2019

Disease Resistance in Wheat and Its Relatives

Ethan James Andersen
South Dakota State University

DISEASE RESISTANCE IN WHEAT AND ITS RELATIVES

BY

ETHAN JAMES ANDERSEN

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy

Major in Biological Sciences

Specialization in Molecular Biology

South Dakota State University

2019

DISEASE RESISTANCE IN WHEAT AND ITS RELATIVES

ETHAN JAMES ANDERSEN

This dissertation is approved as a creditable and independent investigation by a candidate for the Doctor of Philosophy in Biological Sciences degree and is acceptable for meeting the dissertation requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidates are necessarily the conclusions of the major department.

Madhav P. Nepal, Ph.D.
Dissertation Advisor Date

Volker Brozel, Ph.D.
Head, Dept. of Biology & Microbiology Date

Dean, Graduate School Date

This dissertation is dedicated to my wife, Alecia Andersen, and my parents, Kevin Andersen and Chris Andersen.

ACKNOWLEDGEMENTS

This research endeavor would not have been possible without the direct and indirect assistance of many individuals. First, I would like to acknowledge the contributions of my Ph.D. advisor, Dr. Madhav Nepal, who has helped me develop into the research professional I am now. My committee members Dr. Shaukat Ali, Dr. Emmanuel Byamukama, Dr. Stephen Gent, and Dr. Yang Yen have assisted in the formulation of research projects, especially by giving me constructive feedback throughout my program. Dr. Brian Moore and Alan Carter also assisted greatly by helping me to use the High Performance Computing Cluster on campus. Much of this research would not have been possible without that resource to run scripts. Several people have assisted me at various points along this journey with research. Graduate students Benjamin Benson, Surendra Neupane, Achal Neupane, and Jordan Purintun helped at many places throughout my program. Undergraduate students Samantha Shaw, Liz Opitz, Lauren Lindsey, and Matthew Hummel have also provided valuable assistance with projects. Dr. Scott Pedersen and Mr. Andrew Ellis were tremendous mentors to me in teaching undergraduates and other aspects of professional development. I would especially like to acknowledge my wife, Alecia Andersen. Her support throughout this project was beyond what could be expected from any one individual and I greatly appreciate it. My parents, Kevin Andersen and Chris Andersen, have been extremely supportive through this endeavor, providing me with the encouragement to persevere through challenging times. They provided me with the experience of growing up on a cattle ranch, which caused me to fall in love with nature and biology in the first place. Most importantly, I would like to display gratitude to my Lord Jesus Christ, through which all of this was possible.

CONTENTS

ABBREVIATIONS

ABA	Abscisic Acid
AVR	Avirulence
bHLH	Basic Helix-Loop-Helix
CC	Coiled-Coil
CNL	CC-NBS-LRR
DAMP	Damage-Associated Molecular Pattern
DIMBOA	2,4-Dihydroxy-7-Methoxy-1,4-Benzoxazin-3-one
ET	Ethylene
ETI	Effector-Triggered Immunity
HMM	Hidden Markov Model
HR	Hypersensitive Response
HST	Host-Selective Toxin
ID	Integrated Domain
JA	Jasmonic Acid
K_a	Nonsynonymous substitutions per nonsynonymous site
K_s	Synonymous substitutions per synonymous site
LRR	Leucine-Rich Repeat
MAPK	Mitogen-Activated Protein Kinase
MEGA	Molecular Evolutionary Genetics Analysis
MEME	Multiple Expectation maximization for Motif Elicitation
MLA	Mildew Locus A
NBAC	NB-ARC-Containing
NB-ARC	Nucleotide-Binding site found in Apoptotic protease activating factor 1, R genes, and *Caenorhabditis elegans* death-4 protein

NBS	Nucleotide-Binding Site
NLR	Nucleotide-binding site Leucine-rich repeat Receptor
PAMP	Pathogen-Associated Molecular Pattern
P-loop	Phosphate-binding loop
PR	Pathogenesis-Related
PRR	Pattern Recognition Receptor
PTI	PAMP-Triggered Immunity
Ptr	*Pyrenophora tritici-repentis*
QTL	Quantitative Trait Locus
R genes	Resistance genes
RNAi	RNA interference
RNBS	Resistance Nucleotide-Binding Site
RNL	RPW8-NBS-LRR
ROS	Reactive Oxygen Species
RPW8	Resistance to Powdery Mildew 8
SA	Salicylic Acid
TF	Transcription Factor
TIR	Toll/Interleukin-1 Receptor
TNL	TIR-NBS-LRR
TS	Tan Spot
Tsn1	Tan spot necrosis 1
WAK	Wall-Associated Kinase
ZF	Zinc Finger

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

DISEASE RESISTANCE IN WHEAT AND ITS RELATIVES

ETHAN JAMES ANDERSEN

2019

Plants have evolved a complex defense system against pests and pathogens utilizing many types of receptors, signaling factors, and defense compounds to detect pathogen presence and respond effectively. Since many pathogens have evolved immune-suppressing effectors used to reduce plant resistance, plants have evolved a family of receptors that detect pathogenic effectors as a result of an evolutionary arms race. These receptors contain Nucleotide-Binding Site and Leucine-Rich Repeat domains and are called NBS-LRR or NLR proteins. Many grasses possess huge genomes with hundreds of NLR-encoding genes, often found in clusters at the extra-pericentromeric regions of chromosomes, where unequal crossing over causes tandem duplication and a mechanism for resistance (R) gene diversification. R genes also possess domains associated with signaling factors that either serve as baits for pathogen effectors or as active signaling components to initiate defense responses. The objectives of this dissertation project were to: 1) identify R genes in wheat and some of its relatives; 2) assess how they may have evolved in grasses with available genomes, such as wheat, barley, foxtail millet, and rice; 3) investigate integrated domains found in wheat NLRs; and 4) assess differences in gene expression between tan spot resistant and susceptible wheat when exposed to *Pyrenophora tritici-repentis* (*Ptr*), a pathogen that uses wheat R gene *Tsn1* as a susceptibility gene to facilitate infection. Genomic data was analyzed by the construction of Hidden Markov Model profiles and sequence annotation using programs such as

