo* requires a much higher sequencing depth and more optimized hardware. Programs for *de novo* assembly are also very sensitive to sequencing errors and may not be able to distinguish highly similar transcripts (Li et al., 2013). In wheat, *de novo* assembly has been in combination with draft sequences of the A and D genomes to reconstruct a relatively accurate and essentially complete transcriptome for wheat development (Li et al., 2013).

One of the biggest challenges of RNA-seq is assembling the large amount of data generated into a genome-scale transcription map. Both short and long transcriptomic reads present their own challenges when analyzing and aligning the data. Short reads may span exon junctions or contain poly(A) ends, both of which have to be analyzed differently. Reads from larger transcriptomes typically match multiple regions in the genome and have to be assigned correctly. Unique neighboring sequences are most commonly used for assignment (Wang et al., 2009). For most assembly methods, preliminary quality filtering to remove poor quality reads can reduce the computational time and effort for further analysis (Costa et al., 2010). Multiple platforms and protocols have been evaluated for use in RNA-seq transcript assembly. In one comparison, 25 transcript reconstruction protocols were evaluated based on genome alignments and *de novo* assembly. However, no single protocol excelled for all features evaluated. These results also indicated that the state of a reference genome assembly and the associated gene annotation are the two main factors for determining which method to use (Steijger et al., 2013).

Programs such as Trinity, SOAPdenovo-Trans, Trans-ABySS, and Velvet-Oases were developed specifically for *de novo* RNA-seq assembly (Li et al., 2013). Trinity consists of the three software modules Inchworm, Chrysalis, and Butterfly and was designed specifically for transcriptome assembly. When tested on samples from fission yeast, mouse, and whitefly, Trinity was able to reconstruct splice isoforms and transcripts from recently duplicated genes. Trinity was also able to resolve ~99% of the initial sequencing errors and identify allelic variants (Grabherr et al., 2011). Another major benefit of Trinity is that it is able to accommodate strand-specific Illumina paired-end libraries in addition to non-stand specific and single-end-read data (Haas et al., 2013). When compared with SOAPdenovo-Trans, Trans-ABySS, and Velvet-Oases, Trinity was the most efficient and sensitive when using short reads to assemble full-

length transcripts and isoforms from several model organisms (Grabherr et al., 2011). However, Oases has also proven to be useful. Oases was successfully used for the characterization of gene expression of phosphate starvation in wheat. Oases produced the longest transcripts and was able to reconstruct the highest percentage of alignments to wheat full-length cDNAs (Oono et al., 2013).

Genotyping-by-sequencing

Genotyping-by-sequencing (GBS) is quickly becoming an economically feasible and widely used method of high-density single nucleotide polymorphism (SNP) discovery and DNA sequencing, even in organisms without a reference genome available (Elshire et al., 2011). SNPs are the most common form of variation found in genomes (Metzker, 2005). GBS has the potential to allow breeders to characterize germplasm used in their breeding programs without having to develop molecular markers or other molecular tools first. Restriction enzymes are utilized to reduce genome complexity when constructing libraries (Elshire et al., 2011). Barcoded adapters are used to sequence multiple samples in one run and separate them later (Poland et al., 2012a). Using restriction enzymes may also allow researchers to avoid highly repetitive regions of the genome and to reach regions of the genome that may be otherwise inaccessible. GBS may also be used to access regulatory regions located in non-coding DNA (Elshire et al., 2011). The use of different restriction enzymes can also change the marker density attained, meaning the restriction enzyme(s) used can be tailored to the organism to achieve the best results. More frequent recognition sites allows for a greater number of SNPs to be assayed but creates a lower complexity reduction of the genome (Baird et al., 2008, Poland et al., 2012a). However, the need for very accurate DNA quantification has proven to be necessary to reduce sample-to-sample variation in sequence coverage (Elshire et al., 2011). Many of these

markers also make it difficult to distinguish between homozygous and heterozygous individuals in the population (Allen et al., 2013).

The first use of restriction enzymes with multiplex next-generation sequencing for SNP-discovery utilized restriction-site associated DNA (RAD) tags (Baird et al., 2008). RAD tags are markers originally developed for microarrays and are "short fragments of DNA adjacent to each instance of a particular restriction enzyme recognition site." Baird et al. (2008) were able to identify more than 13,000 SNPs in two model organisms by developing new RAD tags for use with an Illumina Genome Analyzer sequencer. They were also able to map three different traits in these two species of stickleback as well as an induced mutation in *Neurospora crassa*. For their method, genomic DNA was digested with a restriction enzyme and an adapter was ligated to the overhanging ends of the fragment. These fragments were then pooled, randomly sheared, and size selected before being ligated to a second, divergent adapter. In order to fine map RAD sequences, a barcode was additionally ligated for each individual.

Since then, several other methods of GBS have been developed and used. A more recent GBS method developed by Elshire et al. (2011) utilizes barcodes included in the adapter sequence ligated just upstream from the restriction enzymes. These barcodes allow multiple samples to be pooled into a single sequencing channel. This method removes the need for multiple sequencing runs and size selection of fragments. This procedure was demonstrated on barley and maize and roughly 25,000 and 200,000 sequence tags were mapped (Elshire et al., 2011). Poland et al. (2012) developed another GBS approach, from which 20,000 SNPs were mapped onto a wheat reference map and 34,000 SNPs were mapped onto a barley reference map. This was accomplished by combining a "common-cutter" restriction enzyme with Y-adapters

and by also including a "rare-cutter" restriction enzyme to simplify quantification of the library prior to sequencing (Poland et al., 2012a).

For GBS, the plant material used is typically from doubled haploid or recombinant inbred lines (Masher et al., 2013; Poland et al., 2012a). Multiplex libraries are created from DNA digested with restriction enzymes, ligated with barcode adapters, pooled, and PCR amplified (Poland et al., 2012a). Libraries are then sequenced using Illumina sequencing or the Ion Torrent sequencing platform (Mascher et al., 2013). SNP calling can be performed using the SAMtools/BCFtools pipeline, the UNEAK pipeline, or the TASSEL GBS pipeline (Li et al., 2009; Lu et al., 2013; Glaubitz et al., 2014). SNP data obtained can then be used for association mapping, genomic selection, and KASP genotyping assays (Poland and Rife, 2012; Uitdewilligen et al., 2013).

Much of the focus of GBS has been on organisms with complex genomes and lack of reference genome. Barley (*Hordeum vulgare* L.) is a diploid plant with a relatively large genome that has the potential to serve as a genetic model for other small-grain, temperate cereals, such as wheat (Mascher et al., 2013). Liu et al. (2014) used GBS to identify the genetic location of the *ari-e.GP* semi-dwarfing gene of cultivated barley. They were able to confirm the location of *ari-e.GP*, which is a recessive mutant allele of the *Breviaristatum-e* (*Ari-e*) gene, on chromosome 5H of barley. They did find the GBS data more difficult to handle and analyze than another multiplex SNP assay technology they normally used. Out of 1,400 informative markers identified, the most closely linked markers spanned a region of 7 cM and were not as closely linked as hoped. This may have been due to a shared haplotype between the two parents in the region of interest (Liu et al., 2014). Another study focused on switchgrass, which has proven to be an especially difficult organism to use for genomic selection and genome-wide association

studies. The reduced representation library created by using GBS allowed identification of 1,242,860 SNPs. The information gathered provided information on the diversity, genomic complexity, population structure, phylogeny, phylogeography, ploidy, and evolutionary dynamics of switchgrass. Two linkage maps were also constructed based on the high conservation of genome structure between switchgrass and foxtail millet (*Setaria italica* (L.) P. Beauv.; Lu et al., 2013). Maize is another important crop for which GBS has been applied. Crossa et al. (2013) applied GBS to maize in order to compare GBS data with that obtained from pedigree methods. They discovered significant gains by combining the two sets of data versus just using pedigree methods. Using multiple traits under varying environmental conditions, they also found increased predictive ability with the use of GBS (Crossa et al., 2013). GBS is not limited to use in monocots and can also be applied to dicotyledonous species. GBS was successfully applied in soybean to identify 10,120 SNPs. Of these SNPs, 60.57% were found in intergenic regions (Sonah et al., 2013).

Objective

Puccinia triticina Eriks causes significant yield losses in wheat worldwide. *Lr16* is a wheat leaf rust resistance gene found in some varieties grown in the Northern Great Plains of the U.S. that provides partial resistance at the seedling stage. *Lr16* is located at the tip of chromosome 2B, and the closest molecular markers have mapped between 1 and 9 cM depending on the population used. The objective of this research was to use RNA-seq and genotyping by sequencing (GBS) technologies to develop new markers linked to *Lr16*. RNA was isolated from seedlings of Thatcher (Tc) and isoline Tc*Lr16* and used to generate cDNA. Using *in silico* subtraction, sequence from Tc*Lr16* was aligned to Thatcher. Non-aligning sequence was assembled and analyzed using BLAST to identify putative gene functions.

Primers were designed to target resistance gene analog sequences not found in Tc. PCR reactions were performed using these primers and previously known SSR markers on 260 F_2 individuals of a CS x Tc*Lr16* mapping population. DNA for GBS was isolated from 22 individuals from a CS x Tc*Lr16* F_2 population. The UNEAK and Tassel pipelines were used along with custom R script to identify and filter putative SNPs. KASP markers were designed using the SNP calls identified, and KASP assays were run on 129 individuals of a Tc x Tc*Lr16* F_2 population.

Chapter 2 - Using RNA-seq and *in silico* subtraction to identify RGA markers for *Lr16*

Introduction

Puccinia triticina Eriks, causal agent of wheat leaf rust, is responsible for significant yield wheat losses worldwide and is the most widespread cereal rust pathogen in North America (McCartney et al., 2005). Leaf rust infects bread wheat (*Triticum aestivum* L.), durum (*T. turgidum* L. var. *durum*), triticale (X *triticosecale* Wittmack), goat grass (*Aegilops cylindrica*), *Aegilops speltoides*, cultivated emmer (*T. dicoccon*), and wild emmer (*T. dicoccoides*; Bolton et al., 2008). Leaf rust can overwinter on winter wheat species in the warmer, Southern states and induce epidemics after the wheat breaks dormancy and resumes growth in the spring (Eversmeyer and Kramer, 2000). The use of fungicides can reduce the impact of leaf rust on wheat, but economic factors prevent broad use of chemicals. R-gene mediated defense in the plant provides the best control in the event of infection (Gururani et al., 2012). Breeding for leaf rust resistance has proven to be one of the most effective methods of control for leaf rust in wheat (Kolmer and Oelke, 2006). Molecular markers linked to resistance genes aid in the process of marker-assisted selection (MAS) and gene pyramiding to allow for more effective resistance (Dholakia et al., 2013).

Microsatellites, or simple sequence repeats (SSRs), are tandem repeats of short sequences, usually 2-6 nucleotides in length (Kuleung et al., 2004). Microsatellites occur in all eukaryotic genomes, and the most common classes are dinucleotide, trinucleotide, and tetranucleotide repeats (Gibson and Muse, 2004). Due to the highly repetitive nature of the wheat genome and low level of polymorphism, microsatellites have been widely used for linkage

analysis studies and MAS (Röder et al., 1998). Microsatellites have also proven to be worthwhile because they are codominant, have high levels of polymorphism, and are predominantly genome specific (Röder et al., 1995). The major disadvantage for the use of SSRs in the large amount of time and labor required for the initial development of functional markers (Röder et al., 1998).

Lr16 is a wheat leaf rust resistance gene that provides partial resistance at the seedling stage (McCartney et al., 2005). Five wheat varieties are known to be sources of *Lr16*. Warden was one of the original sources of *Lr16* (McIntosh et al., 1995). Exchange was produced from a cross containing Warden and was then used in a cross for Selkirk (Anderson, 1961). The other two sources of *Lr16* are the varieties Etoile de Choise and Columbus (Bartos et al., 1969; Samborski and Dyck, 1982). Akin et al. (2013) evaluated 76 winter-facultative wheat cultivars used worldwide for leaf rust resistance genes, including 40 Thatcher leaf rust isolines. *Lr16* was only identified in 7 of these cultivars. The presence of *Lr16* in spring wheat varieties is more common. In an evaluation of 26 hard red spring wheat cultivars, 14 cultivars were postulated to contain *Lr16* (Oelke and Kolmer, 2004). One more of these varieties, Ivan, was later determined to have *Lr16* as well (Kolmer and Oelke, 2006). The spring wheat varieties AC Majestic, AC Karma, and AC Splendor have also been determined to have *Lr16* (Kolmer and Liu, 2002).

Leaf rust symptoms in wheat lines containing the gene *Lr16* are indicated small uredinia surrounded by a necrotic ring (Bolton et al., 2008). As cultivars containing *Lr16* become less predominant in the field, leaf rust races do not maintain their virulence and become less virulent on *Lr16*. This makes the *Lr16* resistance gene a good candidate for gene recycling in breeding programs (Marshall, 1992). *Lr16* has also been shown to interact with the adult-plant resistance genes *Lr13* and *Lr34* to provide higher levels of resistance (Kolmer, 1992; German and Kolmer,

1992). *Lr16* has not been cloned, but it has been mapped to the tip of chromosome 2B. The closest molecular markers have mapped between 1 and 9 cM depending on the population used (Somers et al., 2004). *Xwmc764, Xwmc661,* and *Xbarc35* are all microsatellites and have all been shown to map within 10 cM of *Lr16* in previous studies using doubled haploid (DH) and recombinant inbred (RI) lines (McCartney et al., 2005).

Several classes of plant disease resistance genes exist, and each class has conserved motifs. The most common and most studied class contains a nucleotide-binding site (NBS) and leucine rich repeat (LRR) domain (Ellis and Jones, 2003). Oligonucleotide primers can be designed for these conserved sequences to isolate and map new resistance gene analogs (RGAs) in various species, such as maize (*Zea mays*) and wheat (Collins et al., 1998; Maleki et al., 2003). One of the first instances of the use of conserved sequences in resistance genes to design primers was in potato. The use of primers designed from conserved sequences of the virus resistance gene *N* in tobacco and *RPS2* in *Arabidopsis thaliana* led to the discovery of an amplification product absolutely linked to the nematode resistance locus *Gro1* and the *Phytopthora infestans* resistance locus *R7* in potato (Leister et al., 1996). Dodds et al. (2001) further confirmed the utility of this approach by isolating an RGA clone that co-segregated with the *N* locus of flax through the use of oligonucleotide primers based on the conserved domains of the NBS domain. Spielmeyer et al. (1998) used RGAs containing sequences solely from the NBS-LRR class of resistance genes as probes to map and identify RGA loci in wheat. Chen et al. (1998) used primers derived from LRR, NBS, and protein kinase genes to demonstrate that polymorphisms associated with these conserved domains were inherited as single loci in wheat, rice, and barley. Primers may also be designed for conserved sequences between particular cloned *R* genes. After the *Cre3* locus for cereal cyst nematode (CCN) resistance was cloned by

Lagudah et al. (1997), Seah et al. (1998) used sequences conserved between the *Cre3* locus in wheat and other known *R* genes to clone and map a family of RGAs in wheat and barley. Exploiting these conserved sequences can also lead to the discovery of new resistance genes. The rice blast (*Magnaporthe oryzae*) resistance gene *Pid3* was identified using molecular markers designed for pseudogene members of rice (*Oryza sativa*) NBS-LRR gene families (Shuang et al., 2009).

New technologies like RNA-seq are allowing for even more in-depth analysis of genomes. RNA-seq can be used for organisms with a reference genome and allows analysis of splice sites and isoforms of both known and novel genes (Haas and Zody, 2010). For organisms like wheat without an annotated genome, RNA-seq allows analysis of the entire transcriptome without the need for cloning. Even genes that are expressed at low levels, such as resistance genes, can be sequenced through the use of RNA-seq (Grabherr et al., 2011). Next-generation sequencing technologies such as Illumina, Roche 454, or AB SOLiD sequencing are traditionally used for sequencing of cDNA fragments (Marguerat and Bahler, 2010). More recently Ozsolak et al. (2009) detailed a method for direct sequencing of RNA fragments, which is more useful for RNA samples that are degraded or have low concentrations. Overall, RNA-seq allows for unprecedented insight into the transcriptomes of all organisms.

Objective

Leaf rust, caused by *Puccinia triticina*, causes significant losses in wheat worldwide. Timely application of fungicides can reduce the impact of leaf rust, but resistant varieties are still the best method of control. The use of molecular markers closely linked to resistance genes can significantly aid in the process of marker-assisted selection (MAS). SSRs have been widely used in breeding programs, largely because of their success in the highly repetitive wheat genome.

The major advantages of SSRs are that they are mostly genome specific, have levels of polymorphism, and are codominant. *Lr16* is a desirable gene to incorporate into breeding programs because it provides partial resistance at the seedling stage, even in the presence of virulent leaf rust populations. This resistance in enhanced in the presence of *Lr13, Lr23,* and *Lr34. R*esearch has also indicated that *Lr16* is a good candidate for gene recycling due to the reduced virulence of leaf rust populations after cultivars containing *Lr16* become less predominant in the field. *Lr16* is also always associated with the stem rust resistant gene *Sr23*. The closest SSR to *Lr16* mapped between 1 and 9 cm away, depending on the population used. Other markers closely linked to *Lr16* to best incorporate the gene using MAS into breeding programs.

The objective of this research is to use resistance gene analogs (RGAs) identified from RNA-seq data to develop new markers linked to *Lr16*. RNA was isolated from seedlings of Thatcher (Tc) and isoline Tc*Lr16* and used to generate cDNA. Using *in silico* subtraction, sequence from Tc*Lr16* was aligned to Thatcher. Non-aligning sequence was assembled and analyzed using BLAST to identify putative gene functions. Primers were designed to target resistance gene analog sequences not found in Tc. PCR reactions were performed using these primers and previously known SSR markers on 262 F_2 individuals of a CSxTc*Lr16* mapping population including the parental lines. Two new RGA markers were found to map near *Lr16* and known SSR markers.

Materials and Methods

Plant Material and Growth Conditions

A cross was made between two spring wheat (*Triticum aestivum*) lines, Chinese Spring (CS) and the Thatcher (Tc) isoline containing *Lr16* (RL6005). The F_1 generation was self-

pollinated to produce the F_2 population. Individual F_2 seeds were germinated in 3.8-centimeter (cm) cone-tainers filled with Metro-Mix 360 (Sun-Gro Horticulture, Vancouver, Canada) and later transferred to one-gallon pots filled with Metro-Mix 360 at a density of two plants per pot. Plants were kept in a greenhouse with supplemental high-pressure sodium lighting set for 16-hour days and eight-hour nights. Daytime temperatures were set at 20°C, and nighttime temperatures were 18°C. Plants were grown until maturity, dried down, and F_3 seed harvested.

Plant Inoculation and Screening

F_2 plants were inoculated with Race 1, BBBD of *P. triticina* at the seedling 2-3 leaf stage. Urediniospores were suspended in Soltrol 170 (Phillips Petroleum, Bartlesville, OK) and applied with an atomizer at 20 psi. The plants were left overnight in a mist chamber at 18°C with 100% relative humidity and then moved back into the greenhouse. Infection types were evaluated at 10-14 days post inoculation and scored as either resistant or susceptible. A resistant, low infection type consisted of the presence of a small pustule surrounded by a distinct yellow halo (Figure 1a). Susceptible, high infection types consisted of large pustules with no halos (Figure 1b). To determine the zygosity of the F_2, $F_{2:3}$ families were evaluated for their infection types. Sixteen seeds from each line were planted in 8.9 cm pots containing Metro-Mix 360. These families were inoculated with leaf rust Race 1, BBBD and scored 10-14 days after inoculation as either resistant, heterozygous, or susceptible. Chi-square analysis was used to evaluate the inheritance of the leaf rust resistance gene, *Lr16,* in both the F_2 and F_3 populations.

DNA Isolation

Three leaf tips 3.8 cm long from newly emerged leaves were collected for DNA isolation from individual F_2 seedlings and placed in collection plates consisting of 96-2 mL collection tubes with a 3.9688 mm steel bead (Abbott Ball Company, West Hartford, CT). Samples were

quick frozen in liquid nitrogen and stored at -80°C until lyophilization and extraction. Tissue samples were lyophilized in a Labconco (Labconco, Kansas City, MO) lyophilizer for 2-3 days at -50°C and 0.050 mBar. The lyophilized samples were ground using a Retsch MM 400 (Retsch, Haan, Germany) and DNA was isolated following a modified version of the protocol "Purification of Plant DNA using the BioSprint 96" from the BioSprint DNA Plant Handbook, version 03/2005, (Qiagen, Valencia, CA). Tissue collection racks were placed in the Tissue Lyser and homogenized at 29 Hz instead of 30 Hz for 1 minute. After ensuring all collection microtubes were being adequately ground, samples were homogenized for 1 additional minute at 29 Hz. In step 12, after 300 µL of Buffer RLT was pipetted into each collection microtube, sealed, and mildly shaken back and forth, the racks were centrifuged at 5600 x *g* for 10 minutes at 21°C in a Hermle Labnet Z323K universal centrifuge (Labnet International, Woodbridge, NJ) instead of 6000 x *g* for 5 minutes. Two hundred µL of supernatant was transferred into an S-Block (Qiagen) followed by 200 µL of isopropanol. MagAttract Suspension G was vortexed for 3 minutes before adding 20 µL to each well of the S-Block. The remainder of the extraction was carried out via the "BS96 DNA Plant" protocol on the BioSprint 96 workstation. Slots 1-5 and 7 were all loaded as noted in the protocol, and 100% ethanol was used in slots 3 & 4. The volume of elution buffer loaded in the microplate in slot 6 was modified to have 100 µL of EB buffer instead of 200 µL to increase the DNA concentration. Upon program completion, the elution microplate was sealed and stored in the -20°C freezer until needed.