InterProScan and MEME (Multiple Expectation maximization for Motif Elicitation). R genes were used to construct phylogenetic trees, chromosomal maps, exon-intron diagrams, and syntenic maps. The interaction between wheat and *Ptr* was investigated through a greenhouse experiment, where resistant and susceptible wheat was inoculated with *Ptr* spores or directly infiltrated with the Ptr ToxA protein. Genome-wide identification of R genes revealed that *Hordeum vulgare*, *Setaria italica*, *Aegilops tauschii*, and *Triticum aestivum* had 175, 202, 402, and 802 NLRs with N-terminal Coiled-Coil domains (CNLs), respectively. In each species studied, R genes formed clusters, many containing highly similar genes, providing evidence of tandem duplication. CNLs in wheat and wheat relatives formed an expansion of the CNL-C clade that showed evidence of purifying selection. R gene sequences necessary for effector detection diversified, while domains necessary for signaling remained conserved. Wheat NLRs contained integrated domains (IDs) associated with kinase, transcription factor, and other signaling mechanisms. Wheat NLR-ID genes encoded multiple transcripts, indicating that wheat is able to include or exclude IDs through alternative splicing. Greenhouse experiments revealed several groups of genes that differed in expression between tan spot resistant and susceptible cultivars. Wheat resistant to Ptr showed increased expression of genes associated with resistance: chitinases, signaling factors (i.e. transcription factors and kinases), resistance receptors, and enzymes associated with phytoalexin production. These results all showed that resistance in wheat and its relatives relies on a complex network of factors, and that NLRs have diversified to be a variable family of components that initiate defense responses when triggered by pathogenic effectors.

INTRODUCTION

Plants have evolved numerous strategies to detect and respond to biotic stress imposed by pests and pathogens [1-3]. Elicitors from pathogens trigger host receptor proteins, which utilize signaling mechanisms to initiate gene expression linked to defense responses [4]. Plants utilize several types of receptors to detect the following elicitors: pathogen components, damage caused by infection, and effectors that pathogens use to inhibit host defenses [5-7]. These receptors then initiate signal transduction pathways, often including kinase cascades and transcription factors, in order to trigger the expression of genes associated with various resistance responses, such as hypersensitive response or phytoalexin production [8-10]. This dissertation research focuses on genes involved in disease resistance mechanisms. **Chapter 1** of this dissertation provides a literature review of the components required for pathogen detection, signal transduction, and defense response. This chapter is published in the journal *Genes* [11].

Receptors that detect pathogenic effectors are often encoded by genes referred to generally as resistance genes (R genes), and more specifically based upon domains they encode for: Nucleotide-Binding Sites and Leucine-Rich Repeats (NBS-LRRs or NLRs) [4]. The dynamic nature of the plant immune system's response to quickly evolving pathogens has driven a rapid coevolution of the receptors and corresponding effectors [12-14]. Agriculture relies heavily on cultivars with these genes to provide genetic resistance against diseases that cause yield loss. Grasses play a significant role in agriculture, and wheat, for instance, makes up a large percentage of the human food supply. Other grass crops include barley, rice, corn, and millet. Due to the economic importance of these species, their genomes have recently been sequenced and made available in public

databases [15]. The genomes of some wild relatives of crop species, such as the wheat progenitor *Aegilops tauschii,* have also been made available [16]. **Chapter 2**, **3**, and **4** focus on a genome-wide identification of R genes in *Hordeum vulgare*, *Setaria italica*, and *Aegilops tauschii*, respectively. Several gene annotation programs were employed to identify, map, align, and compare R gene sequences, and multiple programs were used to assess phylogenetic relationships, evidence of duplications, and genome synteny. InterProScan [17] and MEME [18] were used to annotate domains and motifs within R genes. The programs ClustalW [19] and MEGA [20] were used to align and construct phylogenetic trees. SyMAP [21] was used to generate syntenic maps comparing whole chromosome sequences between species. The Gene Structure Display Server [22] and Geneious [23] were used for data visualization, such as exon-intron diagrams and chromosomal clustering. DNAsp [24] was used to measure the ratio of nonsynonymous substitutions per nonsynonymous site to synonymous substitutions per synonymous site. The findings showed that 175, 202, and 402 R genes occur in *H. vulgare*, *S. italica*, and *A. tauschii*, respectively. These genes occurred both in isolation and in clusters, providing evidence of tandem and segmental duplication, respectively. An expansion of CNLs (NLRs with Coiled-Coil motifs) was consistent across all the species analyzed in this study, particularly the CNL-C clade, with members undergoing purifying selection. These three chapters (**2**, **3**, and **4**) are published in the journals *Evolutionary Bioinformatics* [25], *Plant Gene* [26], and *Proceedings of the South Dakota Academy of Science* [27], respectively.

Due to the importance of wheat to the human diet, pathogens reducing wheat yield have become major topics of research. **Chapter 5** focuses on identification and characterization of R genes in the *Triticum aestivum* genome. In addition to the methods

implemented in earlier chapters, Circa (http://omgenomics.com/circa) and Circoletto [28] were used to map genes to chromosomes and generate similarity diagrams between clustered genes, respectively. A total of 2151 potential R genes were found in wheat, with 802 belonging to the CNL class. These genes formed 547 clusters, providing evidence of tandem duplications. In **Chapter 6,** further domain analysis of R genes with potential roles in signal transduction or as baits for pathogenic effectors was elaborated. Gene annotation using InterProScan was used to identify integrated domains, and genomic data regarding alternative transcripts were analyzed to understand possible alternative splicing. Wheat R genes that contain integrated domains (IDs) show evidence that alternative splicing is used to include or exclude IDs, which, based upon sequence homology, are likely functional as signaling domains or baits for pathogenic effectors. **Chapter 6** has been accepted for publication in the *Journal of Botanical Research*.

Chapter 7 focuses on the disease tan spot, caused by the fungal pathogen *Pyrenophora tritici-repentis* (*Ptr*), a necrotrophic fungus (Phylum Ascomycota) [29,30]. This pathogen secretes effectors that cause leaf chlorosis and necrosis, leading to the tan spot symptoms. Wheat cultivars possess varying levels of resistance to *Ptr*, at least partially due to the presence of susceptibility genes in the wheat genome. Wheat cultivars possessing the susceptibility gene *Tsn1* are susceptible to the *Ptr* effector ToxA and develop necrotic lesions [31]. Tsn1 is a cytoplasmic NLR protein, indicating that Ptr ToxA hijacks the wheat defense system to lead to symptom development [32]. Therefore, while NLRs are generally described as resistance genes, some pathogens have developed mechanisms to use them as susceptibility genes. Research presented in **Chapter 7** examined differences between resistant and susceptible wheat cultivars after either

infection with *Ptr* or direct infiltration with an isolate of the Ptr ToxA. Gene expression was analyzed in susceptible wheat cultivar Glenlea and resistant wheat cultivar Salamouni with a greenhouse experiment in which plants were inoculated with Ptr ToxA-producing race 2 or directly infiltrated with Ptr ToxA isolate. The programs FASTQC [33], Btrim [34], Hisat [35], and Htseq [36] were used to quality check, trim, map, and assemble the RNA sequencing data to the pathogen and host genomes. The program R [37] with the Bioconductor package and DESeq2 [38] were used to assess differential gene expression. The results showed that *Ptr*-resistant wheat expressed genes associated with resistance functions, such as chitinases, signaling factors (i.e. transcription factors and kinases), resistance receptors, and enzymes associated with phytoalexin production.

Overall, this dissertation research shows that disease resistance in wheat and its relatives relies on a complex network of factors, with R genes that have evolved to encode a highly variable set of effector-detecting receptors. These findings show that R genes generally exist in clusters at the extra-pericentromeric regions of chromosomes and experience purifying selection. The transcripts that R genes encode are alternatively spliced to include additional domains that extend R protein function as signal transduction factors and baits for pathogenic effector modification. Wheat-pathogen interactions involve the expression of multiple classes of genes that have roles in detection, signaling, and response (e.g. NLRs, kinases, and chitinases). These findings serve as a framework for future analyses of R genes in grasses and will have implications for the development of wheat cultivars with durable resistance.

REFERENCES

1. Dangl JL, Jones JD. Plant pathogens and integrated defence responses to infection. *nature.* 2001;411(6839):826-833.
2. Gururani MA, Venkatesh J, Upadhyaya CP, Nookaraju A, Pandey SK, Park SW. Plant disease resistance genes: current status and future directions. *Physiological and Molecular Plant Pathology.* 2012;78:51-65.
3. Hammond-Kosack KE, Jones JD. Plant disease resistance genes. *Annual review of plant biology.* 1997;48(1):575-607.
4. Jones JD, Dangl JL. The plant immune system. *Nature.* 2006;444(7117):323-329.
5. de Jonge R, van Esse HP, Kombrink A, et al. Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. *Science.* 2010;329(5994):953-955.
6. Sun Y, Li L, Macho AP, et al. Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex. *Science.* 2013;342(6158):624-628.
7. Niehl A, Wyrsch I, Boller T, Heinlein M. Double‐stranded RNAs induce a pattern‐triggered immune signaling pathway in plants. *New Phytologist.* 2016;211(3):1008-1019.
8. Brutus A, Sicilia F, Macone A, Cervone F, De Lorenzo G. A domain swap approach reveals a role of the plant wall-associated kinase 1 (WAK1) as a receptor of oligogalacturonides. *Proceedings of the National Academy of Sciences.* 2010;107(20):9452-9457.
9. Galletti R, Ferrari S, De Lorenzo G. Arabidopsis MPK3 and MPK6 play different roles in basal and oligogalacturonide-or flagellin-induced resistance against Botrytis cinerea. *Plant Physiology.* 2011;157(2):804-814.

10. Kohorn BD, Kohorn SL, Todorova T, Baptiste G, Stansky K, McCullough M. A dominant allele of Arabidopsis pectin-binding wall-associated kinase induces a stress response suppressed by MPK6 but not MPK3 mutations. *Molecular plant.* 2012;5(4):841-851.

11. Andersen E, Ali S, Byamukama E, Yen Y, Nepal M. Disease resistance mechanisms in plants. *Genes.* 2018;9(7):339.

12. Michelmore RW, Meyers BC. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Research.* 1998;8(11):1113-1130.

13. Bergelson J, Kreitman M, Stahl EA, Tian D. Evolutionary dynamics of plant R-genes. *Science.* 2001;292(5525):2281-2285.

14. McGrann GR, Stavrinides A, Russell J, et al. A trade off between mlo resistance to powdery mildew and increased susceptibility of barley to a newly important disease, Ramularia leaf spot. *Journal of experimental botany.* 2014:ert452.

15. Marcussen T, Sandve SR, Heier L, et al. Ancient hybridizations among the ancestral genomes of bread wheat. *Science.* 2014;345(6194):1250092.

16. Jia J, Zhao S, Kong X, et al. Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature.* 2013;496(7443):91-95.

17. Jones P, Binns D, Chang H-Y, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014:btu031.

18. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in bipolymers. Department of Computer Science and Engineering, University of California, San Diego; 1994.

19. Larkin MA, Blackshields G, Brown N, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23(21):2947-2948.

20. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution.* 2011;28(10):2731-2739.

21. Soderlund C, Bomhoff M, Nelson WM. SyMAP v3. 4: a turnkey synteny system with application to plant genomes. *Nucleic acids research.* 2011:gkr123.

22. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics.* 2014:btu817.

23. Kearse M, Moir R, Wilson A, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012;28(12):1647-1649.

24. Rozas J. DNA sequence polymorphism analysis using DnaSP. *Bioinformatics for DNA sequence analysis*: Springer; 2009:337-350.

25. Andersen EJ, Ali S, Reese RN, Yen Y, Neupane S, Nepal MP. Diversity and Evolution of Disease Resistance Genes in Barley (Hordeum vulgare L.). *Evolutionary bioinformatics online.* 2016;12:99.

26. Andersen EJ, Nepal MP. Genetic diversity of disease resistance genes in foxtail millet (Setaria italica L.). *Plant Gene.* 2017;10:8-16.

27. Andersen EJ, Shaw SR, Nepal MP. Identification of Disease Resistance Genes in Aegilops Tauschii Coss. (Poaceae). *Proceedings of the South Dakota Academy of Science.* 2015;94:273-287.

28. Darzentas N. Circoletto: visualizing sequence similarity with Circos. *Bioinformatics.* 2010;26(20).

29. Murray TD, Bockus, W.W., Bowden, R.L., Hunger, R.M., and Smiley, R.W., collators. Diseases of Wheat (Triticum spp. L.). *Common Names of Plant Diseases* 2015; http://www.apsnet.org/publications/commonnames/Pages/Wheat.aspx. Accessed July 28, 2017.

30. Rees R, Platz G. Effects of yellow spot on wheat: comparison of epidemics at different stages of crop development. *Crop and Pasture Science.* 1983;34(1):39-46.

31. Faris JD, Liu Z, Xu SS. Genetics of tan spot resistance in wheat. *Theoretical and Applied Genetics.* 2013;126(9):2197-2217.