Simple Sequence Repeat Analysis

Primers were made for SSR markers known to map near the distal end of chromosome 2BS (Röder et al., 1998; Somers et al., 2004; Song et al., 2002; Song et al., 2005). The GrainGenes (USDA-ARS) website was used to identify loci nearby *Lr16* and to obtain the

correlated SSR primer sequences. Data for mapping was collected using polymerase chain reaction (PCR), M13 tailing, and fluorescent capillary electrophoresis. Each 25 μL PCR consisted of the following: 200 ng genomic DNA, 5.0 picomoles (pmol) reverse primer, 1.0 pmol of M13-tailed forward primer (Schuelke, 2000), 5.0 pmol of M13 primers labeled with 6-FAM or VIC (Applied Biosystems, San Diego, CA), 1× PCR Buffer (Sigma, St. Louis, MO), 2.5 mM $MgCl_2$, 2.5 mM dNTPs, 1.25 U Taq DNA polymerase (Sigma). MJ Research (Watertown, MA) PTC-225 and PTC-100 thermal cyclers were used for the selective amplification reactions of the markers *Xwmc661* and *Xwmc764* and programmed with the following conditions: 92°C for 3 minutes; 35 cycles of 92°C for 1 minute, 60°C for 1 minute 30 seconds, and 72°C for 2 minutes; and finally 72°C for 10 minutes (Bremenkamp-Barrett et al., 2008). For the *Xbarc35* marker, selective amplification was performed on an MJ Research PTC-225 with the following conditions: 95°C for 3 minutes; 36 cycles of 94°C for 40 seconds, 55°C for 40 seconds, and 72°C for 1 minute; and 72°C for 10 minutes (Song et al., 2005). An Applied Biosystems 3730 DNA analyzer was used for capillary analysis of PCR fragments. Samples were prepared for genotyping by pooling 2 μL of two different PCR reactions, each with either 6-FAM or VIC fluorescent tags, and 2 μL of ddH_2O. One μL of the pooled PCR was added to 6 μL of formamide (Applied Biosystems), 3 μL of ddH_2O, and 0.1 μL GeneScan 500-LIZ size standard (Applied Biosystems) in a 96-well semi-skirted plate. Plates were then incubated in an MJ Research PTC-200 at 95°C for 5 minutes followed by 10°C for 5 minutes to denature the DNA. Raw data files from the ABI 3730 were imported into GeneMarker V1.85 (SoftGenetics, State College, PA) and analyzed for polymorphisms.

RNA Isolation

Seeds for Thatcher (Tc), Tc*Lr10*, Tc*Lr21*, and Tc*Lr16* were germinated in 8.9 cm pots filled with Metro-Mix 360 (Sun-Gro Horticulture, Agawam, MA). Four pots were planted for each line, and four seeds were planted in each pot. Plants were grown in a Percival (Percival Scientific, Perry, IA) growth chamber maintained at 18°C with 16-hour days. Seven to ten centimeters of leaf tissue were collected at the two-leaf stage from all four lines for RNA isolation. The tissue was immediately frozen in liquid nitrogen and stored at -80°C until use. RNA isolations were performed using mirVana™ miRNA Isolation Kit (Ambion, Austin, TX) following the procedures for Organic Extraction and Final RNA Isolation. RNA was isolated from three replicates of each of the four lines. Three leaf tips 3.8 cm long were taken from the previously collected leaf tissue and ground in 1.5 mL microcentrifuge tubes in liquid nitrogen using a micropestle cooled in liquid nitrogen. After each sample was ground into a fine powder, 500 µL of Lysis/Binding buffer was added, mixed by inverting tubes until the clumps were dispersed, and the samples were placed on ice. Fifty µL of miRNA Homogenate Additive were added to each sample, mixed thoroughly by inverting tubes, and iced for ten minutes. Five hundred µL of Acid-Phenol:Chloroform (Ambion) was added, vortexed for 60 seconds, and centrifuged in an IEC Micromax benchtop centrifuge (Thermo Scientific, Waltham, MA) for five minutes at 17,000 x *g* to separate the aqueous and organic phases.

Four hundred µL of the upper, aqueous phase was removed and transferred to a fresh tube. Then 500 µL of room temperature 100% alcohol was added to the fresh tube with the aqueous phase. A filter cartridge was added to each new collection tube, and 700 µL of the lysate/ethanol mixture was pipetted into the filter cartridge. Each sample was centrifuged for 15 seconds at 9,300 x *g*. The filter was washed by adding 700 µL miRNA Wash Solution 1 and centrifuged for ten seconds. The filter was washed two additional times by adding 500 µL of

Wash Solution 2/3 and centrifuged for ten seconds. To ensure that any residual fluid was removed from the filter, the assembly was centrifuged for one minute at 9,300 x *g*. After the filter cartridge was transferred to a fresh collection tube, 100 µL of Elution Solution preheated to 95°C was added to the center of the filter and spun at 17,000 x *g* for 30 seconds to recover the RNA. RNA presence was checked on a 2% agarose gel and concentrations were determined by a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE). The RNA samples were sent to Cofactor Genomics (St. Louis, MO) for sequencing.

RNA Sequencing and In Silico Subtraction

Cofactor Genomics (St. Louis, MO) used the total RNA for sequencing using the protocol summarized by Bruce et al. (2013). In brief, the cDNA was sequenced using Illumina 60 bp paired-end reads (Cofactor). The cDNA was fragmented in vitro, sequenced, and then reassembled in silico. A reference EST set was *de novo* assembled for Thatcher using Inchworm (Broad Institute, Cambridge, MA). Tc*Lr10*, Tc*Lr21*, and Tc*Lr16* sequences were aligned back to Thatcher using Novocraft novoalign v2.07.10 (Novocraft, Selangor, Malaysia), and fragments that did not align were assembled into contigs using Inchworm. Contigs were Batch Blasted using blastcl3, blastx, and blastn. Contigs with alignments to NBS-LRR, NB-ARC, LRR, kinase, and *R* gene domains were identified (Appendix A). The complete conserved domains for *Lr10* (accession AY270157) and *Lr21* (accession FJ876280) from the NCBI nucleotide database were used for BLAST analysis against the contigs from the Tc isolines. Primers were designed (Appendix B) using MacVector (MacVector, Inc., Cary, NC) from the contigs with sequence similarities to resistance genes in order to target RGA sequences not found in Tc. The parameters for the primer characteristics were a length of 18-25 bases, between 45 and 55%

G+C, and a Tm of 50 to 65°C using the PCR Primer Pairs command. Primer names correlate with the number of the contig from which the primer was designed.

Primers were also designed using expression data analyzed in Cofactor's ActiveSite Viewer. Sequences with a better possibility of containing *Lr16* could be sorted out by changing the number of normalized counts for Tc*Lr16* sequences versus Tc, Tc*Lr10*, and Tc*Lr21*. BLASTn and BLASTx algorithms were used to determine putative gene functions of the samples with higher than 40 counts in the Tc*Lr16* expression data. Sequences that aligned to conserved domains of resistance genes, such as NBS-LRRs, and to known RGAs and resistance genes were used to design primers.

Resistance Gene Analog Analysis

PCR reactions were performed on the F_2 CSxTc*Lr16* mapping population using 137 primer pairs to identify polymorphisms between CS and Tc*Lr16* and among resistant and susceptible plants. All primer pairs were initially tested on CS and Tc*Lr16* DNA. Each 25 µL reaction consisted of the following reagents: 200 ng genomic DNA, 5.0 pmol reverse primer, 5.0 pmol of forward primer, 1× PCR Buffer (Sigma), 2.5 mM $MgCl_2$, 2.5 mM dNTPs, 1.25 U Taq DNA polymerase (Sigma), and 14.3 µL ddH_2O. MJ Research (Watertown, MA) PTC-225 and PTC-200 thermal cyclers were used for the selective amplification reactions with the following conditions: 92°C for 3 minutes; 35 cycles of 94°C for 40 seconds, 60°C for 40 seconds, and 72°C for 1 min; 72°C for 10 minutes. Any primers that did not amplify were redesigned. If polymorphisms were detected between the CS and Tc*Lr16* amplified fragments, the PCR was expanded to include 8 susceptible plants and 8 resistant plants. If polymorphisms were consistent among the resistant and susceptible lines, the PCR was further expanded to include the entire population.

RGA Cloning

RGA markers that were run on the entire population were cloned for sequencing. Each 25 µL PCR reaction consisted of the following reagents: 200 ng genomic DNA, 5.0 pmol reverse primer, 5.0 pmol of forward primer, 1× PCR Buffer (Sigma), 2.5 mM $MgCl_2$, 2.5 mM dNTPs, 1.25 U Taq DNA polymerase (Sigma), and 14.3 µL ddH_2O. MJ Research (Watertown, MA) PTC-225 and PTC- 200 thermal cyclers were used for the selective amplification reactions with the following conditions: 92°C for 3 minutes; 35 cycles of 94°C for 40 seconds, 60°C for 40 seconds, and 72°C for 1 min; 72°C for 10 minutes. For markers that amplified two bands, the polymorphic band was excised from a 2% HiRes agarose gel using the QIAquick Gel Extraction Kit (Qiagen) following a modified version of the microcentrifuge protocol in the QIAquick® Spin Handbook (May 2012). Incubation temperatures in step 3, centrifugation speeds, and the volume of Buffer EB added in step 12 were modified from the original protocol. The gel slice weighed 149.8 mg, so 450 µL of Buffer QG was added. The sample was incubated for 10 minutes at 40°C and mixed every 2-3 minutes. Isopropanol (150 µL) was added to the sample and mixed. The sample was pipetted into the QIAquick spin column and centrifuged at 17,000 x *g* for 1 minute. After the flow-through was discarded, 750 µL of Buffer PE was added, and the sample was centrifuged for 1 minute at 17,000 x *g*. The flow-through was discarded and the samples were centrifuged again for 1 minute at 15,800 x *g*. Buffer EB (10mM Tris-Cl, pH 8.5, 30 µL) was added to the spin column in a clean 1.5 mL microcentrifuge tube. The column sat for 4 minutes before centrifuging for 1 minute at 17,000 x *g*.

PCR fragments were ligated into the pCR2.1 vector (Invitrogen, Carlsbad, CA) and transformed into OneShot® INVαF´chemically competent *E. coli* cells following the protocol provided with the TA cloning kit. Blue/white colony screening was used. Colonies were grown on LB agar plates containing 0.1 mg/mL ampicillin, and white colonies were chosen for

screening. Eight colonies from each ligation were PCR amplified to check for an insert. Four clones with the gene of interest from each ligation were used for minipreps. Plasmid DNA was extracted from the bacteria using the QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA) following the protocol provided by the manufacturer. Samples were sequenced at the Kansas State University Sequencing Facility.

Linkage Analysis

Marker data from the RGAs was combined with the SSR data and mapped using MapMaker V3.0 (Lander et al., 1987). The RAW file was created as an F_2 intercross with 262 progeny, six genetic loci, and zero quantitative traits. The CS and Tc*Lr16* parents were included the mapping. A hyphen was used to indicate missing data in the raw file for both phenotypic and marker data. The 'cent func k' command was used to change the mapping function from Haldane to Kosambi mapping units. Linkage groups were formed with a LOD threshold of 3.00. The compare, sequence, and map functions were used to determine the best order of the loci, and the 'ripple' command was used to verify the order.

Results

Lr16 Segregation

The presence of *Lr16* in a wheat plant is indicated by a distinct phenotype consisting of a small pustule with a yellow halo (Figure 2.1a). Susceptible plants without *Lr16* do not have this halo and will have larger pustules (Figure 2.1b). For Chi-square analysis of the F_2 population, the null hypothesis was that the phenotypes of the population will fit a ratio of three resistant to one susceptible, and the alternative hypothesis was that the phenotypes would not fit the same 3:1 ratio. Out of 260 lines, 192 lines were scored as resistant, and 68 lines were scored as susceptible. Parental lines were not included in Chi-square analysis. The Chi-square value was

0.18 for a P value between 0.95 and 0.50 at 1 degree of freedom (DF). The null hypothesis failed to be rejected. The null hypothesis for the F_3 populations was that the phenotypes would fit a 1:2:1 ratio of resistant:heterozygous:susceptible, and the alternate hypothesis was that the phenotypes would not fit this 1:2:1 ratio. Some of the lines did not produce enough seed to plant in the F_3 or the seeds did not germinate, so a lower number of lines were scored in the F_3. A total of 241 lines were scored: 69 lines were resistant, 115 were heterozygous, and 57 were susceptible. The Chi-square value was 1.6971 in the F_3 giving a P value between 0.50 and 0.20 with 2 DF. The null hypothesis failed to be rejected.

RNA-seq and RGA Analysis

A total of 7,032,164 RNA-seq reference bases were sequenced using the Illumina platform. Out of 22,362 reference sequences generated, the Tc, Tc*Lr10*, Tc*Lr16*, and Tc*Lr21* hit 8,214, 4,073, 4,619, and 5,199 targets, respectively. The *de novo* assembly consisted of 1,451,648 contigs with a mean length of 88.5 bases. BLASTx analysis of the Tc reference found 17 alignments for NBS-LRRs, seven alignments for NB-ARC domains, 12 alignments for LRR domains, and 15 alignments for serine/threonine and receptor kinases. The alignments to kinases includes alignments for the barley stem rust resistance gene *Rpg1*, the tomato resistance gene *Pti*, the wheat resistance genes *Tsn1* for tan spot resistance and LrK10 associated with leaf rust resistance, and the rice bacterial blight resistance gene *Xa21*.

The genes *Lr10* and *Lr21* have both been cloned so their complete sequences are available (Feuillet et al., 2003; Huang et al., 2003). During the *in silico* subtraction, all of the resistance genes should have been assembled into contigs and not aligned to the Tc reference. BLAST analysis of the assembled contigs should have identified matches for the cloned *Lr10* and *Lr21* complete coding DNA sequences (cds). However, no significant alignments to *Lr10*

were identified in the assembled contigs >100 bases in length. A total of five significant alignments to the *Lr21* complete cds were identified in the Tc*Lr21*, Tc, and Tc*Lr16* assembled contigs. Three of the alignments were to Tc matching 99% of 447 bases, 90% of 265 bases, and 85% of 177 bases. In Tc*Lr16*, 85% of 207 bases aligned to the Tc*Lr21* complete cds, and Tc*Lr21* aligned to 100% of 71 nucleotides.

Batch BLAST results of the contigs assembled from Tc*Lr16* identified 134 resistance gene analogs. The contigs had similarities to known resistance genes, other RGAs, NBS-LRR proteins, kinases, LRRs, and NB-ARC domains (Table 2.1). Expression analysis revealed seven more contigs that appeared to more highly express *Lr16* based on the number of normalized counts. The difference in expression levels between Tc and Tc*Lr16* for these contigs ranged from two-fold to almost 13-fold. The three contigs with alignments to RGAs were a205149, a211607, and 1213409. The a205149, a211607, and a213409 contigs had almost 13-fold, nine-fold, and six-fold higher expression levels of Tc*Lr16* than Tc, respectively. BLASTx analysis revealed alignments to NBS-LRRs for a205149 and a213409, RGA4-like proteins for a211607, and a conserved NB-ARC domain for the a213409 sequence. The other four contigs did not align to any RGAs. Primers were designed for all 134 RGAs from the *in silico* subtraction and the three RGAs from the expression analysis for a total of 137 primer pairs (Table 2.2).

Of the 137 RGA primer pairs 34 were polymorphic between CS and Tc*Lr16.* Of the 34, three were consistently polymorphic in the resistant/susceptible screen of 16 lines. The markers *XRGA22128* and *XRGA266585* were each polymorphic for all but one of the resistant and susceptible samples and were both used in PCR on the entire population. The primers for marker *XRGA22128* amplified a band at 375 bases for both CS and Tc*Lr16*. A second band was amplified at 275 bases in CS but not Tc*Lr16* (Figure 2.2a) and was used for scoring and mapping

purposes. This 275bp band did not amplify in Tc DNA. After cloning and sequencing, the NCBI BLASTx results included two RGAs for cysteine-rich receptor-like protein kinases. The marker *XRGA266585* amplified a band that was less than 100 bases in Tc*Lr16* and no band in CS (Figure 2.2b). This band was not amplified in Tc making it polymorphic between Tc and Tc*Lr16*. BLASTx analysis for the cloned fragment for *XRGA266585* did not return any significant results, most likely because it was only 55 bases long. *XRGA142887* was polymorphic for four of the resistant and susceptible samples (Figure 2.2c) and was initially used in PCR on the entire population. Upon further visual evaluation of the agarose gels, the band at 125 bp in CS amplified in the majority of samples, including resistant lines. It was concluded that *XRGA142887* was not closely linked to *Lr16* and was disregarded in the mapping.

SSR Marker Analysis

A total of eight SSR markers were previously screened on a different CSxTc*Lr16* mapping population. These markers were *Xwmc661, Xwmc764, Xbarc35, Xbarc124, Xgwm210, Xwmc489, Xwmc382,* and *Xgwm614*. Analysis of these SSR markers indicated polymorphisms with the three markers *Xwmc661, Xwmc764,* and *Xbarc35*. The peaks on the other five markers did not show consistent peaks between the parents and among progeny that could be used for scoring (unpublished data). An initial screen of the markers for *Xbarc45* on the current population as well indicated that *Xbarc45* was not polymorphic on either population. Analysis of the capillary analysis data in GeneMarker indicated that the CS samples had a peak at 284 bases, and the Tc*Lr16* samples had a peak at 276 bases with the *Xbarc35* marker (Figure 2.3c). With *Xwmc661*, CS had a peak at 203 bases, and Tc*Lr16* had a peak at 240 bases (Figure 2.3b). For both markers, heterozygous plants showed both peaks. When the *Xwmc764* marker was analyzed, four peaks were identified between CS and Tc*Lr16*. CS samples had peaks at 159 and

197 bases, and Tc*Lr16* had peaks at 150 and 189 bases. Heterozygous samples showed all four peaks (Figure 2.3a).

Linkage Analysis

All five markers mapped proximal to *Lr16* (Figure 2.4). Based on the agarose gel scores, *XRGA266585* mapped nearest to *Lr16* of all the RGA and SSR markers at a distance of 1.2 cM. *XRGA22128* was the furthest marker from *Lr16* mapping 25.6 cM away. Using the data from the capillary analysis of the SSR markers, *Xwmc764, Xwmc661,* and *Xbarc35* mapped between the two RGA markers at distances from *Lr16* of 4.2 cM, 10.9 cM, and 16.4 cM, respectively.

Discussion

RNA-seq allows analysis of the entire transcriptome of organisms. In this case, RNA-seq allowed identification and sequencing of expressed resistance gene analogs without intronic sequences. *Lr16, Lr10,* and *Lr21* are all seedling resistance genes. The tissue for RNA isolation was collected at the two-leaf stage because the resistance gene should be expressed at that stage. The plants were also grown in highly controlled environments, and tissue was only collected if all plants were at the same growth stage. This helped reduce environmental and biological factors that would increase variability in gene expression. The use of Thatcher isolines also provides a more uniform background for evaluation of resistance in lines containing the gene of interest.

Lr10 and *Lr21* are two wheat leaf rust resistance genes that have previously been cloned (Feuillet et al., 2003; Huang et al., 2003). Because *Lr10* and *Lr21* had been cloned, Tc*Lr10* and Tc*Lr21* were used as a reference in the RNA-seq analysis. BLAST analysis of the complete *Lr10* and *Lr21* nucleotide sequence against the custom local BLAST database should have found matches to the RNA contigs from the Thatcher isolines. One hit for *Lr21* was a 100% match to

71 bases from a 270 bp fragment in Tc*Lr21* indicating that at least part of the *Lr21* gene was isolated and sequenced using RNA-seq. BLAST analysis for *Lr10* did not return any matches. An *Lr10* homologue may have been filtered out in the initial RNA-seq analysis. Multiple alignments to NBS-LRRs were identified in the Thatcher reference making this a possibility. In this case, the resistance gene sequences from the Tc*Lr10* and Tc*Lr21* isolines would align back to Thatcher and would not be in the assembled contigs.

Ideally, one of the assembled contigs would have been the *Lr16* gene. The recent cloning of *Sr33* for stem rust resistance in wheat from an *Aegilops tauschii* RGA proves this possibility (Periyannan et al., 2013). However, a marker perfectly correlated with *Lr16* was not identified. The resistance gene class to which *Lr16* belongs is unknown for sure. The domains that make up the *Lr16* resistance gene may different than the kinase, NBS-LRR, or LRR domains searched for using the BLAST analysis, and the contig may have been overlooked as an RGA. Thatcher may also contain a non-functional *Lr16* homolog. In this case, the cDNA fragments from Tc*Lr16* would have aligned to the Tc reference preventing the resistance gene from filtering out for contig assembly.

Two markers linked to *Lr16* were identified though. In this population, *XRGA266585* was more tightly linked to *Lr16* than any of SSR markers tested. *XRGA266585* was also polymorphic between Tc and Tc*Lr16*, making this marker combination a good candidate for further evaluation on other populations and potentially for MAS. The use of *XRGA266585* does have some disadvantages. The marker is not codominant, so heterozygous and homozygous resistant individuals cannot be distinguished from each other. Also, because the progeny were scored on either the presence or absence of the band <100 bp, false negatives could easily be scored as susceptible progeny instead of resistant/heterozygous. Some of the PCRs had to be run

three separate times to ensure correct scoring of the progeny. The markers for *XRGA22128* may be less useful, partially because they are not codominant and amplify the susceptible allele. *XRGA22128* also mapped the furthest from *Lr16* of the five loci evaluated, and the markers were not polymorphic between Tc and Tc*Lr16*. In a TcxTc*Lr16* population, this marker most likely would not have been evaluated past the initial parental screen because of the lack of polymorphism between the two parents. The advantage of the *XRGA22128* marker was that a band was always amplified in CS and Tc*Lr16*, so the possibility of false negatives due to PCR not working was greatly reduced. The BLASTx results from the cloned *XRGA22128* fragment sequence also indicate that an RGA is being amplified for a cysteine-rich receptor-like protein kinase. Multiple resistance genes have been identified as receptor–like kinases including the rice gene *Xa21* for bacterial blight resistance and the barley resistance gene *Rpg1* for stem rust resistance (Song et al., 1995; Brueggeman et al., 2002). *Rpg1* also has NBS and LRR domains (Brueggeman et al., 2002). These results indicate that *Lr16* may have a similar structure. However, the polymorphic band amplified in CS DNA may instead be linked to a RGA in the CS genome that is not present in the Thatcher genome instead of the *Lr16* resistance gene.

SSRs have been widely used for marker-assisted selection due to their abundance, codominance, high rate of polymorphism, and ease of use due PCR (Kuleung et al., 2003). The SSRs *Xwmc661, Xwmc764, Xbarc35, Xbarc124, Xgwm210, Xwmc489, Xwmc382,* and *Xgwm614* had previously been found to be closely linked to *Lr16* in multiple recombinant inbred and doubled haploid populations (McCartney et al., 2005, Somers et al., 2004). Markers for the three loci used in this experiment, *Xwmc661, Xwmc764, Xbarc35,* were used to confirm the usefulness of the CSxTc*Lr16* F_2 population for mapping new RGA markers. The SSRs previously screened on a separate CSxTc*Lr16* F_2 population were disregarded for mapping in this study because they

did not show consistent polymorphisms in the population tested. That population was discarded because it did not fit the expected Mendelian ratios in the F_2 or F_3 generations (unpublished data). This may have skewed the results from the SSR marker analysis and made the markers appear less useful. Because the CSxTc*Lr16* population used for mapping in this study proved to be more useful, the disregarded SSR markers could be screened for polymorphisms on this population as well.

In the current population, the three SSRs mapped similarly to the order expected based on the previous mapping studies (Somers et al., 2004; McCartney et al., 2005). In this population, *Xwmc764* mapped well within the expected range of 1 to 9 cM found in the three populations mapped by McCartney et al. (2005) at 4.1 cM. *Xwmc764* also mapped closest to *Lr16* in the population used in this study, in two out of three populations mapped by McCartney et al. (2005), and in the microsatellite consensus map by Somers et al. (2004). The order of *Xbarc35* and *Xwmc661* varied depending on the population. The wheat microsatellite consensus map places *Xbarc35* between *Xwmc765* and *Xwmc661* (Somers et al., 2004). McCartney et al. (2005) only mapped *Xbarc35* on one population, and it was mapped proximal to *Xwmc764* and *Xwmc661*, which agrees with our mapping order. *Xwmc661* and *Xbarc35* were both outside of the expected 10 cM range, but this may have been due to less homogeneity in the F_3 intercross population versus a recombinant inbred or doubled haploid population. The peak at 197 bases in the *Xwmc764* capillary fragment analysis data closely matches the length of 198 bp reported by Somers and Isaac (2004) after the addition of the 18 bp M13 tail. The peaks in GeneMarker used for scoring varied more for *Xwmc661*. The expected size given by Somers and Isaac (2004) in the GrainGenes database is 226 for CS, so a peak at 244 should have been seen in CS after the addition of the 18 bases for the M13 tail. However, the size of the CS peak that was consistently

scored was 203 bases. The peak for Tc*Lr16* was closer to the expected size at 240 bases. This discrepancy may have been due to a difference in the Chinese Spring line used for the cross and may also account for the more distant mapping distance. The expected size for *Xbarc35* was not given in the GrainGenes database, but the sequence for the *Xbarc35* sequence tag site includes 608 bases. This is much larger than the 284 bp peak in CS used for this mapping study.

In conclusion, RNA-seq can be useful to develop new molecular markers that will map to a desired region. Two new molecular markers were developed for mapping *Lr16* through the use of RNA-seq and RGAs. Flanking markers are needed to most confidently ensure the incorporation of *Lr16* into breeding programs using MAS. We had hoped to identify flanking markers using the RNA-seq data, but both markers mapped proximal to *Lr16*. The use of SSRs closely linked to *Lr16* confirmed the utility of our mapping population and the new markers. The *XRGA266585* marker will more useful based on its close linkage to *Lr16* and because it is polymorphic between Tc and Tc*Lr16*. Future efforts should include screening this marker for polymorphisms on other populations containing *Lr16*.

Figure 2.1 Representative phenotypes at 10 days post inoculation of a) resistant plants with *Lr16* and b) susceptible plants.

a)

b)

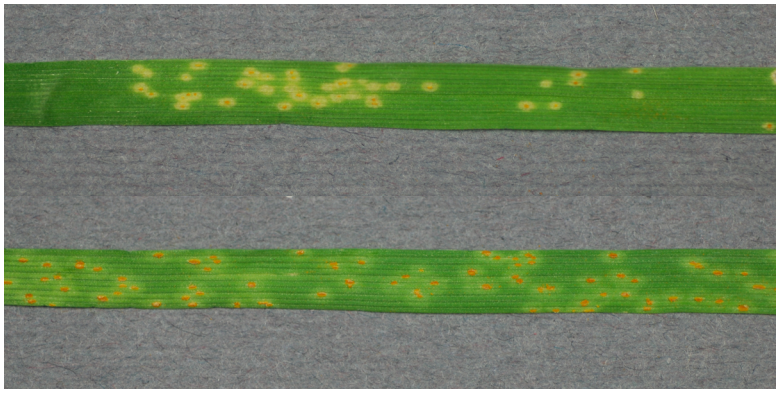


Figure 2.2 Gel images from PCR using DNA from 8 susceptible plants, CS (C), Tc*Lr16* (T), and 8 resistant plants on 2% agarose, high-resolution gels using the markers a) RGA22128, b) RGA266585, and c) RGA142887. Size standard (L) is a 100 bp ladder.

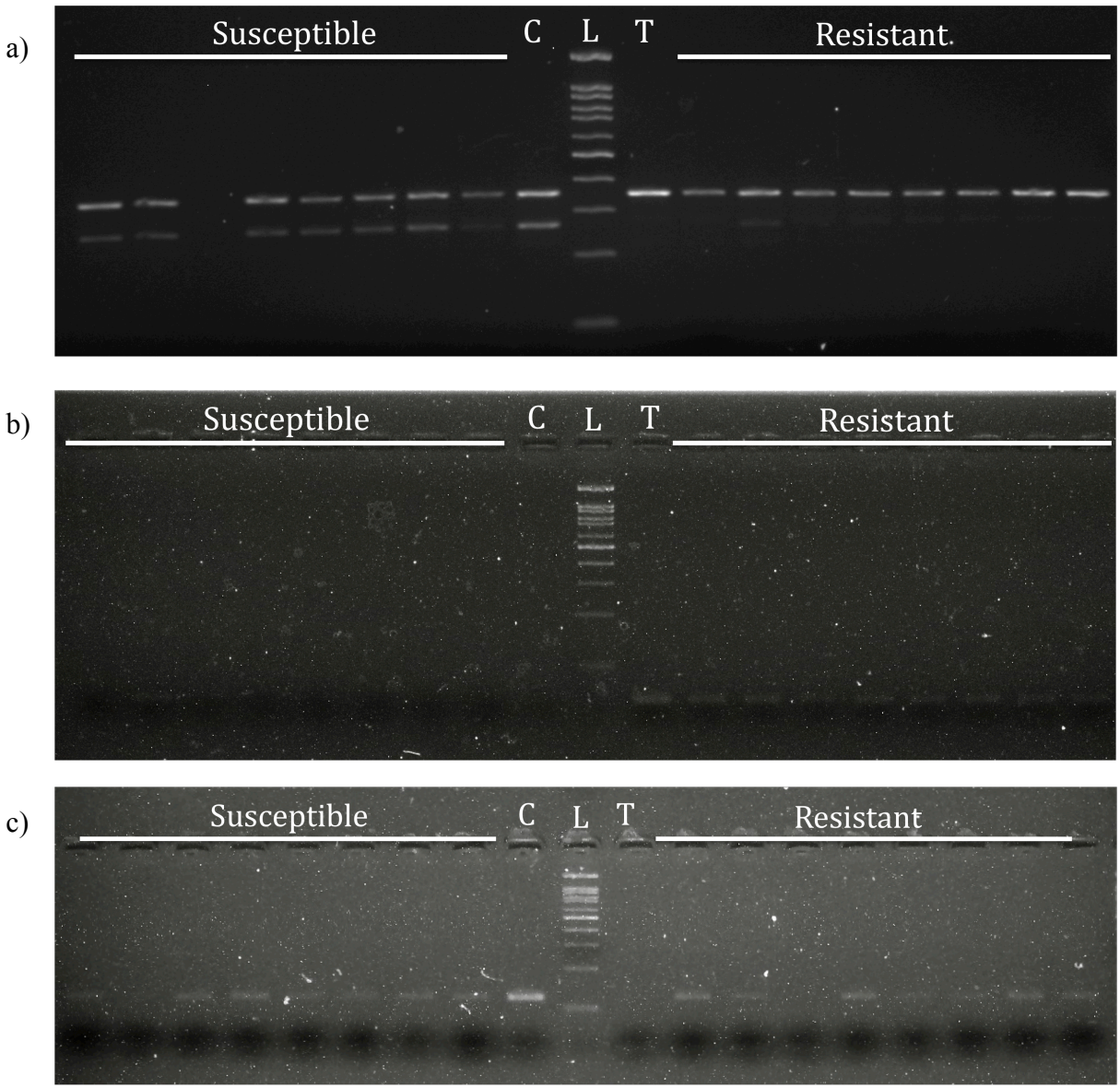
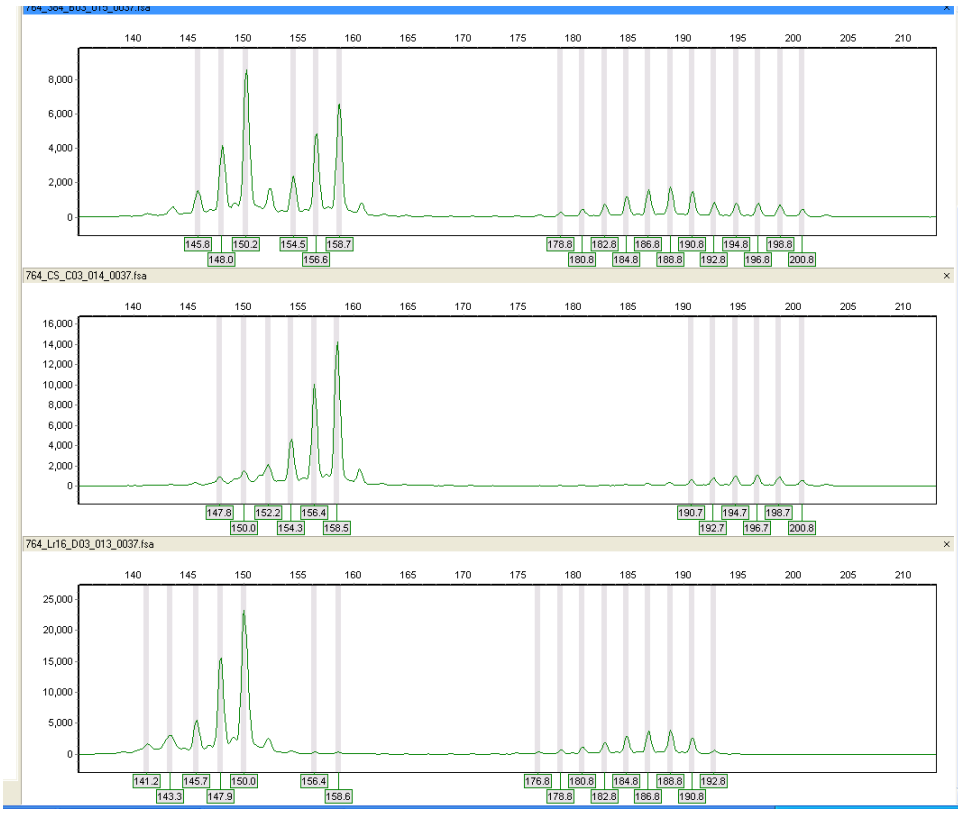
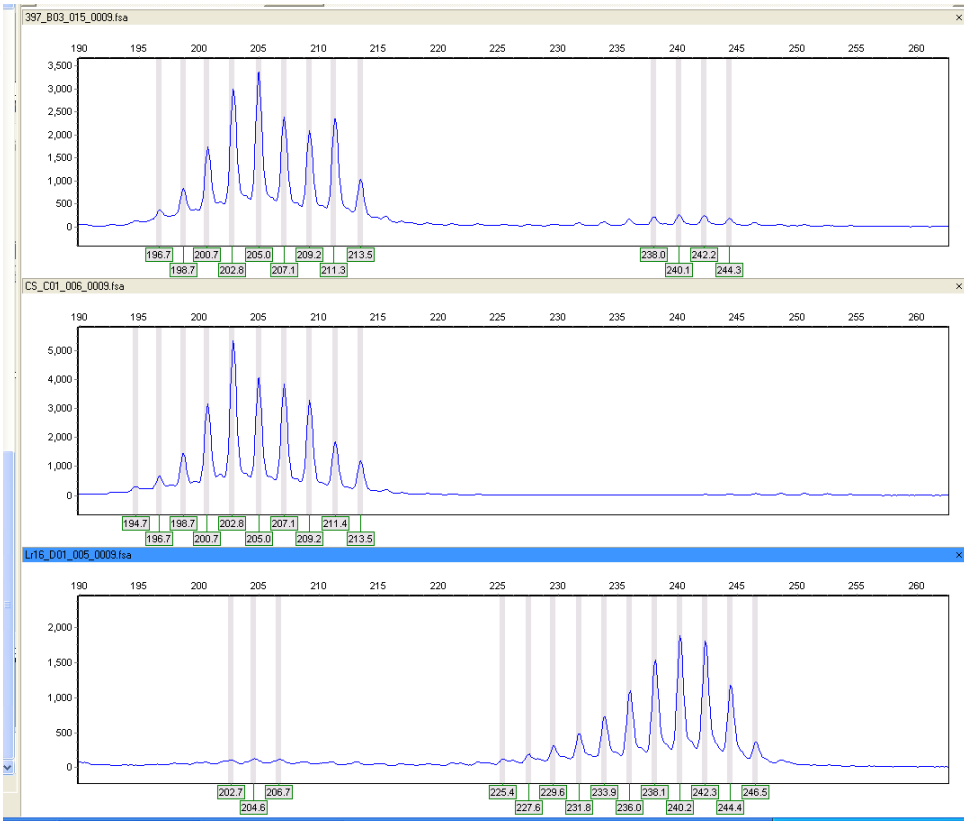


Figure 2.3 GeneMarker alleles heterozygous (top), susceptible (middle), and resistant (bottom) individuals for a) *Xwmc764*, b) *Xwmc661*, and c) *Xbarc35*.

a)



b)



c)

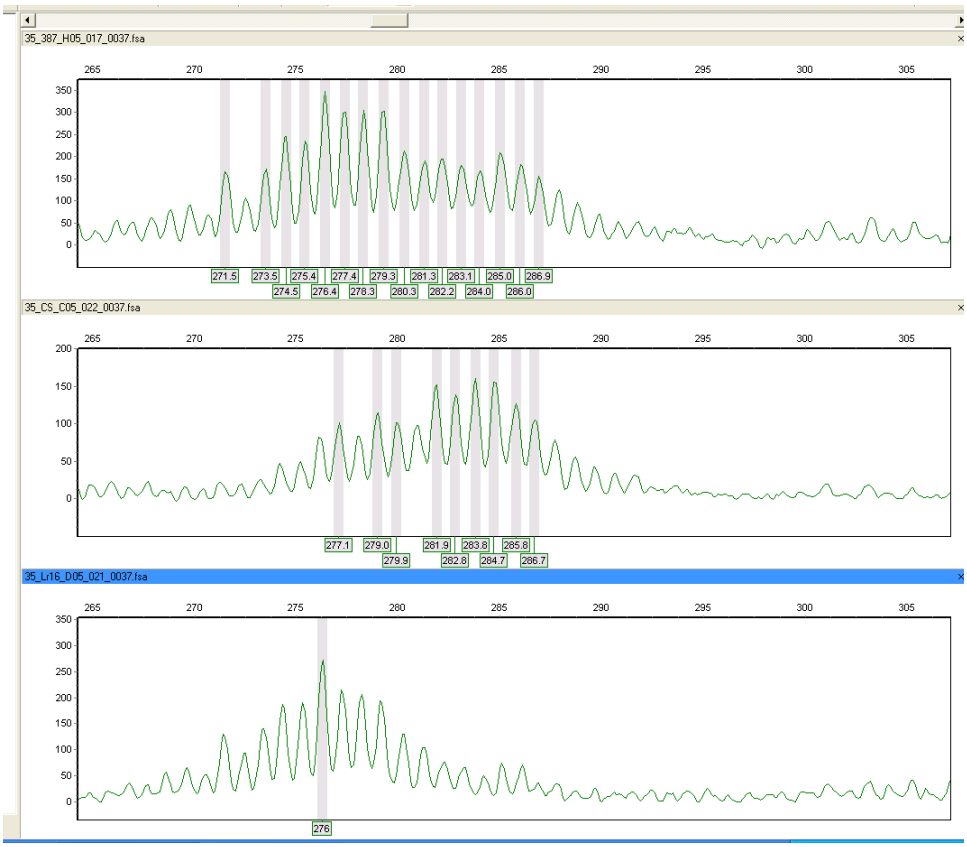
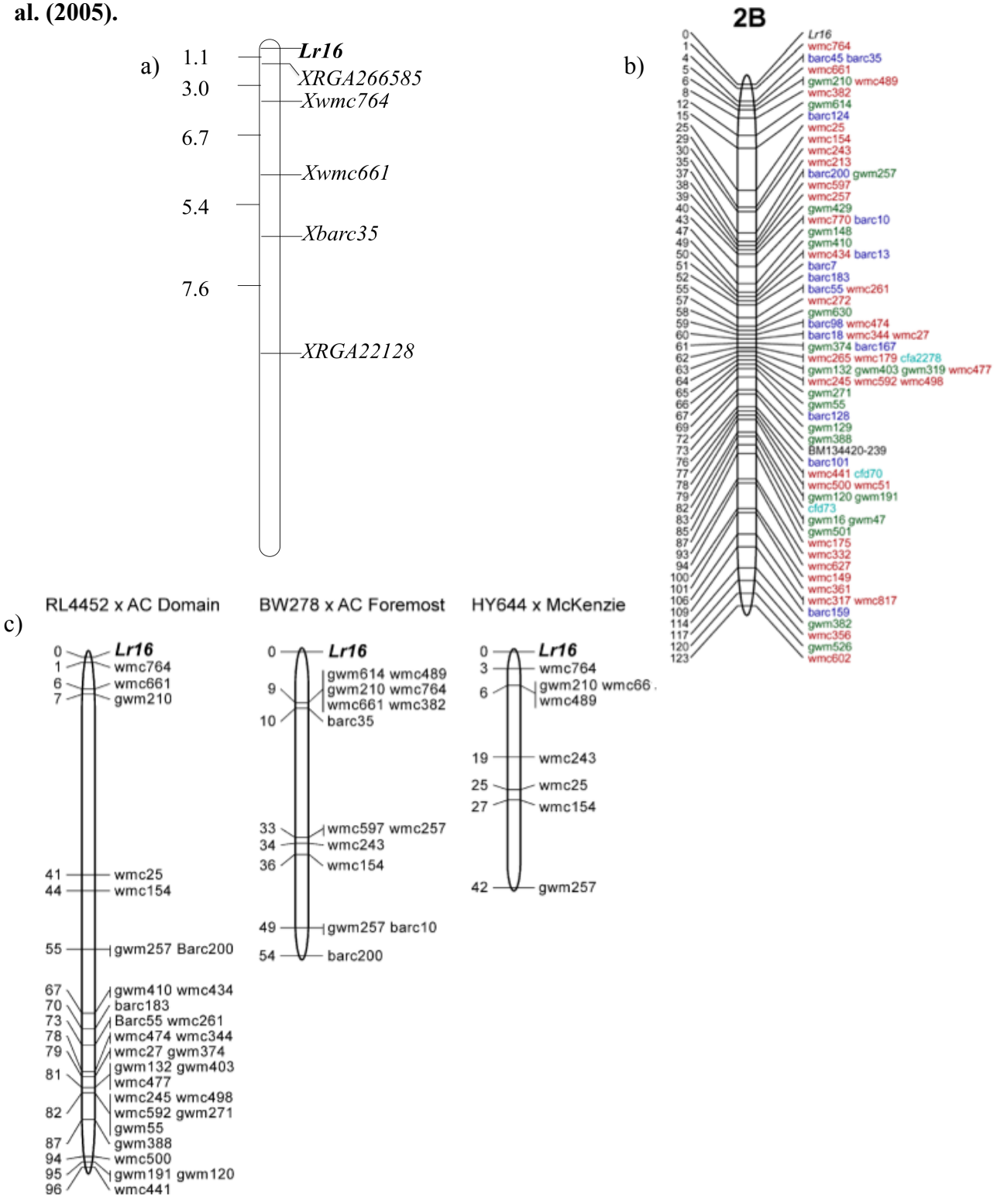

35_387_H05_017_0037.fsa
35_CS_C05_022_0037.fsa
35_Lr16_D05_021_0037.fsa

Figure 2.4 Linkage maps for a) SSR and RGA markers near *Lr16* in the CSxTc*Lr16* population, b) SSR markers on chromosome 2B by Somers et al. (2004), and c) SSR markers near *Lr16* in the recombinant inbred and doubled haploid lines by McCartney et al. (2005).