32. Shi G, Zhang Z, Friesen TL, et al. The hijacking of a receptor kinase–driven pathway by a wheat fungal pathogen leads to disease. *Science advances.* 2016;2(10):e1600822.

33. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010.

34. Kong Y. Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics.* 2011;98(2):152-153.

35. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nature methods.* 2015;12(4):357.

36. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015;31(2):166-169.

37. Team RC. R: A language and environment for statistical computing. 2013.

38. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology.* 2014;15(12):550.

CHAPTER 1: PATHOGEN RESISTANCE MECHANISMS IN PLANTS

This chapter is published in the journal *Genes*:

Andersen, E.J., E. Byamukama, S. Ali, Y. Yen, and M.P. Nepal. 2018. Disease Resistance Mechanisms in Plants. *Genes. 9*(7): 339; doi: 10.3390/genes9070339

Abstract: Plants have developed a complex defense system against diverse pests and pathogens. Once pathogens overcome mechanical barriers to infection, plant receptors initiate signaling pathways driving the expression of defense response genes. Plant immune systems rely on their ability to recognize enemy molecules, carry out signal transduction, and respond defensively through pathways involving many genes and their products. Pathogens actively attempt to evade and interfere with response pathways, selecting for a decentralized, multicomponent immune system. Recent advances in molecular techniques have greatly expanded overall understanding of plant immunity, largely driven by potential application to agricultural systems. Here, the major plant immune system components, state of the art knowledge, and future direction of research on plant-pathogen interactions are reviewed. In this review, one major focus is how the decentralization of plant immune systems have provided both increased evolutionary opportunity for pathogen resistance as well as additional mechanisms for pathogen inhibition of such defense responses. Conclusions include that the rapid advances in bioinformatics and molecular biology are driving an explosion of data and information at a much faster rate, contributing to agricultural production and useful in addressing complex problems in evolutionary biology in general.

1. INTRODUCTION

1.1. Plant Disease Resistance has Emerged as a Complex, Multicomponent System

The agricultural revolution of approximately 10,000 years ago intensified humanity's relationship with plants. Since survival largely depended on yield, early farmers would have selected plants based on multiple factors, including their ability to resist disease. Multiple plant diseases recorded in ancient times were often attributed to supernatural causes, but phytopathological observations can be traced back to the third and fourth centuries B.C. in the writings of Aristotle's student Theophrastus [1]. Microbiology during the enlightenment enabled systematic classification of pathogenic organisms, newly visible under compound microscopes. Early experiments in the 19th century demonstrated the efficacy of fungicides, such as the Bordeaux mixture of copper sulfate and calcium oxide [2]. Also in the 19th century, pathogenic microbes were demonstrated to be the causal agents of plant diseases by Heinrich Anton de Bary [3]. Understanding of heritability and genetics developed in the early 20th century allowed researchers to identify sources of heritable resistance, call Resistance genes (R genes) [4,5]. R genes were further described by H. H. Flor's ground-breaking Gene-for-Gene model [6], correlated with avirulence genes present in the pathogen that create an incompatible response. Mechanisms for resistance utilizing R genes were then elucidated following advances in chemistry and molecular biology later in the 20th century [7,8]. The advances in molecular techniques and genomics of the early 21st century drove the discovery of numerous classes of genes that encode regulators of disease resistance and susceptibility [9]. R genes were found to be only one set of participants in a web of interacting factors. Recent molecular research has revealed that plant resistance relies on a complex

regulatory system that controls plant defense responses, greatly building upon the simple structure of H. H. Flor's gene-for-gene model. Plant immune system components participate in pathogen detection, signal transduction, or defense response. Selection pressures drive the evolution of plants with complex detection systems and pathogens with sophisticated evasion techniques, as outlined in the Zig-Zag model [10]. This review focuses on these components and how they are involved in immunity. This review does not aim to completely exhaust every aspect of each component, as whole reviews can and are dedicated to a single class of one factor. Instead, the conceptual framework of phytopathogen resistance is presented, as supported by examples. The first area this review describes is which components are involved in this plant immune system and how pathogens have evolved to subvert defenses.

1.2. Plant Defense Relies on Detection and Response

Investigation into the molecular basis of pathogen resistance reveals a suite of cellular receptors that performs direct detection of pathogenic molecules. Pattern Recognition Receptors (PRRs) within the cell membrane detect pathogen-associated molecular patterns (PAMPs) and Wall-Associated Kinases (WAKs) detect damage-associated molecular patterns (DAMPs) that result from cellular damage during infection [11,12]. Receptors with Nucleotide-binding domains and Leucine-rich Repeats (NLRs) detect effectors that pathogens use to facilitate infection [13]. PRRs, WAKs, and NLRs initiate one of many signaling cascades that have yet to be completely elucidated. Mitogen-Activated Protein Kinases (MAPKs), G-proteins, ubiquitin, calcium, hormones,

Transcription Factors (TFs), and epigenetic modifications regulate the expression of pathogenesis-related (PR) genes [14-18]. This leads to various responses that prevent further infection: Hypersensitive Response (HR), production of Reactive Oxygen Species (ROS), cell wall modification, closure of stomata, or the production of various anti-pest proteins and compounds (e.g. chitinases, protease inhibitors, defensins, and phytoalexins) [19,20]. As now understood from molecular techniques, pathogen resistance in plants involves various organelles and classes of both proteins and non-protein compounds, each of which are required to regulate defense response (see **Figure 1.1**). Factors in each of these roles affect various other signaling systems, such as growth and abiotic stress response. An improved understanding of plant-pathogen interaction requires a full description of these molecular interactions that take place when a compatible pathogen interacts with plant tissue. First, however, the pathogens that elicit these responses must be described and how their evolution has led to the complex immune system that plants possess.