Chapter 3 - Using Genotyping-by-Sequencing to identify SNPs and KASP markers for *Lr16*

Introduction

Leaf rust, caused by *Puccinia triticina*, is one of the most widespread fungal diseases of wheat worldwide (Neelam et al., 2013). Leaf rust primarily affects *Triticum* species, including durum, emmer, and *triticale* spp. (Neu et al., 2003). Losses due to leaf rust can range from 1 to 20% nationwide in wheat (Kolmer, 2013). Breeding for leaf rust resistance is one of the most effective methods of control (Krattinger et al., 2009). Traditional breeding relies on the use of phenotypic evaluation and the use of effective pedigrees to incorporate resistance into cultivars (Oelke and Kolmer, 2004). The use of perfect or closely linked molecular markers for resistance genes can greatly facilitate the incorporation of desired genes into breeding programs (Dholakia et al., 2013).

Lr16 is a leaf rust resistance gene that provides partial resistance at the seedling stage (McCartney et al., 2005). *Lr16* originated in hexaploid bread wheat and is commonly found in hard red spring varieties grown in the northern Great Plains (McIntosh et al., 1995). Resistance to *Lr16* has not been completely overcome, and as resistant varieties containing *Lr16* become less prevalent in the field, the pathogen population becomes less virulent to the gene. This indicates that *Lr16* may be a good candidate for gene recycling due to a fitness cost in the pathogen population associated with virulence on *Lr16* (Marshall, 1992). The *Lr16* locus has been mapped to the distal end of chromosome 2BS using microsatellites. The closest marker mapped 1, 3, and 9 cM away from the locus depending on the population used (Somers et al., 2004; McCartney et al., 2005).

New developments in next-generation sequencing technologies have enabled the discovery, sequencing, and genotyping of thousands of markers in hundreds of individuals (Davey et al., 2011). The use of restriction enzymes has proven to be useful for the discovery of new molecular markers. Amplified fragment length polymorphisms (AFLPs) and restriction fragment length polymorphisms (RFLPs) have been widely used as molecular markers, and both methods rely on restriction enzymes to create a reduced representation of the genome (Vos et al., 1995; Karp et al., 1996). However, the success rate of AFLPs and RFLPs for marker-assisted selection (MAS) in wheat has been relatively low due to the limited level of polymorphism in the wheat genome (Song et al., 2005). The use of microsatellites and, more recently, single nucleotide polymorphisms (SNPs) as molecular markers has proven to be more successful in wheat (Somers et al., 2004; Poland et al., 2012b). SNPs have the best potential as molecular markers due to their abundance in genomes and because they are polymorphic between wheat varieties (Wicks et al., 2001; Allen et al., 2013). One of the first instances of the use of next-generation sequencing for SNP discovery in large and complex genomes allowed for the discovery of 195,631 putative SNPs in gene sequences, 155,580 putative SNPs in uncharacterized single-copy regions, and 145,907 putative SNPs in repeat junctions in the *Aegilops tauschii* genome (You et al., 2011).

Restriction-site associated DNA (RAD) tagging was one of the first methods of SNP discovery developed that uses restriction enzymes to produce a reduced representation of the target genome by targeting the sequences flanking the restriction sites (Miller et al., 2007). RAD tags are short fragments of DNA that represent every site of particular restriction enzyme at a genome-wide level (Baird et al., 2008). RAD markers were developed as a low-cost microarray genotyping resource and as a way to screen almost every restriction site of a particular enzyme in

parallel instead of just a small subset (Miller et al., 2007). The RAD genotyping platform was further advanced by the development of a nucleotide barcoding system for sample multiplexing and its use with next-generation sequencers (Baird et al., 2008).

Genotyping-by-sequencing (GBS) is similar to RAD tagging and uses restriction enzymes to produce a reduced representation of the genome being analyzed. However, many of the steps are more straightforward and less complicated. For example, the genomic DNA digestion and adapter ligation are performed in a single well resulting in less sample handling. In addition, DNA samples do not have to be sheared and size selected, and there are fewer DNA purification steps (Elshire et al., 2011). The initial GBS protocol described by Baird et al. (2008) uses a single restriction enzyme. Poland et al. (2012a) presented a modified protocol that uses both a "rare-cutter" and a "common-cutter" restriction enzyme along with Y-adapters. Library quantification is greatly simplified in this protocol due to all amplified fragments consisting of the barcoded forward adapter and the common reverse adapter (Poland et al., 2012a).

Even with simplified protocols, genome-wide SNP discovery through the use of next-generation sequencing (NGS) platforms has proven to be challenging in plants with large, complex, and highly repetitive genomes such as wheat, barley, and rye. Alignment and clustering of short reads is especially difficult due to the large number of repetitive sequences generated (You et al., 2011). The need for consistent DNA quantification for the GBS protocol also presents a challenge. A precise and high-throughput DNA quantification system will reduce sample-to-sample variation in sequence coverage and limit quantification as a procedural bottleneck (Elshire et al., 2011). Analysis of the SNP markers requires a cost-effective SNP genotyping assay to implement the SNPs for linkage analysis (Cortés et al., 2011). Several assays are available including, but not limited to, TaqMan and KASPar assays (Shen et al., 2009;

Semagn et al., 2013). The KASPar assay is a PCR-based method for SNP genotyping that uses fluorescently labeled SNPs (Hiremath et al., 2012). The KASPar assay provides the opportunity for high-throughput genotyping and the ability to eliminate the post-PCR handling necessary with other markers (Neelam et al., 2013).

KASPar assays have successfully been implemented in wheat in multiple experiments. Allen et al. (2011) reported the first linkage map of wheat consisting of several hundred SNPs. This was also the first experiment in which KASPar technology was used to genotype wheat varieties and to generate a linkage map. Validation of a subset of putative varietal SNPs revealed a conversion rate of ~67% using the KASPar genotyping platform (Allen et al., 2011). A KASPar assay was developed to facilitate marker-assisted deployment of the leaf rust resistance gene *Lr21*. *Lr21* has been cloned and still provides effective resistance in the field. After aligning sequence information from the *Lr21* locus of 67 North American Spring wheat cultivars and 10 winter wheat cultivars, a SNP was detected in the NBS region of *Lr21* and was targeted for use in the KASPar assay. The marker designed from this SNP could successfully be used to differentiate between susceptible and resistant genotypes (Neelam et al., 2013). The KASPar assay has also been used in chickpea. Four different approaches were used to identify 2,486 SNPs, from which 2,005 KASP assay markers were validated (Hiremath et al., 2012).

Objective

Genome-wide SNP discovery in wheat has proven to be particularly difficult due to the large, complex, polyploid nature of wheat. The development of new sequencing technologies has more readily allowed for the identification of SNPs in many species, especially through the use of restriction enzymes to create reduced-representation libraries. The use of barcodes for multiplexing further increases the utility of this approach by allowing sequencing of multiple

samples in a single run. Genotyping-by-sequencing allows discovery of SNPs without the need for cloning or previously developed markers. This is largely possible through the use of restriction enzymes to create a reduced representation of the genome in combination with next-generation sequencing technologies. The UNEAK and Tassel pipelines have been developed to process the GBS data into SNP genotypes. KASP assays can then be used to validate and map SNPs discovered. The objective of this research was to use genotyping-by-sequencing and KASP assays to identify SNP markers closely linked to the leaf rust resistance gene *Lr16*.

Materials and Methods

Plant Material and Growth Conditions

Two separate crosses were used in this experiment. The first cross was made between two spring wheat (*Triticum aestivum*) lines, Chinese Spring (CS) and the Thatcher (Tc) isoline containing *Lr16* (RL6005). A second cross was made between Tc and the RL6005 Tc isoline (Tc*Lr16*). In both populations, the F_1 generation was self-pollinated to produce the F_2 population. Individual F_2 seeds were germinated in 3.8-centimeter (cm) cone-tainers filled with Metro-Mix 360 (Sun-Gro Horticulture, Vancouver, Canada). The CSxT*cLr16* F_2 seedlings were later transferred to one-gallon pots filled with Metro-Mix 360 at a density of two plants per pot. The TcxTc*Lr16* F_2 plants were not transferred from the cone-tainers. Plants were kept in a greenhouse with supplemental high-pressure sodium lighting set for 16-hour days and eight-hour nights. Daytime temperatures were set at 20°C, and nighttime temperatures were 18°C. Plants were grown until maturity, dried down, and F_3 seed harvested.

Plant Inoculation and Screening

F_2 plants of both populations were inoculated with Race 1, BBBD of *P. triticina* at the seedling 2-3 leaf stage. Urediniospores suspended in Soltrol 170 (Phillips Petroleum,

Bartlesville, OK) were applied with an atomizer at 20 psi. The plants were left overnight in a mist chamber with 100% relative humidity at 18°C. Plants were moved back into the greenhouse, and infection types were evaluated at 10-14 days post inoculation and scored as either resistant or susceptible. Plants were considered to be resistant if they displayed a low infection type consisting of the presence of a small pustule surrounded by a distinct yellow halo (Figure 3.1a). High infection types consisted of large pustules with no halos and were considered susceptible phenotypes (Figure 3.1b). $F_{2:3}$ families were evaluated for their infection types to determine the zygosity of the F_2. Sixteen seeds from each line were planted in 8.9 cm pots containing Metro-Mix 360. These families were inoculated with leaf rust Race 1, BBBD and scored 10-14 days after inoculation as either resistant, heterozygous, or susceptible. Chi-square (χ^2) analysis ($P < 0.05$) was used to test for goodness of fit to the expected Mendelian ratio of 3:1 in the CSxTc*Lr16* and TcxTc*Lr16* F_2 populations and 1:2:1 in the F_3 populations.

DNA Isolation

DNA isolation was performed using the same protocol outlined on pages 36 and 37 of Chapter 2 from F_2 lines from both the CSxTc*Lr16* and TcxTc*Lr16* populations.

DNA Quantification

DNA samples were initially quantified using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE) and diluted to 100 ng/µL in ddH_2O. Diluted DNA from both parental lines along with DNA from 22 plants that scored resistant and 22 plants that scored as susceptible in the CSxTc*Lr16* F_2 and F_3 populations was quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies, Carlsbad, CA). Standard samples were prepared using λ-DNA and 1x TE in a serial dilution at DNA concentrations of 100, 80, 60, 40, 30, 20, 10, and 0 ng/ µL. In a separate plate, 5 µL of diluted DNA was added to 95 µL of 1x TE.

One hundred µL of a mix of 10 mL 1x TE and 85 mL PicoGreen was added to each well of the standard samples and DNA samples, mixed, and covered to protect the plates from light. DNA concentrations had to be between 20 ng/µL and 150 ng/µL to accurately quantify DNA to 200 ng in a total volume of 20 µL.

Genotyping-by-Sequencing

DNA from 22 resistant plants, 22 susceptible plants, and one of each parent from the CSxTc*Lr16* F_2 population was used to create the genotyping-by-sequencing (GBS) library. A double-digest with the restriction enzymes *Pst1* and *Msp1* was used to make the GBS library as described by Poland et al. (2012a). The *Pst1* overhang corresponds to the barcoded forward adapter, Adapter 1 (Figure 3.2a), and the *Msp1* overhang corresponds to the common Y-adapter, Adapter 2 (Figure 3.2b; Poland et al., 2012a). Two hundred ng (20 µL at 10 ng/µL) of genomic DNA was digested in 10.0 µL of restriction master mix consisting of 2.0 µL 10x NEB Buffer 4, 8 U (0.4 µL) *Pst1*-HF (High-Fidelity), 8 U (0.4 µL) *Msp1* (New England BioLabs Inc., Ipswich, MA) and 7.2 µL H_2O. The digestion was conducted at 37°C for 2 hours, 65°C for 20 minutes in a Bio-Rad Dyad Peltier Thermal Cycler (Bio-Rad Laboratories, Hercules, CA). Adapters were ligated to individual samples using a ligation master mix prepared with 1x (2.0 µL 10x) NEB Buffer 4, 100 mM (0.4 µL) 1 ATP, 200 U (0.5 µL) T4 DNA ligase, 12.1 µL H_2O per sample for a total volume of 15 µL per sample. The 15 µL of restriction master mix was mixed with 20 µL of the restriction digest, 0.1 pmol Adapter 1, and 15 pmol Adapter 2 (5 µL of adapters). The ligation was completed at 22°C for 2 hours and 65°C for 20 minutes in a Bio-Rad Dyad Peltier Thermal Cycler.

The multiplex library was prepared by pooling 5 µL from each sample ligation into a single tube with 20 µL of ddH_2O. Two cleanups were performed on the library using the

QIAquick PCR Purification Kit (Qiagen, Valencia, CA). In two separate tubes, 200 µL of pooled ligation DNA was combined with 1000 µL of Buffer PB, added in increments of 600 µL to spin columns, and spun down in a MIDSCI microcentrifuge (MIDSCI, St. Louis, MO). The rest of the cleanup followed steps 6-9 of the QIAquick PCR Purification Kit Protocol using a microcentrifuge from the QIAquick Spin Handbook, version 03/2008 (Qiagen). The protocol was modified slightly to elute the DNA in 45 µL of Buffer EB instead of 50 µL. After the cleanup protocol was completed, the flow-through from both tubes was combined into one tube and used for library amplification. Six PCR reactions were made for each library that consisted of 10 µL cleaned and digested DNA library, 5 µL NEB MasterMix, 2 µL of 10 µM Illumina F and R paired-end primers, and 8 µL H_2O. A Bio-Rad Dyad Peltier Thermal Cycler was used for selective amplification and programmed with the following conditions: 95°C for 30 seconds; 16 cycles of 95°C for 30 seconds, 62°C for 20 seconds, and 68°C for 90 seconds; and 72°C for 5 minutes. A short extension time is used to enrich for fragments in the 200-500 bp range. All six PCR reactions were pooled, and the second cleanup was completed. A modified version of the protocol on page 7-8 was used. A volume of 750 µL Buffer PB was added instead of 1,000 µL, and the DNA was resuspended in 30 µL Buffer EB. After the library was quantified, samples were submitted to the University of Missouri for sequencing. The libraries were sequenced on an Illumina HiSeq 2000 sequencing system.

GBS Data Analysis

The raw Illumina data was processed following the protocol outlined in Poland et al. (2012a). The barcode sequence was used to assign unfiltered qseq data to individual samples. The sequences were then trimmed to 64 bp using a custom Java script (maizegenetics.net; sourceforge.net/projects/tassel/). The UNEAK and Tassel pipelines were used to process the raw

GBS sequence data into SNP genotypes (Lu et al., 2013; Glaubitz et al., 2014). Because the Tassel pipeline is designed for species with a reference genome, custom java script was used to finish the SNP discovery. Custom R script was used align sequence tags from both pipelines (Appendix C). In brief, data from the either the UNEAK or Tassel pipeline and the phenotypic data were imported into R and used to make "geno" and "pheno" matrices. The two matrices were then merged. The two alleles for each SNP were designated as allele1 or allele2. A true/false array was used to compare the phenotypes with allele1 and allele2. SNPs were filtered out based on the number of mismatches allowed in the true/false array. The allowable mismatch for the UNEAK pipeline data was 0 mismatches per 64 bp sequence and was 0 to 1 mismatches per 64 bp sequence for the Tassel pipeline data.

KASP Marker Design

The blastn algorithm was used to search the National Center for Biotechnology Information (NCBI) database for RGA matches to the sequence tags with the allowable number of mismatches. BLAST was also used to search for matches between the sequence tags and the Chinese Spring draft genome assembly (Wilkinson et al., 2012). If a match was identified with either algorithm, the FASTA sequence from the contig or accession with the best match was used to design Kompetitive Allele Specific PCR (KASP; LGC Genomics, Beverly, MA) primers. The FASTA sequence was searched for the 64 bp fragment to ensure the SNP was included, and at least 50 bases on both sides of the SNP were imported into MacVector (MacVector, Inc., Cary, NC). Two allele-specific forward primers were designed with FAM (GAAGGTGACCAAGTTCATGCT) and HEX (GAAGGTCGGAGTCAACGGATT). A common reverse primer was also designed. All three primers were designed with a length between 18 and 25 nucleotides excluding FAM and HEX, an average Tm between 50°C and

70°C but as close to 65°C as possible, and with less than 5° difference in the Tm. The SNP was positioned at the 3' end of the primer and the sequence for FAM or HEX was on the 5' end (Appendix D).

KASP Genotyping Assay

DNA from the TcxTc*Lr16* F_2 population was used for the KASP genotyping assays. Each assay was comprised of KASP V4.0 2x Mastermix (LGC Genomics) and 72x primer assay mix. To make the 72x primer assay mix, dry primers were first diluted to 100 µM using Tris (10 mM, pH 8.0). The two allele-specific forward primers and common reverse primer were all pooled with final concentrations of 12 pm/µL for each of the two forward primers and 30 pm/µL for the reverse primer in the assay mix. Each reaction contained 80 ng genomic DNA and 1x concentrations of both the primer assay mix and KASP reaction mix in a total reaction volume of 7.8 µL. A CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad Laboratories) was used for selective amplification of the KASP assay and was programmed with the following conditions: 94°C for 15 minutes, 36 cycles of 94°C for 10 seconds and 60°C for 1 minute, and 35°C for 30 seconds.

Markers were originally screened on a total of 15 DNA samples including three resistant, three heterozygous, three susceptible progeny along with three replicates each of Tc and *TcLr16* DNA. Three non-template controls (NTC) consisting of the same ratio of ddH_2O, Buffer TE, and KASP mix as the DNA samples were included to determine background level of fluorescence. F_2 and F_3 phenotypes were used to determine which progeny to include. If the KASP assay results showed distinct clusters of allele 1, allele 2, and heterozygous SNP calls that matched the phenotype, the assay was expanded to include samples from the entire population. Multiple runs were performed to ensure the best accuracy of the allele call. KASP markers were

used for mapping if the distinct clusters for allele 1, allele 2, and heterozygous samples were still evident.

Linkage Analysis

KASP marker data was mapped using MapMaker V3.0 (Lander et al., 1987). The RAW file was created as an F_2 intercross with 144 progeny, two genetic loci, and zero quantitative traits. Multiple parental samples and the NTCs were included, which accounted for the additional progeny. The 'cent func k' command was used to change the mapping function from Haldane to Kosambi mapping units. Linkage groups were formed with a LOD threshold of 3.00. The sequence and map functions were used to determine the distance between the two loci.

Results

Lr16 Segregation

In the TcxTc*Lr16* mapping population, 129 plants were scored in the F_2 population. Ninety-one of the plants were scored as resistant and 38 were susceptible. The chi-square value was 1.37 for a P value between 0.50 and 0.20 at one DF. The null hypothesis failed to be rejected because the population fits the expected 3:1 ratio. Out of 126 families scored in the F_3 population, 39 were resistant, 62 were heterozygous, and 25 were susceptible. The population had a chi-square value of 3.14 and a P value between 0.50 and 0.20 at two DF. Because this population fits the expected 1:2:1 ratio, the null hypothesis failed to be rejected.

SNP Discovery, Genotyping, and Linkage Analysis

A total of 63 putative SNPs were detected using both the Tassel and UNEAK pipelines. Fewer SNPs were detected using the Tassel pipeline, so 0 to 1 mismatches were allowed. Thirty-eight putative SNPs were identified from three output files using the Tassel pipeline, and ten of

these had 0 mismatches and were perfectly correlated with the phenotype. The other 28 SNPs identified had only a single mismatch, meaning the phenotype from one of the 44 plants did not correlate with the SNP call. The UNEAK pipeline identified a greater number of putative SNPs, and 25 were identified when allowing for 0 mismatches and perfect correlation. Not all of the SNPs from both pipelines were unique, and 24 were duplicates of SNPs already found either in another pipeline or another Tassel output file. A total of 39 primer combinations were designed from all of the unique SNPs identified. Only two primers, *XSNP16_155* and *XSNP16_TP1456*, showed distinct clusters for allele 1, allele 2, and heterozygous samples and were screened on the entire population. Only the *XSNP16_TP1456* marker showed the same clusters when screened on the entire population and was used for mapping (Figure 3.3). Three separate KASP assay runs with the *XSNP16_TP1456* marker were analyzed to determine a consensus genotype for each $F_{2:3}$ line to be used for mapping. Based on this information and the F_3 phenotypes, the *XSNP16_TP1456* marker mapped 0.7 cM from the *Lr16* locus (Figure 3.4). No blastn hit were detected for the 64 bp TP1456 sequence tag, but a 97% match to 45 out of 46 matches was detected in CS.

Discussion

Even though the Tassel GBS pipeline is designed for species with a reference genome, custom java script could be used to complete the pipeline and identify putative SNPs (Glaubitz et al., 2014). The UNEAK pipeline was designed for species like wheat without a reference genome and could be completed following the protocol in the reference material (Lu et al., 2013). Some duplicate SNPs were detected, but both pipelines proved to be useful for detecting unique putative SNPs. Discovery of 39 perfectly correlated SNPs from the two pipelines indicates that GBS can be useful in the identification of SNPs linked to resistance genes.

The KASP assay was less successful because only one SNP could be validated. This is most likely a result of running the KASP markers on a TcxTc*Lr16* population instead of a CSxTc*Lr16* like that from which the markers were designed rather than how the KASP assays were set up and run. Many of the SNPs identified may be associated with polymorphisms in the Chinese Spring background but still closely linked to *Lr16*. Screening of the other KASP markers on the CSxTc*Lr16* population discussed in Chapter 2 may also allow for validation of other SNPs because the background would be the same as that used for the GBS run. The TcxTc*Lr16* mapping population would have fewer polymorphisms detectable with the KASP assay due to both parents having the same background. The Kansas State Integrated Genomics Facility optimized the amplification program used in this experiment for a Bio-Rad CFX96. However, the KASP assay reaction set up used in this experiment differs slightly from that used by the Kansas State Integrated Genomics Facility. The Integrated Genomics Facility uses a volume of 4.11 μL, instead of 3.78 μL, of primer mix and 2x reaction mix. Even with the differences a SNP closely linked to *Lr16* was discovered.

The SNP in the *XSNP16_TP1456* marker was identified by allowing 0 mismatches between the phenotypes and genotypes in the UNEAK pipeline. This same SNP was also identified in the Tassel pipeline with 0 mismatches. *XSNP16_TP1456* has a C/T SNP at nucleotide 17. The *XSNP16_TP1456* marker appears to be the most closely linked marker to *Lr16* to date. However, no other markers were run on the population to validate the KASP marker. Other KASP markers from the Cerealsdb database may be used to determine if *XSNP16_TP1456* is more closely linked than other known SNPs. Mapping of known SSRs on the same population would confirm the utility of this TcxTc*Lr16* population and marker as well. The *XSNP16_TP1456* marker could also be mapped on the CSxTc*Lr16* mapping population

discussed in Chapter 2. This would allow the SNP to be mapped with other known markers to see if it truly does map closer to *Lr16* than other known markers.

The number of useful KASP markers identified in this study is not as high as originally hoped. However, the success of validating one SNP as a marker for *Lr16* indicates that this methodology does have the potential to be useful with further research and optimization. Because the SNPs were identified using a CSxTc*Lr16* mapping population, it makes sense that the SNP marker validated in the TcxTc*Lr16* population would be closely linked to *Lr16*. Other SNPs would most likely be associated with differences in the CS background and may not be as closely linked to *Lr16* as *XSNP16_TP1456.*

Figure 3.1 Representative phenotypes at 10 days post inoculation of a) resistant plants with *Lr16* and b) susceptible plants.

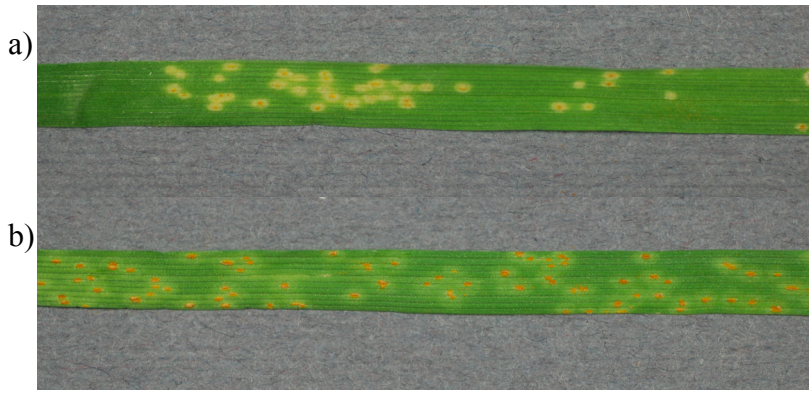


Figure 3.2 Adapters used to create the GBS library.

a) The barcoded forward adapter (Adapter 1) with a *Pst*I restriction site after the barcode. The barcode is designated by XXXXX, and genomic DNA is designated by NNNNNNN.

b) The barcoded reverse adapter (Adapter 2) with an *Msp*I restriction site. Genomic DNA is designated by nnnnnnnn. The unpaired tail of the Y-adapter is underlined.

a) 5’ CACGACGCTCTTCCGATCTXXXXXTGCA GNNNNNNN 3’

3’ GTGCTGCGAGAAGGCTAGAXXXXX TGCACNNNNNNN 5’

b) 5’ nnnnnnnnC CGAGATCGGAAGAGCGGGGACTTTAAGC

3’ nnnnnnnnGGC TCTAGCCTTCTCGCCAAGTCGTCCTTACGGCTCTGGCTAG

Figure 3.3 Scatterplot of RFU values from the *XSNP16_TP1456* genotyping assay

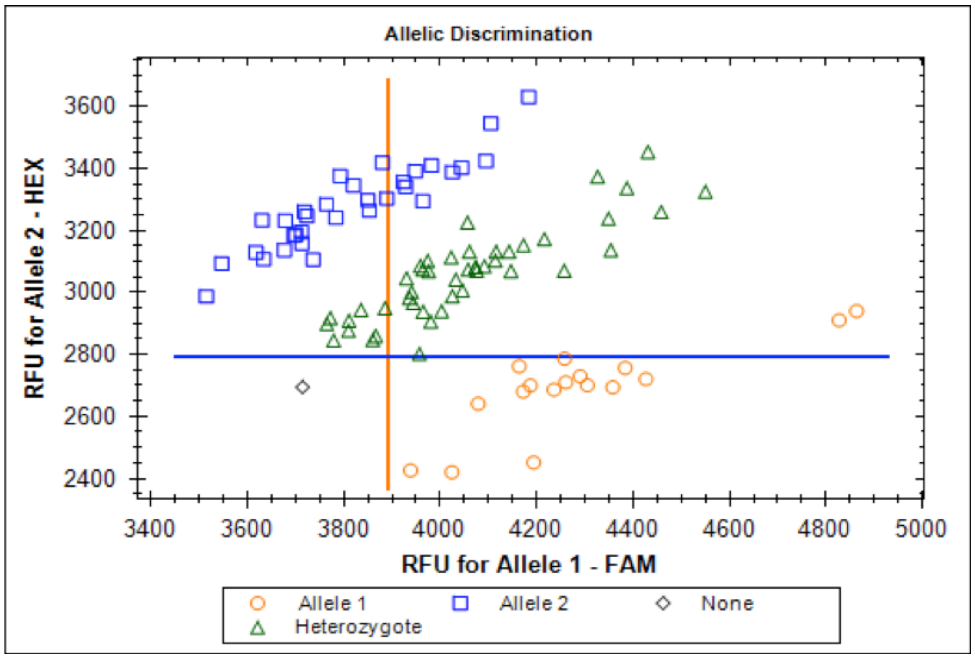
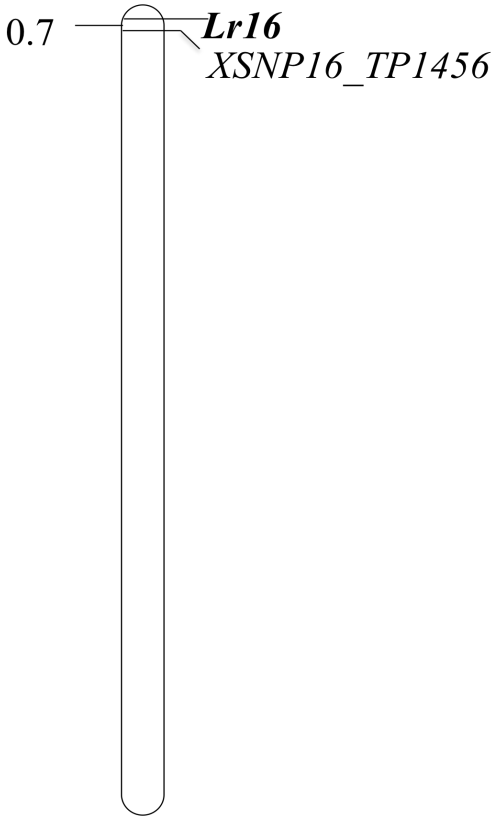


Figure 3.4 Linkage map of *XSNP16_TP1456*



Chapter 4 - Summary

Ideally either flanking markers or a marker perfectly correlated with *Lr16* would have been identified using either of the two approaches outlined in Chapter 2 and Chapter 3. Neither a flanking marker or a perfect marker was identified, but three markers linked to *Lr16* were identified. Two of these linked markers identified are resistance gene analog (RGA) markers developed through the use of RNA-seq and *in silico* subtraction. In this population, *XRGA266585* was actually more tightly linked to *Lr16* than any of the SSR markers tested. *XRGA22128* was the most distantly linked marker from *Lr16* but was still linked to the gene. The third marker developed was a Kompetitive Allele Specific PCR (KASP) marker developed through the use of genotyping-by-sequencing (GBS) on an F_2 population. Being able to use an F_2 population for GBS can save time in population development and reduce the amount of time necessary to complete a project. Based solely on the distances in their respective mapping populations, the KASP marker identified was even more closely linked than any other marker tested and appears to be the most closely linked marker to date. However, running this marker with the other five RGA and SSR markers on the same CSxTc*Lr16* population used in Chapter 2 will be necessary to validate this. Overall, both techniques are viable methods for developing new molecular markers. The results of these experiments indicate that RNA-Seq and *in silico* subtraction can be used to develop RGA markers and that GBS can be used to develop KASP markers linked to a particular disease resistance gene.

The methods and markers developed in these protocols could be used for several future projects. Both the RGA markers and the KASP marker that were mapped should be screened for polymorphisms on other populations containing *Lr16*. Also, most of the RGA markers

developed from the Tc*Lr16* isolines were not linked to *Lr16*. The markers could be screened on different mapping populations to determine if they may be linked to a different resistance gene. The KASP markers developed should also be screened on the CSxTc*Lr16* mapping population used in Chapter 2 because it is very similar to the population used for the GBS run and for identification of single nucleotide polymorphisms (SNPs). More of these markers may be polymorphic on this population and could be used for mapping. The *XSNP16_TP1456* KASP marker should also be mapped on this same CSxTc*Lr16* F_2 population to determine how closely linked the marker is in this population and to validate it with the other RGA and SSR markers. Both projects utilized Thatcher isolines, so developing other crosses using Thatcher and a Thatcher isoline with a different resistance gene introgressed could be useful for developing RGA markers linked to other resistance genes through the use of RNA-seq and *in silico* subtraction. These crosses could also be used for GBS to identify SNPs associated with a particular disease resistance gene and to develop KASP markers from these SNPs. Also, the techniques used in this experiment are not limited to wheat, so both methods could be expanded to other species.

References

AC't Hoen, P., Y. Ariyurek, H. H. Thygesen, E. Vreugdenhil, R. Vossen, R. X. de Menezes, J. M. Boer, G. B. van Ommen, and J. T. den Dunnen. 2008. Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucl. Acids Res.* 36: e141-e141.

Akın, B., S. Yüce, R. Sing, H. Braun, N. Zencirci, and A. Morgunov. 2013. Leaf rust (Puccinia triticina) resistance in winter-facultative wheat (Triticum aestivum L.) cultivars from different countries. *Open Science Repository Agriculture* open-access: e23050406.

Allen, A. M., G. L.A. Barker, S. T. Berry, J. A. Coghill, R. Gwilliam, S. Kirby, P. Robinson et al. 2011. Transcript‐specific, single‐nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (Triticum aestivum L.). *Plant biotechnology journal* 9: 1086-1099.

Allen, A. M., G. L.A. Barker, P. Wilkinson, A. Burridge, M. Winfield, J. Coghill, C. Uauy, S. Griffiths, P. Jack, S. Berry, P. Werner, J. P. E. Melichar, J. McDougall, R. Gwilliam, P. Robinson, K. J. Edwards. 2013. Discovery and development of exome‐based, co‐dominant single nucleotide polymorphism markers in hexaploid wheat (Triticum aestivum L.). *Plant biotechnology journal*, 11: 279-295.

Altieri, E., B. McCallum, D. J. Somers, and F. Sacco. 2008. Inheritance and genetic mapping of leaf rust resistance genes in the wheat cultivar Buck Manantial. In *The 11th International Wheat Genetics Symposium proceedings Edited by Rudi Appels Russell Eastwood Evans Lagudah Peter Langridge Michael Mackay Lynne*. Sydney University Press.

Anderson, R. G. 1961. The inheritance of leaf rust resistance in seven varieties of common wheat. *Can. J. Plant Sci.* 41: 342-359.

Appel, J. A., E. DeWolf, W. W. Bockus, and T. Todd. 2012. *Kansas Cooperative Plant Disease Survey Report: Preliminary 2012 Kansas Wheat Disease Loss Estimates*. https://agriculture.ks.gov/docs/default-source/pp-disease-reports-2012/2012-ks-wheat-disease-loss-estimates.pdf?sfvrsn=3 (accessed 19 Mar. 2013).

Appel, J. A., E. DeWolf, W. W. Bockus, and T. Todd. 2013. *Kansas Cooperative Plant Disease Survey Report: Preliminary 2013 Kansas Wheat Disease Loss Estimates*. https://agriculture.ks.gov/docs/default-source/pp-disease-reports-2012/kansas-cooperative-plant-disease-2013.pdf?sfvrsn=2 (accessed 7 Jan. 2014).

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one* 3: e3376.

Bartoš, P., D. J. Samborski, and P. L. Dyck. 1969. Leaf rust resistance of some European varieties of wheat. *Can. J. Bot.* 47: 543-546.

Bent, A. F., B. N. Kunkel, D. Dahlbeck, K. L. Brown, R. Schmidt, J. Giraudat, J. Leung, and B. J. Staskawicz. 1994. RPS2 of Arabidopsis thaliana: a leucine-rich repeat class of plant disease resistance genes. *Science* 265: 1856-1860.

Bockus, W. W., R. L. Bowden, R. M. Hunger, T. D. Murray, and R. W. Smiley. 2010. Compendium of wheat diseases and pests (No. Ed. 3). American Phytopathological Society (APS Press).

Bolton, M. D., J. A. Kolmer, and D. F. Garvin. 2008. Wheat leaf rust caused by Puccinia triticina. *Molecular Plant Pathology* 9: 563-575.

Bremenkamp-Barrett, B., J. D. Faris, and J. P. Fellers. 2008. Molecular Mapping of the Leaf Rust Resistance Gene *Lr17a* in Wheat. *Crop science* 48: 1124-1128.

Brenchley, R., M. Spannagl, M. Pfeifer, G. LA Barker, R. D'Amore, A. M. Allen, N McKenzie et al. 2012. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* 491: 705-710.

Bruce, M., K. A. Neugebauer, D. L. Joly, P. Migeon, C. A. Cuomo, S. Wang, E. Akhunov, G. Bakkeren, J. A. Kolmer, and J. P. Fellers. 2013. Using transcription of six Puccinia triticina races to identify the effective secretome during infection of wheat. *Frontiers in plant science* 4.

Brueggeman, R., N. Rostoks, D. Kudrna, A. Kilian, F. Han, J. Chen, A. Druka, B. Steffenson, and A. Kleinhofs. 2002. The barley stem rust-resistance gene Rpg1 is a novel disease-resistance gene with homology to receptor kinases. *PNAS* 99: 9328-9333.

Brydon, E. 2012. *Bread wheat's large and complex genome is revealed*. Retrieved from http://www.cshl.edu/Archive/bread-wheats-large-and-complex-genome-is-revealed (accessed 25 Nov. 2013).

Cadalen, T., C. Boeuf, S. Bernard, and M. Bernard. 1997. An intervarietal molecular marker map in Triticum aestivum L. Em. Thell. and comparison with a map from a wide cross. *Theor Appl Genet* 94: 367-377.

Calenge, F., C. G. Van der Linden, E. Van de Weg, H. J. Schouten, G. Van Arkel, C. Denance, and C-E. Durel. 2005. Resistance gene analogues identified through the NBS-profiling method map close to major genes and QTL for disease resistance in apple. *Theor Appl Genet* 110: 660-668.

Chen, X., T. Coram, X. Huang, M. Wang, and A. Dolezal. 2013. Understanding molecular mechanisms of durable and non-durable resistance to stripe rust in wheat using a transcriptomics approach. *Current genomics* 14: 111.

Chen, X. M., R. F. Line, and H. Leung. 1998. Genome scanning for resistance-gene analogs in rice, barley, and wheat by high-resolution electrophoresis. *Theor Appl Genet* 97: 345-355.

Choulet, F., T. Wicker, C. Rustenholz, E. Paux, J. Salse, P. Leroy, S. Schlub et al. 2010. Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. *The Plant Cell Online* 22: 1686-1701.

Clark, B., R. Bryson, L. Tonguc, C. Kelly, and G. Jellis. Rothamsted Research Ltd. 2009. *The Encylopaedia of Cereal Diseases: Brown (Leaf) Rust*. Retrieved from http://archive.hgca.com/document.aspx?fn=load&media_id=8506&publicationId=4488 (accessed 6 Feb. 2014).

Cloutier, S., B. D. McCallum, C. Loutre, T. W. Banks, T. Wicker, C. Feuillet, B. Keller, and M. C. Jordan. 2007. Leaf rust resistance gene Lr1, isolated from bread wheat (Triticum aestivum L.) is a member of the large psr567 gene family. *Plant molecular biology* 65: 93-106.

Collard, B. C. Y., and D. J. Mackill. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363: 557-572.

Collins, N. C., C. A. Webb, S. Seah, J. G. Ellis, S. H. Hulbert, and A. Pryor. 1998. The isolation and mapping of disease resistance gene analogs in maize. *MPMI* 11: 968-978.

Cortés, A. J., M. C. Chavarro, and M. W. Blair. 2011. SNP marker diversity in common bean (Phaseolus vulgaris L.). *Theor Appl Genet* 123: 827-845.

Costa, V., C. Angelini, I. De Feis, and A. Ciccodicola. 2010. Uncovering the Complexity of Transcriptomes with RNA-Seq. *J Biomed Biotechnol.*

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

Crossa, J., J. Burgueno, S. Dreisigacker, M. Vargas, S. A. Herrera-Foessel, M. Lillemo, R. P. Singh et al. 2007. Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. *Genetics* 177: 1889-1913.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12: 499-510.

Deslandes, L., J. Olivier, F. Theulières, J. Hirsch, D. Xin Feng, P. Bittner-Eddy, J. Beynon, and Y. Marco. 2002. Resistance to Ralstonia solanacearum in Arabidopsis thaliana is conferred by the recessive RRS1-R gene, a member of a novel family of resistance genes. *PNAS* 99: 2404-2409.

Dholakia, B. B., A. V. Rajwade, P. Hosmani, R. R. Khan, S. Chavan, D. M. R. Reddy, M. D. Lagu, U. K. Bansal, R. G. Saini, and V. S. Gupta. 2013. Molecular mapping of leaf rust resistance gene Lr15 in hexaploid wheat. *Mol Breeding*: 1-5.

Dilbirligi, Muharrem, and Kulvinder S. Gill. 2003. Identification and analysis of expressed resistance gene sequences in wheat. *Plant molecular biology* 53: 771-787.

Dinesh-Kumar, S. P., W. Tham, and B. J. Baker. 2000. Structure–function analysis of the tobacco mosaic virus resistance gene N. *PNAS* 97: 14789-14794.

Dixon, M. S., D. A. Jones, J. S. Keddie, C. M. Thomas, K. Harrison, and J. DG Jones. 1996. The Tomato *Cf-2* Disease Resistance Locus Comprises Two Functional Genes Encoding Leucine-Rich Repeat Proteins. *Cell* 84: 451-459.

Dodds, P. N., G. J. Lawrence, and J. G. Ellis. 2001. Contrasting modes of evolution acting on the complex N locus for rust resistance in flax. *The Plant Journal* 27: 439-453.

Dubcovsky, J., and Jan Dvorak. 2007. Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* 316: 1862-1866.

Dudoit, S., Y. H. Yang, M. J. Callow, and T. P. Speed. 2002. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica sinica* 12: 111-140.

Dyck, P. L., and D. J. Samborski. 1968. Genetics of resistance to leaf rust in the common wheat varieties Webster, Loros, Brevit, Carina, Malakof and Centenario. *Can. J. Genet. Cytol.* 10: 7-17.

Dyck, P. L., and E. R. Kerber. 1971. Chromosome location of three genes for leaf rust resistance in common wheat. *Can. J. Genet. Cytol.* 13: 480-483.

Dyck, P. L., and E. R. Kerber. 1985. Resistance of the race-specific type. *The cereal rusts* 2: 469-500.

Dyck, P. L., and R. Johnson. 1983. Temperature sensitivity of genes for resistance in wheat to Puccinia recondita. *Can. J. Plant Pathol.* 5: 229-234.

Eastwood, R. F., E. S. Lagudah, and R. Appels. 1994. A directed search for DNA sequences tightly linked to cereal cyst nematode resistance genes in Triticum tauschii. *Genome* 37: 311-319.

Ellis, J., & Jones, D. A. 2003. Plant disease resistance genes. p. 27-45. *In* J. A. Hoffman and R. A. Ezecowitz (eds.) *Infectious Disease: Innate Immunity*. Humana Press, Totowa, NJ. Humana Press.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one* 6: e19379.

Eversmeyer, M. G., and C. L. Kramer. 2000. Epidemiology of Wheat Leaf and Stem Rust in the Central Great Plains of the USA. *Annu. Rev. Phytopathol.* 38: 491-513.

FAOSTAT. 2014. *FAOSTAT crop production*. Retrieved from http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567 (accessed 6 Jan. 2014).

Feuillet, C., G. Schachermayr, and B. Keller. 1997. Molecular cloning of a new receptor‐like kinase gene encoded at the Lr10 disease resistance locus of wheat. *The Plant Journal* 11: 45-52.

Feuillet, C., S. Travella, N. Stein, L. Albar, A. Nublat, and B. Keller. 2003. Map-based isolation of the leaf rust disease resistance gene Lr10 from the hexaploid wheat (Triticum aestivum L.) genome. *PNAS* 100: 15253-15258.

Flor, H. H. 1971. Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.* 9: 275-296.

Fradin, E. F., Z. Zhang, J. C. Juarez Ayala, C. DM Castroverde, R. N. Nazar, J. Robb, C. Liu, and B. PHJ Thomma. 2009. Genetic dissection of Verticillium wilt resistance mediated by tomato Ve1. *Plant Physiology* 150: 320-332.

Fu, D., C. Uauy, A. Distelfeld, A. Blechl, L. Epstein, X. Chen, H. Sela, T. Fahima, and J. Dubcovsky. 2009. A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. *Science* 323: 1357-1360.

German, S. E., and J. A. Kolmer. 1992. Effect of gene Lr34 in the enhancement of resistance to leaf rust of wheat. *Theor Appl Genet* 84: 97-105.

Gibson, G., and S. Muse. 2004. *A Primer of Genome Science*. Sunderland, MA: Sinauer Associates, Inc. Publishers.

Glaubitz, J. C., T. M. Casstevens, F. Lu, J. Harriman, R. J. Elshire, Q. Sun, and E. S. Buckler. 2014. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PloS one* 9: e90346.

Göhre, V., and Silke Robatzek. 2008. Breaking the barriers: microbial effector molecules subvert plant immunity. *Annu. Rev. Phytopathol.* 46: 189-215.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644-652.