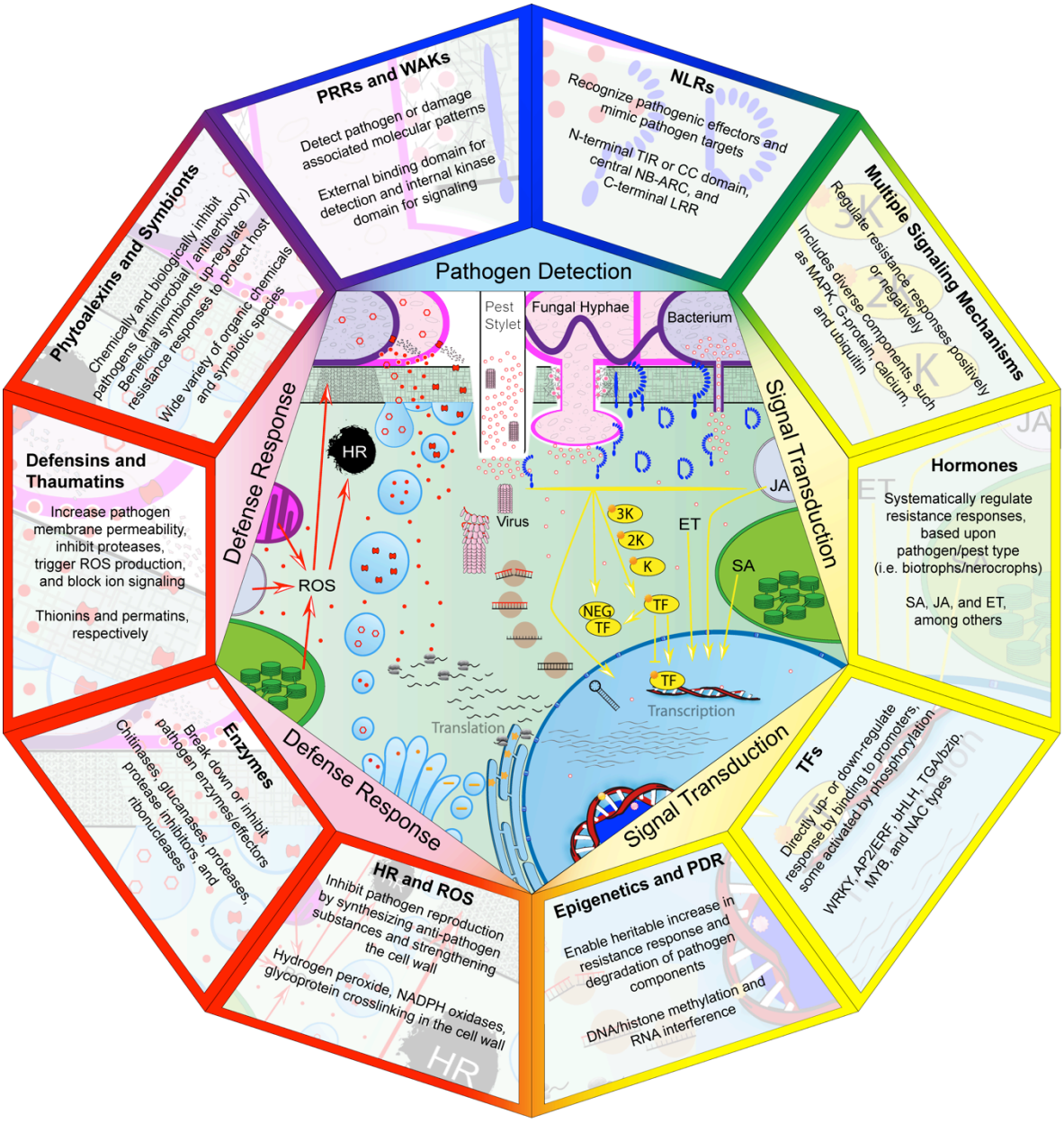


Figure 1.1 Components of plant pathogen resistance pathways involved in pathogen detection, signal transduction, and defense response (detection in the upper center and progressing around clockwise, ending in defense response in the upper left). Pathogenic elicitors (cell components or effectors) produced by bacteria, fungi, insects, nematodes, or viruses trigger plant receptors to initiate signaling cascades. Activated receptors (blue) then initiate one of many signal transduction pathways or

directly act as transcription factors (TFs). Signal transduction pathways (yellow) include MAPK cascades, calcium ion signaling, hormone production, TF activity, and epigenetic regulation. These factors trigger the expression of genes associated with defense responses, such as those regulating the production of ROS, antimicrobial enzymes, defensins, and phytoalexins. These defense-related compounds (red) actively inhibit pathogen reproduction or make further infection more difficult. Breakdown of pathogenic cell components by defense compounds leads to further release of receptor-triggering elicitors, increasing the resistance response. Multiple organelles are involved in defense response, including chloroplasts and peroxisomes for hormone production and the nucleus, endoplasmic reticulum, and Golgi apparatus for antimicrobial protein production.

1.3. Pathogen Virulence Relies on Attack and Evasion

Diverse species and races of pathogens drive plant populations to evolve a highly varied set of immune receptors and modes of response [8,21,22]. Pathogens respond by evolving mechanisms to evade host perception or negate defense responses [10]. Pathogens often subvert PAMP-Triggered Immunity (PTI) through effectors, which can be secreted through various systems. The bacterium *Xanthomonas translucens*, for example, secretes 20-40 effectors through a type III secretion system, resembling a tube that transfers effectors into wheat cytoplasm [23]. Other members of the *Xanthomonas* genus utilize a type VI secretion system which transfers effectors extracellularly [24]. Pathogens can also increase virulence by transfering effector genes from one species to another, called horizontal gene transfer. Such an event has been proposed to describe the transfer of the

necrosis-inducing *ToxA* gene from *Parastagonospora nodorum* to *Pyrenophora tritici-repentis*, which causes the disease tan spot in wheat [25]. Plant Effector-Triggered Immunity (ETI) detects these effectors and initiates a defense response. Such responses must also be varied as pathogens differ in how they extract nutrients: biotrophs from living tissue and necrotrophs from dead tissue. HR intended to inhibit biotrophs only facilitates necrotroph infection [26]. Hemi-biotrophs trick the plant by switching from biotrophy to necrotrophy [27], producing effectors in waves that differ by tissue and infection stage [28]. The complex system leading to pathogen resistance described in **Figure 1** has been subverted at various levels, with pathogens evolving effectors that compromise many of the various signaling steps. As reviewed next, the first stage of pathogen detection relies upon the recognition of molecules unique to pathogens or resulting from pathogen infection.