Grant, M. R., L. Godiard, E. Straube, T. Ashfield, J. Lewald, A. Sattler, R. W. Innes, and J. L. Dangl. 1995. Structure of the Arabidopsis RPM1 gene enabling dual specificity disease resistance. *Science* 269: 843-846.

Grube, R. C., E. R. Radwanski, and M. Jahn. 2000. Comparative genetics of disease resistance within the Solanaceae. *Genetics* 155: 873-887.

Gu, Y., Q. Zhao, D. Sun, and W. Song. 2008. A genetic linkage map based on AFLP and NBS markers in cauliflower (Brassica oleracea var. botrytis). *Botanical Studies* 49: 93-99.

Gupta, P., H. Balyan, K. Edwards, P. Isaac, V. Korzun, M. Röder, M.–F. Gautier et al. 2002. Genetic mapping of 66 new microsatellite (SSR) loci in bread wheat. *Theor Appl Genet* 105: 413-422.

Gururani, M. A., J. Venkatesh, C. P. Upadhyaya, A. Nookaraju, S. Kumar Pandey, and S. W. Park. 2012. Plant disease resistance genes: Current status and future directions. *Physiological and Molecular Plant Pathology* 78: 51-65.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* 8: 1494-1512.

Haas, B. J., and M. C. Zody. 2010. Advancing RNA-Seq analysis. *Nature Biotechnology* 28: 421-423.

Haudry, A., A. Cenci, C. Ravel, T. Bataillon, D. Brunel, C. Poncet, I. Hochu et al. 2007. Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Mol. Biol. Evol.* 24: 1506-1517.

Hiremath, P. J., A. Kumar, R. V. Penmetsa, A. Farmer, J. A. Schlueter, S. K. Chamarthi, A. M. Whaley et al. 2012. Large‐scale development of cost‐effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. *Plant biotechnology journal* 10: 716-732.

Huang, Li, S.A. Brooks, W. Li, J. P. Fellers, H. N. Trick, and B. S. Gill. 2003. Map-based cloning of leaf rust resistance gene Lr21 from the large and polyploid genome of bread wheat. *Genetics* 164: 655-664.

Herrera-Foessel, S. A., E. S. Lagudah, J. Huerta-Espino, M. J. Hayden, H. S. Bariana, D. Singh, and R. P. Singh. 2011. New slow-rusting leaf rust and stripe rust resistance genes Lr67 and Yr46 in wheat are pleiotropic or closely linked. *Theor Appl Genet* 122: 239-249.

Herrera-Foessel, S. A., R. P. Singh, J. Huerta-Espino, G. M. Rosewarne, S. K. Periyannan, L. Viccars, V. Calvo-Salazar, C. Lan, and E. S. Lagudah. 2012. Lr68: a new gene conferring slow rusting resistance to leaf rust in wheat. *Theor Appl Genet* 124: 1475-1486.

Huang, L., S. A. Brooks, W. Li, J. P. Fellers, H. N. Trick, and B. S. Gill. 2003. Map-based cloning of leaf rust resistance gene Lr21 from the large and polyploid genome of bread wheat. *Genetics* 164: 655-664.

Huang, S., A. Sirikhachornkit, X. Su, J. Faris, B. Gill, R. Haselkorn, and P. Gornicki. 2002. Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the Triticum/Aegilops complex and the evolutionary history of polyploid wheat. *PNAS* 99: 8133-8138.

Huerta-Espino, J., R. P. Singh, S. Germán, B. D. McCallum, R. F. Park, W. Q. Chen, S. C. Bhardwaj, and H. Goyeau. 2011. Global status of wheat leaf rust caused by Puccinia triticina. *Euphytica* 179: 143-160.

International Development Research Center (IDRC). 2013. *FACTS & FIGURES on Food and Biodiversity*. Retrieved from http://www.idrc.ca/EN/Resources/Publications/Pages/ArticleDetails.aspx?PublicationID=565 (accessed 6 Jan. 2014).

Isaac, P. 2004. *The wheat microsatellite consortium*. Retrieved from http://wheat.pw.usda.gov/ggpages/SSR/WMC/WMC_TheInsideStory.html (accessed 27 Jan. 2013).

Johal, G. S., and S. P. Briggs. 1992. Reductase activity encoded by the HM1 disease resistance gene in maize. *Science* 258: 985-987.

Jones, J. DG, and J. L. Dangl. 2006. The plant immune system. *Nature* 444: 323-329.

Kanazin, V., L. F. Marek, and R. C. Shoemaker. 1996. Resistance gene analogs are conserved and clustered in soybean. *PNAS* 93: 11746-11750.

Kansas Wheat. 2013. *Classes of Wheat*. Retrieved from http://www.kswheat.com/consumerspageid260_ClassesofWheat.shtml (accessed 23 Feb. 2014).

Karp, A., O. L. E. Seberg, and M. Buiatti. 1996. Molecular techniques in the assessment of botanical diversity. *Annals of Botany* 78: 143-149.

Keller, B., E. S. Lagudah, L. L. Selter, J. M. Risk, C. Harsh, and S. G. Krattinger. 2012. How has Lr34/Yr18 conferred effective rust resistance in wheat for so long?. In *Proceedings of Borlaug Global Rust Initiative Technical Workshop, Beijing, China*.

Kochieva, E. Z., and N. N. Ryzhova. 2009. Analysis of resistance gene family diversity in pepper (Capsicum annuum). In *Doklady Biochemistry and Biophysics* 425:73-75. MAIK Nauka/Interperiodica.

Kolmer, J. 2013. Leaf Rust of Wheat: Pathogen Biology, Variation and Host Resistance. *Forests* 4: 70-84.

Kolmer, J. A. 1992. Enhanced leaf rust resistance in wheat conditioned by resistance gene pairs with Lr13. *Euphytica* 61: 123-130.

Kolmer, J. A. 1996. Genetics of resistance to wheat leaf rust. *Ann. Rev. Phytopathol.* 34: 435-455.

Kolmer, J. A., and J. Q. Liu. 2002. Inheritance of leaf rust resistance in the wheat cultivars AC Majestic, AC Splendor, and AC Karma. *Can. J. Plant Pathol.* 24: 327-331.

Kolmer, J. A., D. L. Long, and M. E. Hughes. 2009. Physiologic specialization of Puccinia triticina on wheat in the United States in 2007. *Plant Disease* 93: 538-544.

Kolmer, J. A., D. L. Long, and M. E. Hughes. 2011. Physiologic Specialization of Puccinia triticina on Wheat in the United States in 2009. *Plant Disease* 95: 935-940.

Kolmer, J. A., and L. M. Oelke. 2006. Genetics of leaf rust resistance in the spring wheats 'Ivan'and 'Knudson'spring wheat. *Can. J. Plant Pathol.* 28: 223-229.

Krattinger, S. G., E. S. Lagudah, W. Spielmeyer, R. P. Singh, J. Huerta-Espino, H. McFadden, E. Bossolini, L. L. Selter, and B. Keller. 2009. A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. *Science* 323: 1360-1363.

Kuleung, C., P. S. Baenziger, and I. Dweikat. 2004. Transferability of SSR markers among wheat, rye, and triticale. *Theor Appl Genet* 108: 1147-1150.

Lagudah, E. S., O Moullet, and R. Appels. 1997. Map-based cloning of a gene sequence encoding a nucleotide-binding domain and a leucine-rich region at the Cre3 nematode resistance locus of wheat. *Genome* 40: 659-665.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly, S. E. Lincoln, and L. Newburg. 1987. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1: 174-181.

Laroche, A., M. M. Frick, R. Huel, C. Nykiforuk, B. Conner, and A. Kuzyk. 2000. Molecular identification of the wheat stripe rust resistance gene Yr10, the first full-length leucine zipper-nucleotide binding site-leucine-rich-repeat resistance gene in cereals. http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=11990496

Lawrence, G. J., E. J. Finnegan, M. A. Ayliffe, and J. G. Ellis. 1995. The L6 gene for flax rust resistance is related to the Arabidopsis bacterial resistance gene RPS2 and the tobacco viral resistance gene N. *The Plant Cell Online* 7: 1195-1206.

Leister, D., A. Ballvora, F. Salamini, and C. Gebhardt. 1996. A PCR–based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. *Nature genetics* 14: 421-429.

Li, H., X. Gao, X. Li, Q. Chen, J. Dong, and W. Zhao. 2013. Evaluation of Assembly Strategies Using RNA-Seq Data Associated with Grain Development of Wheat (Triticum aestivum L.). *PloS one* 8: e83530.

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

Li, W., P. Zhang, J. P. Fellers, B. Friebe, and B. S. Gill. 2004. Sequence composition, organization, and evolution of the core Triticeae genome. *The Plant Journal* 40: 500-511.

Liu, H., M. Bayer, A. Druka, J. R. Russell, C. A. Hackett, J. Poland, L. Ramsay, P. E. Hedley, and R. Waugh. 2014. An evaluation of genotyping by sequencing (GBS) to map the Breviaristatum-e (ari-e) locus in cultivated barley. *BMC genomics* 15: 104.

Lokossou, A. A., T. Park, G. van Arkel, M. Arens, C. Ruyter-Spira, J. Morales, S. C. Whisson et al. 2009. Exploiting knowledge of R/Avr genes to rapidly clone a new LZ-NBS-LRR family of late blight resistance genes from potato linkage group IV. *MPMI* 22: 630-641.

Long, D. L., and J. A. Kolmer. 1989. A North American system of nomenclature for Puccinia recondita f. sp. tritici. *Phytopathology* 79: 525-529.

López-Maury, L., S. Marguerat, and J. Bähler. 2008. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nature Reviews Genetics* 9: 583-593.

Lu, F., A. E. Lipka, J. Glaubitz, R. Elshire, J. H. Cherney, M. D. Casler, E. S. Buckler, and D. E. Costich. 2013. Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. *PLoS genetics* 9: e1003215.

Lu, T., G. Lu, D. Fan, C. Zhu, W. Li, Q. Zhao, Q. Feng et al. 2010. Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. *Genome research* 20: 1238-1249.

Lynch, M., and B. Milligan. 1994. Analysis of population genetic structure with RAPD markers. *Molecular Ecology* 3: 91-99.

Mains, E. B., C. E. Leighty, and C. O. Johnston. 1926. Inheritance of resistance to leaf rust, Puccinia triticina Erikss, in crosses of common wheat, Triticum vulgare vill. Authority of the Secretary of Agriculture.

Maleki, L., J. D. Faris, R. L. Bowden, B. S. Gill, and J. P. Fellers. 2003. Physical and genetic mapping of wheat kinase analogs and NBS-LRR resistance gene analogs. *Crop science* 43: 660-670.

Mantovani, P., G. van der Linden, M. Maccaferri, M. C. Sanguineti, and R. Tuberosa. 2006. Nucleotide-binding site (NBS) profiling of genetic diversity in durum wheat. *Genome* 49: 1473-1480.

Marguerat, S., and J. Bähler. 2010. RNA-seq: from technology to biology. *Cell Mol. Life Sci.* 67: 569-579.

Marioni, J. C., C. E. Mason, S. M. Mane, M. Stephens, and Y. Gilad. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome research* 18: 1509-1517.

Marshall, D. 1992. LEAF RUST GENE DEPLOYMENT AND DURABILITY. *in the Proceedings of the Nineteenth Hard Red Winter Wheat Workers Conference* 19.

Martin, G. B., S. H. Brommonschenkel, J. Chunwongse, A. Frary, M. W. Ganal, R. Spivey, T. Wu, E. D. Earle, and S. D. Tanksley. 1993. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* 262: 1432-1436.

Martin, G. B., M. W. Ganal, and S. D. Tanksley. 1992. Construction of a yeast artificial chromosome library of tomato and identification of cloned segments linked to two disease resistance loci. *Mol Gen Genet* 233: 25-32.

Martinez, F., R. E. Niks, R. P. Singh, and D. Rubiales. 2001. Characterization of Lr46, a gene conferring partial resistance to wheat leaf rust. *Hereditas* 135: 111-114.

Mascher, M., S. Wu, P. St Amand, N. Stein, and J. Poland. 2013. Application of genotyping-by-sequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in barley. *PloS one* 8: e76925.

McCartney, C. A., D. J. Somers, B. D. McCallum, J. Thomas, D. G. Humphreys, J. G. Menzies, and P. D. Brown. 2005. Microsatellite tagging of the leaf rust resistance gene Lr16 on wheat chromosome 2BSc. *Molecular breeding* 15: 329-337.

McFadden, H. G., A. Lehmensiek, and E. S. Lagudah. 2006. Resistance gene analogues of wheat: molecular genetic analysis of ESTs. *Theor Appl Genet* 113: 987-1002.

McIntosh, R. A., and N. H. Luig. 1973. Linkage of genes for reaction to Puccinia graminis f. sp. tritici and P. recondita in Selkirk wheat and related cultivars. *Aust. J. biol. Sci.* 26: 1145-1152.

McIntosh, R. A., C. R. Welling, and R. F Park. 1995. *Wheat rusts: an atlas of resistance genes*. CSIRO PUBLISHING.

McIntosh, R.A., Y. Yamazaki, J. Dubcovsky, W. J. Rogers, C. Morris, R. Appels, K. M. Devos. 2014. Catalogue of Gene Symbols for Wheat. KOMUGI Integrated Wheat Science Database. Available at http://www.shigen.nig.ac.jp/wheat/komugi/genes/symbolClassList.jsp (accessed 9 Jan. 2014).

Metzker, M. L. 2005. Emerging technologies in DNA sequencing. *Genome research* 15: 1767-1776.

Miller, M. R., J. P. Dunham, A. Amores, W. A. Cresko, and E. A. Johnson. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* 17: 240-248.

Mindrinos, M., F. Katagiri, G. L. Yu, and F. M. Ausubel. 1994. The A. thaliana disease resistance gene RPS2 encodes a protein containing a nucleotide-binding site and leucine-rich repeats. *Cell* 78: 1089.

Minnesota Association of Wheat Growers. 1994. *Six basic classes of wheat*. Retrieved from http://www.smallgrains.org/WHFACTS/6classwh.htm (accessed 6 Jan. 2014)

Morgante, M., M. Hanafey, and W. Powell. 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature genetics* 30: 194-200.

Nagalakshmi, U., Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein, and M. Snyder. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320: 1344-1349.

Neelam, K., G. Brown-Guedira, and L. Huang. 2013. Development and validation of a breeder-friendly KASPar marker for wheat leaf rust resistance locus Lr21. *Mol Breeding* 31: 233-237.

Nesbitt, M., and Delwen Samuel. 1998. Wheat domestication: archaeobotanical evidence. *Science* 279: 1431-1431.

Neu, C., B. Keller, and C. Feuillet. 2003. Cytological and molecular analysis of the Hordeum vulgare-Puccinia triticina nonhost interaction. *MPMI* 16: 626-633.

Oelke, L. M., and J. A. Kolmer. 2004. Characterization of leaf rust resistance in hard red spring wheat cultivars. *Plant disease* 88: 1127-1133.

Oono, Y., F. Kobayashi, Y. Kawahara, T. Yazawa, H. Handa, T. Itoh, and T. Matsumoto. 2013. Characterisation of the wheat (triticum aestivum L.) transcriptome by de novo assembly for the discovery of phosphate starvation-responsive genes: gene expression in Pi-stressed wheat. *BMC genomics* 14: 77.

Ozsolak, F., and P. M. Milos. 2011. RNA sequencing: advances, challenges and opportunities. *Nature reviews genetics* 12: 87-98.

Ozsolak, F., A. R. Platt, D. R. Jones, J. G. Reifenberger, L. E. Sass, P. McInerney, J. F. Thompson, J. Bowers, M. Jarosz, and P. M. Milos. 2009. Direct RNA sequencing. *Nature* 461: 814-818.

Padmanabhan, M., P. Cournoyer, and S. P. Dinesh‐Kumar. 2009. The leucine‐rich repeat domain in plant innate immunity: a wealth of possibilities. *Cellular microbiology* 11: 191-198.

Paux, E., D. Roger, E. Badaeva, G. Gay, M. Bernard, P. Sourdille, and C. Feuillet. 2006. Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC‐end sequencing on chromosome 3B. *The Plant Journal* 48: 463-474.

Periyannan, S., J. Moore, M. Ayliffe, U. Bansal, X. Wang, L. Huang, K. Deal et al. 2013. The gene Sr33, an ortholog of barley Mla genes, encodes resistance to wheat stem rust race Ug99. *Science* 341: 786-788.

Poland, J. A., P. J. Brown, M. E. Sorrells, and J. Jannink. 2012a. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PloS one* 7: e32253.

Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, S. Dreisigacker et al. 2012b. Genomic selection in wheat breeding using genotyping-by-sequencing. *The Plant Genome* 5: 103-113.

Poland, J. A., and T. W. Rife. 2012. Genotyping-by-sequencing for plant breeding and genetics. *The Plant Genome* 5: 92-102.

Quirin, E. A., H. Mann, R. S. Meyer, A. Traini, M. L. Chiusano, A. Litt, and J. M. Bradeen. 2012. Evolutionary meta-analysis of solanaceous resistance gene and Solanum resistance gene analog sequences and a practical framework for cross-species comparisons. *MPMI* 25: 603-612.

Richter, T. E., and P. C. Ronald. 2000. The evolution of disease resistance genes. *Plant molecular biology* 42: 195-204.

Roelfs, A. P. 1985. Epidemiology in North America. *The cereal rusts*, *2*, 403-434.

Roelfs, A. P. 1989. Epidemiology of the cereal rusts in North America. *Can. J. Plant Pathol.* 11: 86-90.

Röder, M. S., V. Korzun, K. Wendehake, J. Plaschke, M. Tixier, P. Leroy, and M. W. Ganal. 1998. A microsatellite map of wheat. *Genetics* 149: 2007-2023.

Röder, M. S., J. Plaschke, S. U. König, A. Börner, M. E. Sorrells, S. D. Tanksley, and M. W. Ganal. 1995. Abundance, variability and chromosomal location of microsatellites in wheat. *Mol Gen Genet* 246: 327-333.

Rouse, M. N., E. L. Olson, B. S. Gill, M. O. Pumphrey, and Y. Jin. 2011. Stem Rust Resistance in Germplasm. *Crop Science* 51: 2074-2078.

Rubiales, D., and R. E. Niks. 1995. Characterization of Lr34, a major gene conferring nonhypersensitive resistance to wheat leaf rust. *Plant disease* 79: 1208-1212.

Saintenac, C., W. Zhang, A. Salcedo, M. N. Rouse, H. N. Trick, E. Akhunov, and J. Dubcovsky. 2013. Identification of wheat gene Sr35 that confers resistance to Ug99 stem rust race group. *Science* 341: 783-786.

Salentijn, E., S. Goryunova, N. Bas, I. van der Meer, H. van den Broeck, T. Bastien, L. Gilissen, and M. Smulders. 2009. Tetraploid and hexaploid wheat varieties reveal large differences in expression of alpha-gliadins from homoeologous Gli-2 loci. *BMC genomics* 10: 48.

Samborski, D. J. 1985. Wheat leaf rust. *The cereal rusts* 2: 39-59.

Samborski, D. J., and P. L. Dyck. 1976. Inheritance of virulence in Puccinia recondita on six backcross lines of wheat with single genes for resistance to leaf rust. *Can. J. Bot.* 54: 1666-1671.

Samborski, D. J., and P. L. Dyck. 1982. Enhancement of resistance to Puccinia recondita by interactions of resistance genes in wheat. *Can. J. Plant Pathol.* 4: 152-156.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature biotechnology* 18.

Seah, S., K. Sivasithamparam, A. Karakousis, and E. S. Lagudah. 1998. Cloning and characterisation of a family of disease resistance gene analogs from wheat and barley. *Theor Appl Genet* 97: 937-945.

Semagn, K., R. Babu, S. Hearne, and M. Olsen. 2014. Single nucleotide polymorphism genotyping using Kompetitive allele specific PCR (KASP): Overview of the technology and its application in crop improvement. *Mol Breeding* 33: 1-14.

Shang, J., Y. Tao, X. Chen, Y. Zou, C. Lei, Ji. Wang, X. Li et al. 2009. Identification of a new rice blast resistance gene, Pid3, by genomewide comparison of paired nucleotide-binding site–leucine-rich repeat genes and their pseudogene alleles between the two sequenced rice genomes. *Genetics* 182: 1303-1311.

Shen, G., K. G. Abdullah, and Q. K. Wang. 2009. The TaqMan method for SNP genotyping. In *Single Nucleotide Polymorphisms*, pp. 293-306. Humana Press.

Shewry, P. R. 2009. Wheat. *Journal of Experimental Botany* 60: 1537-1553.

Shi, Z. X., X. M. Chen, R. F. Line, H. Leung, and C. R. Wellings. 2001. Development of resistance gene analog polymorphism markers for the Yr9 gene resistance to wheat stripe rust. *Genome* 44: 509-516.

Simmons, S. R., E. A. Oelke, and P. M. Anderson. 2013. *Growth and development guide for spring wheat*. Retrieved from http://www.extension.umn.edu/agriculture/small-grains/growth-and-development/spring-wheat/ (accessed 6 Jan. 2014).

Singh, R. P., and E. E. Saari. 1992. *Rust diseases of wheat: concepts and methods of disease management*. Edited by Alan P. Roelfs. CIMMYT (free PDF download).

Singh, R. P., D. P. Hodson, J. Huerta-Espino, Y. Jin, S. Bhavani, P. Njau, S. Herrera-Foessel, P. K. Singh, S. Singh, and V. Govindan. 2011. The emergence of Ug99 races of the stem rust fungus is a threat to world wheat production. *Annu. Rev. Phytopathol.* 49: 465-481.

Singh, R. P., J. Huerta-Espino, and H. M. William. 2005. Genetics and breeding for durable resistance to leaf and stripe rusts in wheat. *Turk J Agric For* 29: 121-127.

Smith, D. B. and R. B. Flavell. 1975. Characterisation of the wheat genome by association genetics. *Chromosoma* 50: 223–242

Somers, D. J., P. Isaac, and K. Edwards. 2004. A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). *Theor Appl Genet* 109:1105–1114.

Sonah, H., M. Bastien, E. Iquira, A. Tardivel, G. Légaré, B. Boyle, É. Normandeau et al. 2013. An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PloS one* 8: e54603.

Song, Q. J., E. W. Fickus, and P. B. Cregan. 2002. Characterization of trinucleotide SSR motifs in wheat. *Theor Appl Genet* 104: 286-293.

Song, Q. J., J. R. Shi, S. Singh, E. W. Fickus, J. M. Costa, J. Lewis, B. S. Gill, R. Ward, and P. B. Cregan. 2005. Development and mapping of microsatellite (SSR) markers in wheat. *Theor Appl Genet* 110: 550-560.

Song, W., G. Wang, L. Chen, H. Kim, L. Pi, T. Holsten, J. Gardner et al. 1995. A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* 270: 1804-1806.

Spielmeyer, W., M. Robertson, N. Collins, D. Leister, P. Schulze-Lefert, S. Seah, O. Moullet, and E. S. Lagudah. 1998. A superfamily of disease resistance gene analogs is located on all homoeologous chromosome groups of wheat (Triticum aestivum). *Genome* 41: 782-788.

Steffenson, B. J., P. Olivera, J. K. Roy, Y. Jin, K. P. Smith, and G. J. Muehlbauer. 2007. A walk on the wild side: mining wild wheat and barley collections for rust resistance genes. *Crop and Pasture Science* 58: 532-544.

Steijger, T., J. F. Abril, P. G. Engström, F. Kokocinski, T. J. Hubbard, R. Guigó, J. Harrow, P. Bertone, and RGASP Consortium. 2013. Assessment of transcript reconstruction methods for RNA-seq. *Nature methods*.

Syed, N. H., A. P. Sørensen, R. Antonise, C. van de Wiel, C. G. van der Linden, W. van't Westende, D. AP Hooftman, H. CM den Nijs, and A. J. Flavell. 2006. A detailed linkage map of lettuce based on SSAP, AFLP and NBS markers. *Theor Appl Genet* 112: 517-527.

Tanksley, S. D., N. D. Young, A. H. Paterson, and M. W. Bonierbale. 1989. RFLP mapping in plant breeding: new tools for an old science. *Nature Biotechnology* 7: 257-264.

Thomas, J., N. Fineberg, G. Penner, C. McCartney, T. Aung, I. Wise, and B. McCallum. 2005. Chromosome location and markers of Sm1: a gene of wheat that conditions antibiotic resistance to orange wheat blossom midge. *Molecular Breeding* 15: 183-192.

Thomas, C. M., D. A. Jones, M. Parniske, K. Harrison, P. J. Balint-Kurti, K. Hatzixanthis, and J. D. Jones. 1997. Characterization of the tomato Cf-4 gene for resistance to Cladosporium fulvum identifies sequences that determine recognitional specificity in Cf-4 and Cf-9. *The Plant Cell Online* 9: 2209-2224.

Tsuda, K., and F. Katagiri. 2010. Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Current opinion in plant biology* 13: 459-465.

Uitdewilligen, J. G., A. A. Wolters, B. Bjorn, T. JA Borm, R. GF Visser, and H. J. van Eck. 2013. A Next-Generation Sequencing Method for Genotyping-by-Sequencing of Highly Heterozygous Autotetraploid Potato. *PloS one* 8: e62355.

Ulvé, V., E. Sevin, A. Chéron, and F. Barloy-Hubler. 2007. Identification of chromosomal alpha-proteobacterial small RNAs by comparative genome analysis and detection in Sinorhizobium meliloti strain 1021. *BMC genomics* 8: 467.

USDA-ARS. 2008. Codes used in wheat leaf rust survey databases. Retrieved from http://www.ars.usda.gov/Main/docs.htm?docid=11025 (accessed 10 Jan. 2014).

USDA NASS. 2013. *National statistics for wheat*. Retrieved from http://www.nass.usda.gov/Statistics_by_Subject/index.php?sector=CROPS (accessed 6 Jan. 2014).

van der Linden, C. G., D. CAE Wouters, V. Mihalka, E. Z. Kochieva, M. JM Smulders, and B. Vosman. 2004. Efficient targeting of plant disease resistance loci using NBS profiling. *Theor Appl Genet* 109: 384-393.

Vanzetti, L. S., P. Campos, M. Demichelis, L. A. Lombardo, P. Romina Aurelia, L. M. Vaschetto, C. T. Bainotti, and M. Helguera. 2011. Identification of leaf rust resistance genes in selected Argentinean bread wheat cultivars by gene postulation and molecular markers. *Electronic Journal of Biotechnology* 14: 9-9.

Vinatzer, B. A., A. Patocchi, L. Gianfranceschi, S. Tartarini, H. Zhang, C. Gessler, and S. Sansavini. 2001. Apple contains receptor-like genes homologous to the Cladosporium fulvum resistance gene family of tomato with a cluster of genes cosegregating with Vf apple scab resistance. *MPMI* 14: 508-515.

Vocke, G., and O. Liefert. USDA ERS. 2013. *USDA ERS - Wheat: Background*. Retrieved from http://www.ers.usda.gov/topics/crops/wheat/background.aspx (accessed 4 Feb. 2014).

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van De Lee, M. Hornes, A. Friters et al. 1995. AFLP: a new technique for DNA fingerprinting. *Nucl. Acids Res.* 23: 4407-4414.

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10: 57-63.

Whitham, S., S. P. Dinesh-Kumar, D. Choi, R. Hehl, C. Corr, and B. Baker. 1994. The product of the tobacco mosaic virus resistance gene *N*: Similarity to toll and the interleukin-1 receptor. *Cell* 78: 1101-1115.

Wicks, S. R., R. T. Yeh, W. R. Gish, R. H. Waterston, and R. HA Plasterk. 2001. Rapid gene mapping in Caenorhabditis elegans using a high density polymorphism map. *Nature genetics* 28: 160-164.

Wilkinson, P.A., M. O. Winfield, G. L. A. Barker, A. M. Allen, A. Burridge, J. A. Coghill, A. Burridge, and K. J. Edwards. 2012. CerealsDB 2.0: an integrated resource for plant breeders and scientists. *BMC Bioinformatics* 13: 219.

Xiao, S., S. Ellwood, O. Calis, E. Patrick, T. Li, M. Coleman, and J. G. Turner. 2001. Broad-spectrum mildew resistance in Arabidopsis thaliana mediated by RPW8. *Science* 291: 118-120.

Yahiaoui, N., P. Srichumpa, R. Dudler, and B. Keller. 2004. Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene Pm3b from hexaploid wheat. *The Plant Journal* 37: 528-538.

Yan, G., X. Chen, R. Line, and C. Wellings. 2003. Resistance gene-analog polymorphism markers co-segregating with the Yr5 gene for resistance to wheat stripe rust. *Theor Appl Genet* 106: 636-643.

You, F. M., N. Huo, K. R. Deal, Y. Q. Gu, M. Luo, P. E. McGuire, J. Dvorak, and O. D. Anderson. 2011. Annotation-based genome-wide SNP discovery in the large and complex Aegilops tauschii genome using next-generation sequencing without a reference genome sequence. *BMC genomics* 12: 59.

Zhang, L., R. Chen, J. Zhang, B. Ouyang, J. Xiao, H. Li, and Z. Ye. 2008. Cloning and analysis of resistance gene analogs from pepper (Capsicum annuum L.). *Scientia Agric Sinica* 41: 169-175.

Zipfel, C. 2009. Early molecular events in PAMP-triggered immunity. *Current opinion in plant biology* 12: 414-420.

Appendix A - Alignments to putative resistance gene domains of assembled RNA-seq contigs

Contig Number	Contig Length	Alignment
322	922	putative brown planthopper-induced resistance protein
4832	628	nematode resistance and Hs1-like protein
32248	279	NBS-LRR
6535	802	NBS-type putative resistance protein
11281	1105	nematode resistance and Hs1-like protein
11329	484	receptor-like protein kinase
12371	275	NBS-LRR disease resistance protein
15755	546	RPP8-like protein and NBS-LRR disease resistance protein
16767	325	Leucine-rich repeat protein
18035	244	Leucine-rich repeat protein
18397	256	pleiotropic drug resistance protein 4-like and PDR-type ABC transporter
22128	217	serine/threonine kinase-like protein
25594	268	serine/threonine kinase-like protein
27903	226	NBS-LRR disease resistance protein
29415	331	receptor-like kinase
30219	221	receptor-like kinase
31577	301	rust resistance Rp1-like protein
32334	243	NBS-LRR disease resistance and RGA1-like protein
32161	324	NBS-LRR disease resistance and RGA4-like protein
32451	236	NBS-LRR disease resistance and RGA4-like protein
34740	211	Protein kinase domain containing protein
36389	288	putative rp3 and NBS-LRR protein
37326	220	LRR receptor-like serine/threonine protein kinase
37714	427	receptor-like kinase
38700	426	putative disease resistance protein RGA3-like
40111	293	Leucine-rich repeat protein
40959	249	TIR-NBS-LRR
42928	338	NBS-LRR
47057	297	disease resistance protein-like [Oryza sativa Japonica Group]
49068	243	RPM1-like protein
50394	233	NBS-LRR
54869	263	disease resistance protein-like [Oryza sativa Japonica Group]
55354	211	NBS-LRR
56215	295	RPM1-like and NB-ARC domain containing protein
57859	307	NBS-LRR and Tsn1-like
58583	222	NBS-LRR and Yr10-like protein
58799	207	RPM1-like and RPP13-like protein
58837	327	NBS-LRR
60342	241	NBS-LRR
61359	202	RPM1-like protein

63249	261	NBS-LRR
64081	222	RPM1-like protein
72532	243	NBS-LRR and Tsn1-like
75817	261	NBS-LRR
75791	237	RPM1-like protein
76450	226	putative blight resistance protein
78268	280	NBS-LRR
78997	278	NBS-LRR
80253	438	NBS-LRR
82178	263	NBS-LRR
82823	207	NBS-LRR
85423	200	NBS-LRR
87097	296	Leucine-rich repeat protein
88663	345	NBS-LRR
89015	209	NBS-LRR
89748	237	NBS-LRR and Tsn1-like
89755	200	NBS-LRR
89946	397	RPM1-like NBS-LRR
221748	281	NB-ARC domain containing protein
216052	218	NB-ARC domain containing protein
40014	208	TIR-NBS-LRR
293773	209	NBS-LRR
282725	277	NBS-LRR
267101	277	Pm3 variant
266585	203	probable disease resistance RPP8-like protein 4-like
232092	227	NBS-LRR
228901	272	NBS-LRR
209129	221	NB-ARC domain containing protein
207290	217	NBS-LRR
206858	206	NBS-LRR
206341	253	NBS-LRR
204647	250	NBS-LRR
195055	229	serine-threonine protein kinase
189911	250	NBS-LRR
188349	210	NBS-LRR
187263	220	NBS-LRR
182447	269	serine-threonine protein kinase
172952	256	NBS-LRR
172362	208	NBS-LRR
168999	228	NBS-LRR
166225	201	Leaf rust resistance Lr10
165468	213	NBS-LRR
164720	238	NBS-LRR
162324	212	NBS-LRR
160284	223	Pm3 variant
156447	293	NBS-LRR
156115	224	NBS-LRR

155856	226	NBS-LRR
155001	221	NBS-LRR
153615	244	NBS-LRR
150007	230	NBS-LRR
148307	211	NBS-LRR
148270	201	putative disease resistance RPP13-like protein 3-like, partial
147259	248	NBS-LRR
146074	278	putative disease resistance protein RGA4-like
144794	235	NBS-LRR
144413	265	LRR receptor-like serine/threonine protein kinase
142887	214	receptor-like kinase
134492	201	RPM1-like protein
131838	237	NBS-LRR
131602	311	NBS-LRR
131599	200	NB-ARC domain containing protein
129013	210	NBS-LRR
128876	215	NBS-LRR
128533	442	NB-ARC domain containing protein
125606	242	NBS-LRR
123727	245	Protein kinase domain containing protein
123424	252	NBS-LRR
121908	204	NBS-LRR
121571	203	NBS-LRR
120401	232	NBS-LRR
119638	212	receptor-like protein kinase
118717	218	Mlo-like protein
118591	357	cyst nematode resistance protein-like protein
118416	251	NBS-LRR and Yr10-like protein
116728	226	NBS-LRR
109407	311	disease resistance protein RPM1, putative
103237	213	NBS-LRR
103013	214	NBS-LRR
101038	267	RPM1-like NBS-LRR
100479	230	NBS-LRR
99584	216	NBS-LRR
99577	207	didn't have any hits listed
99333	232	NBS-LRR
94540	229	NBS-LRR
92973	211	NBS-LRR
91787	201	serine-threonine protein kinase
91633	292	NBS-LRR
91604	214	NBS-LRR
91181	241	RPP13-like protein
90663	302	NB-ARC domain containing protein
90175	315	NBS-LRR
90046	238	LRR receptor-like serine/threonine protein kinase
326194	207	NBS-LRR

a205149	525	NBS-LRR
a211607	794	RGA4-like
a213409	169	NBS-LRR/NB-ARC conserved domain

Appendix B - Primers designed from RGA contigs

RGA/Contig #	Forward	Reverse	Product Size (bp)
322	AAAACAGTGACCGATGCCGC	TCTGCCTGCTACTCAAAG	216
4832	TGTCGCAGATGTTCCTGGAG	CGCCACTGAAATGAATCAC	235
32248	TCTCCAGCCTGTTCAATC	GATGCTCCTGACTAAGTG	207
6535	GTAAGAACAGCACCCAAG	CAGAAGGAAATGATAAGCGG	310
11281	AGAAATGGCGACGACCAAGC	TCTGGTGGATGGTGAAGAG	737
11281 (redesigned)	CAAGCAACGAAGTTTCTG	GTCGCCATTTCTTTCAGC	99
11329	ACTCGGAGAACATCATCC	ACTTCTCCGTCAGCATCTG	225
11329 (redesigned)	GTCAAGTCCACCAACATC	CTTCTCCGTCAGCATCTG	198
12371	CTCTTCTGTGCTATTGCTTG	GGATGACTGGGAAAATCTACTG	99
15755	GAGGATGTTGAAGGAGAAG	GACAGTCGCTGATTGATAG	307
16767	CTGAACAACAGCCTAAGAG	GCAGTTTGACAAGGAGATG	153
18035	CTCATCACCATCTCCTTC	CGATGGATTTCGAGCTCC	103
18397	TTGGATAGCCTGATACGG	GTGCTTTCATTGCCTCTG	206
18397 (redesigned)	GGTCTCTTGCTTCTTTGGATAG	CTTGTGCTTTCATTGCCTC	223
22128	TCTGTGTGAGGAATGTCC	GGCTAAGAACTACATCCG	112
25594	GTCCTCACTTGTTCAGATG	TAATGGCAGGAAAGGAGG	133
27903	ACTCCTCAGGTCACTTAG	CTTATCCTCAACTTCGCC	153
29415	CTGAAATGGCTTGTGATTGC	GAGGTTATTGTTGGCACC	165
30219	ATGCGGCTACAGATGATG	GCTGAGATGACAGTGTATTG	163
31577	GTACAGAACTAGCACTACATG	CAGGACAGTTCCAGTCATATTAG	278
32334	GGTGGTATGTTCAGTTGTG	ACTATCGTCTAACGCCAG	158
32161	GCAACCTTCAATGCCAAATC	CAATGTTTCGCCTATTCTGC	242
32451	GCATAACTTCTCACGACTTG	GGTGGTAAATGTATCTGTCTGTCC	75
34740	GGCTTCATCGCATTCAACATAG	CGTCACGACAACATAACG	102
36389	CGTGATAGCAGAACATTTCC	CAACTGACGATAAACCGATG	177
37326	ATGTGAGCGAAAGAGGAG	GAAGGTGACACCACACCAA	193
37714	GTGTTGACAGAGTATGGTAG	AGCCAGGACGATGTTTAC	321
38700	AGTGGAACCTGTGGTAGTGG	ACAAACGAAGCAAGCAGCG	371
40111	GTCAAGAAGACAAGGAAGAG	CCAGCTCCTCGAACAGGAAC	93
40959	CGTCTCTTGATGTCTACC	GGTGTTGATTCCTATGTGTC	169
42928	AGGTGGTCAACAATCTGG	GGACAAGTTCATAGTTGCC	103
47057	ACTGCTCTGTGCCAATACC	ATGGTGGTGGCAAGAGATG	141
49068	ATGACGAGGATGTGAAGC	CGTTGTTGGAGAATGTCC	168
50394	ATCCAGCATCTTGTCAGC	CTCTCTGTTGTTGGTGAAG	198
54869	CCACCACCCTCAAATAAG	GCCAGTTTCCAGAATGTG	169
55354	GGATGATTTGTGCTGAATGG	CGACCTCACTGTTATTGG	146
56215	GAGTTCAAGCATTTGCGTG	TGTGGCTGTTAGGTCCAAC	183
57859	TTGGCAACTCGGTGAAAG	GCATTGTGAACAGTGTGAC	116
57859 (redesigned)	TGGATTGCTCTTGGATTC	CATTGTGAACAGTGTGAC	159
58583	CTTAGTTGTGGAGGTAGG	GTTCTGGGCTACAAGATAG	85
58799	GATGAAGCCATTGACCTG	CAGCAGATACCACAGTTAG	75

58837	TGAGTTGGTTTGGAGGTC	CGCAATGCTCTTAGTGATG	318
60342	TGTTGAGACCAAGGTTCC	TTAGAGGATGTCAGACGC	119
61359	GACTGCTACCAGCACCAGATG	TCGACGTCGTAGCTCATCC[T]	179
63249	GCTAACAAAGTGCGAAGG	GAACTGACTGAACAGAAAGC	119
64081	GCCTTATCTTCTCGTAAAGC	CGTATGGATGATGTTTGGACTC	167
72532	GAGGTATCTTAGTGCTCC	CTCACAATGACGAAGGTTG	97
75817	TGGCAGTCCTCCACAACTTC	CAGCAGTTGAATCGCTTG	80
75791	TCCTTGCCAATCCGATGGTG	ATGAATGAGGTGCGAGAGC	164
76450	CAACCAGGATGCATCAAGG	CATGCATCAAGGCATATGCTAGG	166
78268	CGCTCACTTATCTCAGTC	CTCCAGTTGCTTGACATC	122
78997	CCTGATGTTCTTGGAAGG	TGACAGCAGCCTGAATAG	215
80253	GTCGTAGCATCATCATCAC	TTGTGGAACTTCGTCTGG	268
82178	TTCCGTGACACTGATACG	CCAAGGCTGAGATAATCAATC	181
82823	ATGCCGTGTCATCCAAGAG	ATCTGAGAGGAGCCAAGACC	102
85423	TCTGCCAGGTCATGCATCAG	GGAAACTGGAGGATGTAG	124
87097	TAACCTCACAGACCTTCC	GTTTCCTCGTCGCACACGAC	144
88663	CCGCAATGATGTTGAGAAG	ATTGTTTGCTTGCCACGG	275
89015	ATCCAGAGTCTCCAAGTGC	TGGACTTCCATGCTTTGAGG	173
89748	TTACACTGCGTGCCTCTTC	ATGTTACACCAGCCTTGG	101
89755	AGTTAGGCACTCTGACTG	GCTGCTTAGAGATTCCAG	95
89946	CCAGAACACATTGGACAC	GAGAGAAGAAGCAACGAC	300
221748	CTGGCATAGAGTTTCTCG	CTGGTAGTGGAACAGTAG	193
216052	GAATCACCAGGCATCATC	GTTTCTTCGGTTGTCCATAG	129
40014	AGAGAGCCAGGAGATAAGC	AGAACGAAGTGGTGGAAGC	153
40014 (redesigned)	AGAGCCAGGAGATAAGCAC	AGAACGAAGTGGTGGAAGC	151
293773	CAGCCTAAAGCCACTATC	CAGTGGACATAAGCAAGC	131
282725	TCTCGGTGAGTTCAAGAC	GCTACTCTCTTCACTTCC	209
267101	AGTGTAAAGGTGACGGAGC	GCATCATTGTGGCATCTATC	258
266585	GCTGGGATGAGCAATAAC	GCAACCAACAACATCAGTAG	57
232092	CCATCCTGTAGCATCTTTG	TCGTTTGACAAGCAAGCG	120
228901	ATAACAGGCAGGTCTTGG	TGCTTCTACTGTGCCTTC	180
209129	CATCTTGCTCAATCTCGG	CATCAGGAAGTTGTGTCAG	157
209129 (redesigned)	CTTCCATTCCCATTCCTC	GTTGACACAAGCCATCAG	211
207290	CGTAGTGTATTGCTTGACAGTG	ACTCCTGGTTCGCATCTTCC	175
206858	GATGTAGTTGAGGAGGATTG	ACCTTGCTTCCTTTGCTC	68
206341	GGCACTTGAGAGATTACATC	CAAATGAGGTCTATCGCAAG	112
204647	GACACACTGAAGCAAGTC	GCACCAACATCATCATACC	111
195055	GACAGTAACCAACCAACG	TAGTCTACAAGGGAGTGC	153
189911	CCTACATCTATTGGTTGACTCG	ACCTTCAGGCTCAGGAATGC	81
188349	GGTAAGAGTGGATAGTCAG	GACAACATTGGAACTGTGG	128
187263	CCAATAGAGTCATAGACCC	CAGAAATGTGGTGGCATAC	98
182447	GACAAAGATGCTGGCAAC	CTCCACGACTTCAGAATC	169
172952	ACTGACGGCAGCATCTTTC	AGTTTGGCAGACCACAAC	113
172362	CCTACAAGCATTAAACTCTCC	ACTGAGTCGGTCTACAAGG	116
168999	GAATCATGCACGAGTTCATG	CTTGCTGGACTTGTACGCTG	154
166225	CTATTGTTGGTGTGTTGGC	TCGGCATTGAGCATCTCTC	144
165468	GATTCGGTCTTCAAAGTCAG	GCACCATACAAGTTGGAG	81