2. PATHOGEN DETECTION

2.1. PRRs and WAKs detect PAMPs and DAMPs

PRRs are able to recognize a wide array of microbial components, including fungal carbohydrates [29], bacterial proteins [30], and viral nucleic acids [31]. These receptors often possess: Leucine-Rich Repeats (LRRs) that bind to extracellular ligands, transmembrane domains necessary for their localization in the plasma membrane, and cytoplasmic kinase domains for signal transduction through phosphorylation [11]. LRRs are highly divergent, associated with their ability to bind to diverse elicitors. Many PRRs rely on the regulatory protein Brassinosteroid Insensitive 1-Associated Kinase1 (BAK1) and other Somatic-Embryogenesis Receptor-Like Kinases (SERKs) [32,33]. Extensive signaling

is not always initiated, as some PRRs, upon activation, release kinase domains that travel to the nucleus to trigger transcriptional reprogramming [34]. Molecules detected by PRRs are diverse: bacterial (flagellin, elongation factor EF-Tu, and peptidoglycan) [35-37], fungal (chitin, xylanase) [38,39], oomycete (β-glucan and elicitins) [40,41], viral (double stranded RNA) [31], and insect (aphid-derived elicitors) [33]. Although many of these studies were conducted to elucidate specific molecular interactions in *Arabidopsis*, wheat PRRs TaLRK10, TaRLP1.1, and TaRLK-R1-3 have been associated with resistance to rust (fungi of the genus *Puccinia*) via detection of fungal PAMPs [42-44].

Unlike the PRRs that detect non-self pathogen molecules upon infection, other receptors perceive damage by recognizing cellular components that have been disrupted by pathogenic enzymes. This has been shown in *Arabidopsis* with RKWAK1 perception of oligogalacturonides [45] or DORN1/LecRK-I.9 perception of extracellular ATP [46]. WAKs possess an N-terminal, extracellular galacturonan-binding domain that interacts with pectins in the cell wall and cytoplasmic kinase domains, similar to the structure of PRRs. WAK1 and WAK2 perceive oligogalacturonic acid, resulting from plant cell wall pectin degradation by fungal enzymes [45]. Plant lectins are able to recognize carbohydrates that originate directly from pathogens or from damage incurred during infection [47]. Many PAMPs and DAMPs contain carbohydrates (i.e. lipopolysaccharides, peptidoglycans, oligogalacturonides, and cellulose) and are recognized by PRRs/WAKs with lectin domains, such as lectin receptor-kinases [47]. Plants detect many extracellular molecules that indicate pathogen infection [48], such as extracellular DNA, ATP, and NAD(P). Pathogens have evolved to interfere in the detection of PAMPs and reduce the efficacy of PTI. *Cladosporium fulvum* and *Magnaporthe oryzae* produce chitin-binding proteins in

order to prevent plant perception (i.e. Avr4 and Slp1, respectively) [49-51]. Pathogens also produce effectors to thwart many aspects of plant immunity, which plants have developed ways to overcome, as outlined in the zig-zag model [10]. In order to recognize these infection-facilitating pathogen effectors, plants utilize another, more varied class of proteins.

2.2. NLRs detect pathogen effectors

NLRs are R genes among the fastest evolving gene families. Their products, upon detection of pathogenic effectors, undergo a conformational shift from a condensed, ADP-bound state to an open ATP-bound state with exposed N-terminal domains for the initiation of downstream signaling [52-54]. N-terminal Toll/interleukin-1 receptor-like (TIR) or coiled-coil (CC) domains precede an evolutionarily conserved NB-ARC domain and a highly variable LRR [52]. CNL genes are found in both monocots and dicots but TNL genes are restricted only to the latter [55]. Similar to PRRs, variability in the LRR provides these receptors with the ability to recognize various effector structures. Unlike PRRs, NLRs are generally located in the cytoplasm and possess LRRs at the C-terminal end. The NB-ARC is named for a Nucleotide-Binding domain with homology to Apoptotic protease activating factor 1, plant R-proteins, and *Caenorhabditis elegans* death-4, and contains many conserved motifs: P-loop/Walker-A, RNBS-A, Kinase-2/Walker-B, RNBS-B, RNBS-C, GLPL, RNBS-D, and MHDV. However, not all motifs are required for function, exemplified by the rice Pb1 CNL protein, which lacks a P-loop [56]. Wu et al. (2016 and 2017) have shown that NLRs operate in networks, differentiating sensor and

helper NLRs, along with NLRs required for cell death (NRCs) [57,58]. NLRs have emerged as complex receptors that can detect a variety of changes, both through non-self and modified-self recognition [59]. NLR interactions are not always advantageous, as mismatching of NLRs in hybrids can result in autoimmunity [60]. NB-ARCs similar to those found in NLRs are found in many different species, from the bacterial *Streptomyces coelicolor* genome [61] to those of nematode [62] and human [63], the latter two also involved in programmed cell death. Many NLR genes are located in extra-pericentromeric clusters [64], which experience high rates of chromosomal recombination. These genes evolve quickly through duplications, chromosomal rearrangements, and unequal crossing over [65]. Transposable elements also play a role in the evolution of regulatory sequences, like promoters [66,67]. Translocation of NLR genes to unlinked loci increases the likelihood of functional diversification [57]. Similar to PRRs, many effector-NLR interactions have been elucidated in *Arabidopsis* [68]. NLRs have been shown to become activated by direct interaction with pathogen effectors [69] or other ways by detecting: modification of the effector's target protein [21,70], modification of a target-mimicking decoy protein [71], or modification of the NLR itself [72-74]. One of the most well-studied NLR-effector interactions involves the NLRs RPM1 and RPS2, which perceive the targeting of resistance negative regulator RIN4 by *P. syringae* effectors AvrRpt2 and AvrRpm1 [75]. Unlike other NLRs, RPS2 and RPM1 are located within the plasma membrane, since their guardee, RIN4, is also associated with the plasma membrane [76]. After activation, RPM1 associates with the promoter-binding AtTIP49a negative regulator to up-regulate resistance responses [77]. Targets guarded by resistance proteins appear to be conserved

among plant species, as found in analysis of PBS1 homologs in grass species (see **Appendix I**).

NLRs have diverse structures that may include integrated decoy domains that exist as targets for effectors, such as TF sequences that effectors regularly target. *Arabidopsis* Resistance to Ralstonia Solanacearum 1 (RRS1) has a WRKY TF domain at the C-terminal end to directly bind to promoter sequences [78]. The *Ralstonia solanacearum* effector PopP2 modifies this WRKY domain, which triggers activation of the NLR [79,80]. NLRs are also able to dimerize, such as CNL proteins RGA5 and RGA4 in rice, in which RGA5 directly binds to *Magnaporthe oryzae* effectors Avr-Pia and Avr1-Co39 [81,82]. Much research is still necessary to elucidate how NLRs self-associate or dimerize, specifically genes with economic importance, such as wheat stem rust resistance genes [83]. Similar to PRR reliance on BAK1, NLRs rely on other proteins to transmit signals. CNLs and TNLs associate with Non-race-specific Disease Resistance 1 (NDR1) and Enhanced Disease Susceptibility 1 (EDS1) proteins, respectively [84-86]. NLRs are also able to localize to specific areas of the cell, such as endosomes or the nucleus. Potato CNL protein R3a, when triggered by *Phytopthora infestans* effector $Avr3a^{KI}$, moves to endosomes, where it recruits additional effectors [87]. Barley CNL MLA proteins accumulate in the nucleus to interfere with WRKY TFs to down-regulate immunity [88]. Plants also use exocytosis to move immune receptors to the plasma membrane and secrete antimicrobial substances. As a way to benefit from disruption of antimicrobial compound production, some pathogen effectors interfere with protease secretion [89], vesicular trafficking via proteasome degradation [90], and endocytosis [91]. *Parastagonospora nodorum* (*Septoria nodorum*) and *Pyrenophora tritici-repentis*

effectors SnTox1 and PtrToxA utilize susceptibility genes *Snn1* and *Tsn1*, which encode WAK and NLR proteins, respectively [92-96]. This hijacking of immune components allows necrotrophic pathogens to triggering HR through ROS accumulation [97]. PRRs, WAKs, and NLRs rely on complex signaling mechanisms to initiate defense responses. MAPKs, hormones, TFs, and other components play major rolls in this signal transduction.

3. SIGNAL TRANSDUCTION

3.1. Resistance Involves Multiple Signaling Mechanisms

Receptors activate signaling mechanisms that are common to many cellular processes, including MAPKs, G-proteins, ubiquitin, and calcium fluctuations. In the general model of MAPK signaling, membrane-bound Ras proteins facilitate the conversion of GTP to GDP, phosphorylating MAPKKK (Raf) proteins, which then phosphorylate MAPKK (MEK) proteins, leading to the phosphorylation of MAPK (ERK) proteins [14]. The involvement of MAPK in many cellular processes has led to the identification of MAPK genes in *Arabidopsis*, which contains 60 MAPKKKs, 10 MAPKKs, and 20 MAPKs [98]. Initiated by bacterial flagellin and elongation factor interaction, PRRs FLS2 and EFR dimerize with BAK1 and trigger MAPK signaling [99,100]. Pathogen pectin degradation detected by WAK1 and WAK2 also initiates a MAPK cascade [45,101,102]. Studies in tomato show MAPK genes involved in signal transduction of NLR perception as well [103-105]. Defense responses can also be down-regulated by MAPK signaling [106] and pathogens have developed effectors that interfere with MAPK signaling to suppress resistance responses [107]. Similarly, the heterotrimeric G-protein and G-

protein-coupled receptor (GPCR) system has been heavily studied due to its involvement in numerous cellular processes. Extracellular ligands bind to the transmembrane GPCR, causing the exchange of GDP for GTP in the α subunit of the G-protein complex, allowing a dissociation of the α subunit from the β-γ subunit complex, initiating further signaling [108]. Hydrolysis of GTP by the α subunit then causes the subunits to re-associate. Metazoan systems make more use of G-protein signaling [109-111], but G-proteins possess roles in HR and stomatal closure [15]. Ubiquitination and subsequent protein degradation by the proteasome also has activity in many signaling systems, including defense. Components are regulated positively by the repression of their degradation or negatively by targeted degradation [17]. Pathogens have evolved effectors to interfere with the ubiquitin proteasome system in an attempt to disrupt this signaling and facilitate infection [17]. Small ubiquitin-like modifiers (SUMOs) are also utilized by plants to regulate response and pathogens disrupt this signaling as well [112].

Receptors triggering fluctuations in calcium ions (Ca^{2+}) acts as signaling mechanisms to trigger responses to symbiotic or pathogenic microbes [16,113]. Calmodulin (CaM), calcium-dependent protein kinases (CDPKs), and calcineurin B-like proteins detect calcium to activate diverse families of TFs, including Calmodulin-Binding Transcription Activators (CAMTAs) [113,114]. CaM is involved in ROS production through MAPK cascade initiation [115]. Calcium signaling controls hormone activation and the expression of NDR1 and EDS1 proteins [113]. CDPKs move to the nucleus to phosphorylate WRKY TFs involved in RPS2 and RPM1 ETI [116]. This molecular signal can be transmitted through hormones that have roles in many different stress and developmental responses [14]. Similar to calcium signaling, fluctuations in hormones drive

differential expression of defense response genes. While sustained MAPK activity during ETI allows for less reliance on hormonal regulation, transient MAPK activity during PTI depends heavily upon hormonal signaling [117].

3.2. Plant Hormones Systematically Initiate and Repress Resistance

Hormones operating downstream of pathogen detection provide another layer of regulation and take many forms: salicylic acid (SA), jasmonic acid (JA), ethylene (ET), abscisic acid (ABA), nitric oxide (NO), cytokinins (CK), gibberellin (GA), auxin, and brassinosteroids (BR). Along with affecting a multitude of developmental and response functions, including cross-talk with other hormones, SA plays a central role in local and systemic resistance responses to biotrophic and hemi-biotrophic pathogens [118]. SA and MAPK cascades can act upstream of each other, with some cascades triggering SA activity or SA triggering MAPK cascades [118]. Signaling is transferred from receptors to SA through NDR1 for CNL receptors and a combination of EDS1 and Phytoalexin Deficient 4 (PAD4) for TNL receptors [86,119,120]. Through a complex web of interactions, SA transfers the signal that a pathogen is present through the action of TFs to induce the expression of defense-related genes. After initiated by signaling, SA leads to the reduction of disulfide bonds in the oligomer protein Nonexpressor of Pathogen Resistance Gene 1 (NPR1) by thioredoxins, allowing its constituent monomers to pass from the cytosol into the nucleus, bind to the TF TGA (binding site: TGACG), and up-regulate genes associated with resistance [18,121]. Taking advantage of this system, the *Cochliobolus victoriae* pathogenic effector victorin targets the thioredoxin TRX-h5,

involved in the monomerization of NPR1, and triggers cell death through the action of *Arabidopsis* susceptibility gene LOV1 [122].