164720	TTGACTGAGGTTGGCTTC	CTGGTTGTAATCGTCTCG	205
162324	CACCAACACCTCCTTATC	GAGAGTTCTGTCCATAGG	111
160284	ACAGTTCTTCAGCCATCC	CTCAGGTGATGCAGATGT	148
156447	TCTCTTCGCTTGACTTGG	ATCTGGAGGTTGATAGCCG	199
156115	TCCTTGAAGCAATCCCTCC	GGTGCTTCCTATATGCAGC	145
155856	GACAGCAATACCTGAACTTC	CGCTTATGTTGGTGATTACAC	194
155001	CAGGTTGAAGTTGGTGTC	GAAGATGAAGATGACGCTC	166
153615	AGCGAGAGAACTACACAGG	GGTCAGAGGGATAAGAAACTCAAC	111
153615	GAGTTTCTTATCCCTCTGAC	CCACAGTCTCTACGAGTG	106
150007	TGCCTCATTCATCCAGCAC	CTTATTCGTCTCCTCAGG	152
148307	CAAGAGCATCAACAACAAGC	ATTTCTCGTCCACGCAGCTTG	128
148270	CAGTTGGTGAGAAGGTTC	TAGAGTATGCGAGGTCTG	118
147259	TGCTCCTTGTATCTCCCAG	CTATCAAGAACTTCACCCG	104
146074	ATGGCGTCCTCATCTGCTAC	CAGGTTGAAAGACACTTGG	192
144794	AGTATCTACGGGTGCTTGCC	TTGAATCTCTGCTGGGAGC	143
144413	GATAGTTTGGTGGTGGTC	CTCGTTATGAGAGAGGTC	234
142887	CAGTTGTTCACGGATGATG	GCTCAACATTCTGAACCTG	122
134492	GCCACTTTGGTAATACAGTTCC	TGGACCTGTTTGTGGGTG	116
131838	GGCACATCAAAGAACTGC	TTCGGTCAACTTGCTCCTG	161
131602	GTTCTGGGAAACTGGTAG	CTTGTCATTGTTGGCGTC	238
131599	GCGTGAAAAATGTAGTGAGAGC	ACAACCCGTAGCATTGAGC	142
129013	CTATTGGAACAAGGCTACC	TGGCTGGTGGTATGTTCAG	168
128876	GAGTGAAGGATTTGAACC	GCTGCTTCTTCTTTTGTATG	99
128533	CACTCTCTTATTCTTCGCC	CGCTATTCACTCTTCCTTACTC	376
128533 (redesigned)	GGTCACTCTCTTATTCTTCG	CGCTATTCACTCTTCCTTAC	379
125606	TGTTGCCATTGACCCACGAG	AATCGCCCTTTAGCCTCTGG	197
123727	CCACTTATATGCCAGCTGA	ATGTCCTTCTTTGGCTCG	146
123424	GTGTGGTGGTTATTATCACG	AAGCAGGGAGCATACAAC	162
121908	TGATGAACCCATTCCTCCCG	TCCAAGAACGCCAACACGAC	189
121571	GCTTGTCCACATCAATCC	GAGGAGACTGGTATCTTAC	121
120401	GGCACTTATCTCACTGAAG	AATCCGCTTGAACTCTCG	159
119638	GCTTGGGATACACGATAC	GCCAAACCAAAGTCAGAG	158
118717	CTGCTGCTTCTTCTTCAC	CAGCTATATCACGCTTCCAC	130
118591	GGAATGGCTCATAATCAGC	ATCAGGTGAAGGACTCCAG	260
118416	CCTAATCCACCAAATCCAAC	TGGCATTGACGAGAAGAGC	108
116728	GTCTTGAAGTAGTGCCAG	TCATAGTGACGAGCAGGTC	103
109407	GTCTTTAGATGAGAAGGCAG	CAATGGCTCTTGAAGTGG	216
103237	GCCTGATTTTGACAAGGAAC	ATCTCTCTCCCAACTGAG	102
103013	GATTCTTTAGAGATGGGAGC	ATGAGTGGAGGTTGCTTC	183
101038	CAAGAGGACACTGAAGATAC	CATTGAAACAAGTCAGAGCC	193
101038 (redesigned)	CAAGAGGACACTGAAGATAC	TAATCGGCGAGTTTGTTG	223
100479	CCAATGGTGACTAACTCTG	TGGGAGGATGTTCTGAAG	212
99584	CGACTACCATTAGCTACATC	CACGAAGATGAGTCAGTTG	173
99584 (redesigned)	CTTTCCGACCTCATACTC	TCATAGGGCACGAAGATG	56
99577	CGCACCTACTTCTACTAC	TGGTCGGAAGGTTGAATG	164
99333	CAATCTTTGGAGGCTCAAG	GAATGGCGACTCAATAACC	157
94540	CTGCTGTCCTTCTCTATC	GAAACTGGTGCTGAATGAC	168

92973	CCCAATCCAACTCTTGAC	CCATTTTAGCAGATGAAGCC	182
91787	AGAAGGTGAGGCTCTTGAG	GGTGTCATGTGATCTCTTC	146
91633	GTCTTGTTTCACTCCAGTTG	GGGTGCTTCTTCACTTCATTTTCC	154
91604	CCAGATTTCCAAGGTTCG	GACACTTCATCGTCTGAG	103
91181	TCGAGCCGACGAGGAAGGAA	TCAGGTTCGGCCGGATGTA	123
90663	CTGAGTTTGGAAAGAGGC	GGGAACATTGGCAAGATG	241
90175	CTTACATAGCAACTGAAGCG	TGGGTGAGGAGATTCTTG	206
90046	CGGTGTAGTTGTTGTGATG	AGGCTTCAGTTCTTGGAC	112
326194	CCAAAAAGTGGAAGGCTG	AGTCTCCAACTGCCTTAGC	132
a205149	TAATGGAGAGAGGCTGTG	GAGGTTAGCATCAGAAGG	222
a211607	CAATAGCGAACTCATCACC	CATACAGAAGAAGAAGCGG	217
a213409	TTCCGATTGTTGGTCCAGG	AGCACCTTATCCAGACTG	136

Appendix C - Custom R Script used to align sequence tags and pull out SNPs from the UNEAK and Tassel pipelines

```
## import tassel data
data =
 read.table(file="~/Dropbox/Nicole_Pipeline_Files/Tassel_hap_files/Run_
 3/LR16_dif3_20130604.hap", header=TRUE)
data[1:5,]
## import UNEAK data
data =
 read.table(file="~/Dropbox/Nicole_Pipeline_Files/UNEAK_HapMap_files/Ru
 n_2/HapMap.hmp.txt", header=TRUE, check.names=FALSE)
data[1:5,]

## checking that CS alleles match
data$CS != data$CS.2 & data$CS != "N"  & data$CS.2 != "N"
## remove CS alleles that do not match
data = data[ !(data$CS != data$CS.2 & data$CS != "N"  & data$CS.2 !=
 "N"),]

## change missing data label N to NA
data[data=="N"] = NA

## import tassel phenotype
pheno =
 read.csv(file="~/Dropbox/Nicole_Pipeline_Files/Lr16xCS_Phenotype_Score
 s.csv", header=TRUE)
## import UNEAK phenotype
pheno =
```

```
 read.csv(file="~/Dropbox/Nicole_Pipeline_Files/UNEAK_HapMap_files/Lr16
 xCS_Phenotype_Scores.csv", header=TRUE)

## transpose data and make new matrix 'geno'
geno = t(data)

## assign plant names (labeled Plant_Number) as the row names instead
 of 1,2,3...
rownames(pheno) = pheno$Plant_Number

## merge pheno and geno on row names
mg = merge(pheno,geno,by="row.names")

## make geno array removing 'extra' columns (1-5) from merged data
geno = mg[,-c(1:5)]

## designate SNP alleles as allele1 or allele2
allele1 = substring(data$alleles,1,1)
allele2 = substring(data$alleles,3,3)

## join phenotype to the genotypes (data)
mg = merge(pheno,t(data),by="row.names")

## create TRUE/FALSE array for the phenotypes
res = mg$Nr_F3=="R"
suc = mg$Nr_F3=="S"

## testing marker #1
## make T/F array for marker is allele 1 or 2
```

```
mrk.a1 = geno[,1] == allele1[1]
mrk.a2 = geno[,1] == allele2[1]

##creating array to compare resistant phenotype array and array for
 marker 1 and 2
test.a1 = res == mrk.a1
test.a2 = res == mrk.a2

##gives the sum of the number of resistant phenotypes with allele 1
 with the missing data removed
sum(test.a1, na.rm=TRUE)
##gives the sum of the number of resistant phenotypes with allele2
 with the missing data removed
sum(test.a2, na.rm=TRUE)

### loop through all of the markers
for(i in 1:ncol(geno)){
 mrk.a1 = geno[,i] == allele1[i]
 mrk.a2 = geno[,i] == allele2[i]
 test.a1 = res == mrk.a1
 test.a2 = res == mrk.a2

 s1 = sum(test.a1, na.rm=TRUE)
 s2 = sum(test.a2, na.rm=TRUE)
 ##print(cbind(s1, s2, data[i,1:15]))
 if((s1<=1 | s2<=1) && (s1>5 | s2>5)){
 ##  print(s1)
 ##  print(s2)
 ##  print(data[i,])
```

```
    print(cbind(s1, s2, data[i,1:15]))
 }

}

##loop through all markers and check for perfect correlation
for(i in 1:ncol(geno)){
 mrk.a1 = geno[,i] == allele1[i]
 mrk.a2 = geno[,i] == allele2[i]
 test.a1 = res == mrk.a1
 test.a2 = res == mrk.a2

 s1 = sum(test.a1, na.rm=TRUE)
 s2 = sum(test.a2, na.rm=TRUE)
 ##print(cbind(s1, s2, data[i,1:15]))
 if((s1==0 | s2==0) && (s1>5 | s2>5)){
 ##  print(s1)
 ##  print(s2)
 ##  print(data[i,])
    print(cbind(s1, s2, data[i,1:15]))
 }

}
```

Appendix D - KASP Primer Sequences

GBS Contig	Forward with Fam
SNP16_8	GAAGGTGACCAAGTTCATGCTCTGCAGAACCCGAAGACC
SNP16_92	GAAGGTGACCAAGTTCATGCTGTGTGAGCGTGTGCAACTGAAAAA
SNP16_112	GAAGGTGACCAAGTTCATGCTCTGCAGATTGCCCGTCAGCC
SNP16_126	GAAGGTGACCAAGTTCATGCTATGCTGCTGTTCTTGAGGAGTCT
SNP16_155	GAAGGTGACCAAGTTCATGCTATATGTCTCTTTGGTCTCGC
SNP16_180	GAAGGTGACCAAGTTCATGCTTCCGCAGGTATGCCGCTCC
SNP16_296	GAAGGTGACCAAGTTCATGCTTCAACGTCACTTTCTTAGCC
SNP16_362	GAAGGTGACCAAGTTCATGCTAGTAGTAATAAAGTAAATAAAGTGT
SNP16_420	GAAGGTGACCAAGTTCATGCTGTCCTGCAGGGCACAACCCA
SNP16_128	GAAGGTGACCAAGTTCATGCTAGTTCTTCGAGCAACGAGCG
SNP16_129	GAAGGTGACCAAGTTCATGCTATCTTGTTCCTCCAGTTCTT
SNP16_115	GAAGGTGACCAAGTTCATGCTATCAGTCGTAACTCATCACTA
SNP16_116	GAAGGTGACCAAGTTCATGCTTCAGGTAAGCTCCTCCTCTT
SNP16_206	GAAGGTGACCAAGTTCATGCTTCTCATAAGATTTTTTTTCTCCGTCC
SNP16_207	GAAGGTGACCAAGTTCATGCTTGGGACGGAGAAAAAAAATCT
SNP16_TP1456	GAAGGTGACCAAGTTCATGCTCGTGGGCTGATCCAAAAAAG
SNP16_TP2102	GAAGGTGACCAAGTTCATGCTATTATACTTGAATTGACAGTAGGCG
SNP16_TP4632	GAAGGTGACCAAGTTCATGCTGGACTAGATCCGTGGGGGAAGAAAT
SNP16_TP5771	GAAGGTGACCAAGTTCATGCTGACGGCCGCCCCAGCTTC
SNP16_TP10706	GAAGGTGACCAAGTTCATGCTTGAGAAGGCTGCAGATCC
SNP16_TP15392	GAAGGTGACCAAGTTCATGCTCATGCTGCTGTTCTTGAGGAGTCT
SNP16_TP17021	GAAGGTGACCAAGTTCATGCTAGGCTCCAGGAAGGGGATCTCG
SNP16_TP18483	GAAGGTGACCAAGTTCATGCTGTTTCTGTAGCAGATTTAGAGCAAA
SNP16_TP21706	GAAGGTGACCAAGTTCATGCTAGTAGATTGCCAAGCCGACT
SNP16_TP29515	GAAGGTGACCAAGTTCATGCTGCCATGCGACCCGACCG
SNP16_TP35107	GAAGGTGACCAAGTTCATGCTGTGGCGACGAATAGTACAACCCC
SNP16_TP46099	GAAGGTGACCAAGTTCATGCTACCCACACGTCAGAGTTGAC
SNP16_TP48461	GAAGGTGACCAAGTTCATGCTCAGTCAGTCAGCCAGCTAATAAAAC
SNP16_TP53756	GAAGGTGACCAAGTTCATGCTGACCCCCATTCCCAGAACGC
SNP16_TP57547	GAAGGTGACCAAGTTCATGCTCAGACAGAGACAGAAGCGTGTTTTT
SNP16_TP58540	GAAGGTGACCAAGTTCATGCTCGCCTCACCGACGTGTCT
SNP16_TP58730	GAAGGTGACCAAGTTCATGCTTGCAGGGCTAGATCACTACC
SNP16_TP64737	GAAGGTGACCAAGTTCATGCTGCACCTTGGGGTTCGTCAATC
SNP16_TP65314	GAAGGTGACCAAGTTCATGCTGAATTGATGGCGGCAGAAGAAGAAG
SNP16_TP65824	GAAGGTGACCAAGTTCATGCTGCCTGGGCCACACATAAGA
SNP16_TP67062	GAAGGTGACCAAGTTCATGCTGTGTTAAAAAATACTTGTCAAACAA
SNP16_TP69388	GAAGGTGACCAAGTTCATGCTAACTTGTACATCTCCATGGGTGTC
SNP16_423	GAAGGTGACCAAGTTCATGCTAAATCGCCCCGAGCGTCG
SNP16_537	GAAGGTGACCAAGTTCATGCTCGCCAGCAGCATGGTTGG

GBS Contig	Forward with Hex
SNP16_8	GAAGGTCGGAGTCAACGGATTCTGCAGAACCCGAAGACT
SNP16_92	GAAGGTCGGAGTCAACGGATTGTGTGAGCGTGTGCAACTGAAAAC
SNP16_112	GAAGGTCGGAGTCAACGGATTCTGCAGATTGCCCGTCAGCG
SNP16_126	GAAGGTCGGAGTCAACGGATTATGCTGCTGTTCTTGAGGAGTCG
SNP16_155	GAAGGTCGGAGTCAACGGATTATATGTCTCTTTGGTCTCGT
SNP16_180	GAAGGTCGGAGTCAACGGATTTCCGCAGGTATGCCGCTCT
SNP16_296	GAAGGTCGGAGTCAACGGATTTCAACGTCACTTTCTTAGCT
SNP16_362	GAAGGTCGGAGTCAACGGATTAGTAGTAATAAAGTAAATAAAGTGC
SNP16_420	GAAGGTCGGAGTCAACGGATTGTCCTGCAGGGCACAACCCA
SNP16_128	GAAGGTCGGAGTCAACGGATTAGTTCTTCGAGCAACGAGCA
SNP16_129	GAAGGTCGGAGTCAACGGATTATCTTGTTCCTCCAGTTCTG
SNP16_115	GAAGGTCGGAGTCAACGGATTATCAGTCGTAACTCATCACTG
SNP16_116	GAAGGTCGGAGTCAACGGATTTCAGGTAAGCTCCTCCTCTC
SNP16_206	GAAGGTCGGAGTCAACGGATTTCTCATAAGATTTTTTTTCTCCGTCA
SNP16_207	GAAGGTCGGAGTCAACGGATTTGGGACGGAGAAAAAAAATCC
SNP16_TP1456	GAAGGTCGGAGTCAACGGATTCGTGGGCTGATCCAAAAAAA
SNP16_TP2102	GAAGGTCGGAGTCAACGGATTATTATACTTGAATTGACAGTAGGCA
SNP16_TP4632	GAAGGTCGGAGTCAACGGATTGGACTAGATCCGTGGGGGAAGAAAC
SNP16_TP5771	GAAGGTCGGAGTCAACGGATTGACGGCCGCCCCAGCTTT
SNP16_TP10706	GAAGGTCGGAGTCAACGGATTTGAGAAGGCTGCAGATCT
SNP16_TP15392	GAAGGTCGGAGTCAACGGATTCATGCTGCTGTTCTTGAGGAGTCG
SNP16_TP17021	GAAGGTCGGAGTCAACGGATTAGGCTCCAGGAAGGGGATCTCA
SNP16_TP18483	GAAGGTCGGAGTCAACGGATTGTTTCTGTAGCAGATTTAGAGCAAG
SNP16_TP21706	GAAGGTCGGAGTCAACGGATTAGTAGATTGCCAAGCCGACG
SNP16_TP29515	GAAGGTCGGAGTCAACGGATTGCCATGCGACCCGACCA
SNP16_TP35107	GAAGGTCGGAGTCAACGGATTGTGGCGACGAATAGTACAACCCT
SNP16_TP46099	GAAGGTCGGAGTCAACGGATTACCCACACGTCAGAGTTGAT
SNP16_TP48461	GAAGGTCGGAGTCAACGGATTCAGTCAGTCAGCCAGCTAATAAAAA
SNP16_TP53756	GAAGGTCGGAGTCAACGGATTGACCCCCATTCCCAGAACGT
SNP16_TP57547	GAAGGTCGGAGTCAACGGATTCAGACAGAGACAGAAGCGTGTTTTC
SNP16_TP58540	GAAGGTCGGAGTCAACGGATTCGCCTCACCGACGTGTCC
SNP16_TP58730	GAAGGTCGGAGTCAACGGATTTGCAGGGCTAGATCACTACT
SNP16_TP64737	GAAGGTCGGAGTCAACGGATTGCACCTTGGGGTTCGTCAATT
SNP16_TP65314	GAAGGTCGGAGTCAACGGATTGAATTGATGGCGGCAGAAGAAGAAT
SNP16_TP65824	GAAGGTCGGAGTCAACGGATTGCCTGGGCCACACATAAGC
SNP16_TP67062	GAAGGTCGGAGTCAACGGATTGTGTTAAAAAATACTTGTCAAACAT
SNP16_TP69388	GAAGGTCGGAGTCAACGGATTAACTTGTACATCTCCATGGGTGTG
SNP16_423	GAAGGTCGGAGTCAACGGATTAAATCGCCCCGAGCGTCA
SNP16_537	GAAGGTCGGAGTCAACGGATTCGCCAGCAGCATGGTTGT

GBS Contig	Reverse
SNP16_8	AAATAGAACCGTGGGCTG
SNP16_92	TTCTGGTGTGGTAGCAGCCGAC
SNP16_112	TTGTTCCTCCAGTTCTGCGAGC
SNP16_126	AAGTCCCTGTGCTGCTGAGC
SNP16_155	CCATTCATACCTTTTGTTGGTC
SNP16_180	TGTGGCATCGCAAGCAGAG
SNP16_296	CAATCAAAGGTCGGCATC
SNP16_362	CAGGAGATAAGCCAGAAC
SNP16_420	ATGACGACGACGACGACAGCGTTC
SNP16_128	CGATGGTATCTTGACAAATGGATTC
SNP16_129	TGACAAATGGTTGCGATG
SNP16_115	GGAAGGAAGGAACTCAGG
SNP16_116	CTGAAACTGAACGCCACC
SNP16_206	GCTGCTGCCTTTACAGTTATTTTG
SNP16_207	CGTGTCATCCAACTAACTG
SNP16_TP1456	TACTCTGGTGGCTCTTGC
SNP16_TP2102	TGGGACAACAAGTGCTGC
SNP16_TP4632	AGCAGGTGATTGGGAAACGTG
SNP16_TP5771	AGTAATCGCCCTTGCCTCGG
SNP16_TP10706	TGTGCTTCTGTGGTAACC
SNP16_TP15392	AGTCCCTGTGCTGCTGAGC
SNP16_TP17021	ACAGACGCTGCCAGGAGAGG
SNP16_TP18483	TCTCCCCGACCGACTCTTG
SNP16_TP21706	CCCATTTTCCAATCCTATCACTAAG
SNP16_TP29515	ACCACCCTTCTGCTCTCTGC
SNP16_TP35107	TGGGCTCTGACGGTTATGGC
SNP16_TP46099	TCACTGACAGGTGGGTCCG
SNP16_TP48461	GCGTGATGTCCATTGTGC
SNP16_TP53756	CCGCACTTGACCTTTGGC
SNP16_TP57547	GCAGGGAGTGTTACCGACC
SNP16_TP58540	AGGAAGAAGGGGTCGGGGAC
SNP16_TP58730	CGGACGAGTTCGCAATGG
SNP16_TP64737	TCGGTCATCAACGCTCGC
SNP16_TP65314	AACCCCAGGAGAAGGTGGC
SNP16_TP65824	TACGGGTCGTGGGGAATCG
SNP16_TP67062	GCACCAAAACAAATCACC
SNP16_TP69388	CGGAGGGACCAAGGATTC
SNP16_423	AGATTCGGCAACCTGGAGT
SNP16_537	GGCGATGTCCTCGTCTTTG