JA and ET play key roles in the plant's response to necrotrophic pathogens [18,123-125] and herbivorous insects [126,127]. JA and ET up-regulate the emission of volatile compounds in response to caterpillar herbivory, initiated by caterpillar oral secretions [128]. In order to subvert plant response, insect gut microbes can reduce the JA-mediated defense in plants [129]. Perception of bacterial flagellin enhances the production of ET as a signaling mechanism [130]. Without ET present, TF Ethylene Insensitive 3 (EIN3) is degraded by F-box protein-mediated ubiquitination and proteasome activity. ET inactivates receptors and the Constitutive Triple Response1 (CTR1) protein, which stops the repression of EIN2 and EIN3 and allows up-regulation of ET signaling, expression of defense genes, and necrotroph resistance [18,131]. This is also a target of pathogen interference, as the XopD effector of tomato pathogen *Xanthomonas euvesicatoria* desumoylates the TF SIERF4 to interfere with hormone signaling, specifically suppressing ET production and resistance [132]. Functioning of the hormones ABA, NO, auxin, CK, GA, and BR in immunity and development shows that defense and growth are closely linked [133], often inversely related. ABA is involved in various plant stresses [134], including the repression and promotion of resistance responses during presence and absence of abiotic stress, respectively [135]. Closure of stomata involves ABA signaling to regulate water loss, gas exchange, and pathogen access to tissue [136,137]. GA and ascorbic acid (AA) deficiencies lead to enhanced resistance [138,139]. Since hormones are broadly defined as systemic regulators, peptides can also function as plant hormones. Recently, the small peptide hormone systemin has been shown to be involved in system herbivory

response, leading to changes in gene expression, specifically of neighboring plants that are not exposed to the biotic stress [140]. This indicates that plant hormones can trigger communication between individual plants to increase defense responses. Having discussed some of the major signaling factors, the proteins that directly alter transcription in the nucleus will be reviewed.

3.3. TFs Play an Essential Role in Transcriptional Reprogramming

Transcriptional reprogramming through TF activity plays roles at several levels of resistance: 1) basal expression of resistance components (e.g. receptors, kinases, response suppression proteins), 2) direct TF activity of activated receptor proteins, and 3) TFs activated downstream of receptor initiation (i.e. MAPK cascade leading to TF activation via phosphorylation [141]). Plant TFs are diverse in comparison to Metazoan systems: homeodomain, MADS box, C_2H_2 zinc finger, AP2/ERF, bHLH, TGA/bZIP, MYB, NAC, and WRKY; the latter six families are especially involved in defense [131]. AP2/ERF defense-related TFs (binding to GCC boxes: GCCGCC) are associated with ethylene regulation and involved in positive regulation of rice resistance to *Chilo suppressalis* [142] and *Arabidopsis* resistance to *B. cinerea* [143,144]. In the latter example, Ethylene Response Factor 6 (ERF6) activates Plant Defensin 1.1 and 1.2 after MPK3/MPK6 phosphorylation [145]. An MYC2 TF, NaMYC2 regulates the production of nicotine, an anti-herbivory secondary metabolite from *Nicotiana attenuata* [146]. The chitin-triggered rice bZIP TF OsTGAP1 up-regulates antimicrobial compound synthesis [147]. Abiotic stresses can also impact resistance signaling, like cold, drought, or wounding [148,149]. bHLH TF MYC2

works antagonistic to Ethylene Response Factor 1 (ERF1) by down-regulating resistance and up-regulating wounding responses [150]. Aside from kinase phosphorylation, TFs can also be activated by modifications, such as proteolysis in NAC TF NTL6 [148]. TFs can also target specific aspects of infection, like barley HvNAC6 up-regulating genes involved in resistance to *Blumeria graminis* penetration [151]. In addition to these classes of TFs, WRKYs appear to be involved in many of the thoroughly studied resistance responses.

WRKY TFs bind to W-box regions (TTGACC/T) and possess conserved WRKYGQK amino acid residues along with zinc finger domains [152]. In addition to their involvement in abiotic stresses, like drought and salt tolerance, WRKY TFs positively and negatively regulate the expression of genes associated with defense, including response to viruses [153]. WRKY TFs are classifying into groups I, IIa-e, and III, with analysis of ancestral species (i.e. algae) leading to potential phylogenetic relationships [152]. *Arabidopsis* resistance to *B. cinerea* involves phosphorylation of WRKY33 by MPK3/MPK6, followed by WRKY33 binding to its own promoter and the promoters of components necessary for the synthesis of ethylene [154] and antimicrobial compounds (via Phytoalexin Deficient 3) [155]. In contrast, barley HvWRKY1 and HvWRKY2 repress PTI, disrupted by CNL protein Mildew Locus A10 (MLA10) after detection of the *Blumeria graminis* effector AVR_{A10} [88]. Barley CNL protein MLA1 interacts with WRKY1 and MYB6 TFs, the latter up-regulating *B. graminis* resistance [156]. Rice WRKY45 positively regulates blast resistance when activated by the CNL protein Pb1 [157]. Receptor TF activity is shown in resistance to *Xanthomonas oryzae*, where rice PRR/RLK Xa21 is cleaved and sends a kinase domain to the nucleus that binds to the negative regulator

OsWRKY62 [34,158]. The allelic variants OsWRKY45-1 and OsWRKY45-2 both positively regulate resistance to *Magnaporthe grisea*, but down-regulate and up-regulate resistance to *Xanthomonas oryzae*, respectively [159]. WRKY TFs also interact with other families of TFs, such as bHLH TFs AtBZR1 and AtBES1/BZR2, working in conjunction with a WRKY TFs to regulate BR signaling [160]. Wheat TFs TaWRKY49 and TaWRKY62 have been demonstrated to play roles in resistance to *P. striiformis*, affecting the expression of genes associated with SA, JA, ET, and ROS [161].

Pathogens have evolved effectors that interfere with transcriptional reprogramming by blocking plant TF activity or directly promoting plant gene expression. *Phytophthora infestans* effectors prevent potato NAC TFs from moving from the endoplasmic reticulum to the nucleus [162]. *Pseudomonas syringae* effector HopD1 interacts with the membrane-bound TF NTL9, suppressing effector-triggered immunity [163] and Tobacco Mosaic Virus (TMV) also appears to interfere with the NAC TF ATAF2 to suppress resistance [164]. Pathogens also produce effectors that act as plant TFs [28]. *Xanthomonas* and *Rastonia* bacteria produce transcriptional activator-like effectors (TALEs) that bind to plant susceptibility gene promoters [23,165]. Rice *Xa10* and pepper *Bs3* genes possess TALE-binding promoter sites for TALEs AvrXa10 and AvrBs3 involved in resistance to *Xanthomonas oryzae* pv. *oryzae* and *Xanthomonas campestris* pv. *vesicatoria*, respectively [166,167]. While TFs directly trigger transcriptional reprogramming, an additional layer of regulation exists for response-related genes.

3.4. Epigenetic Regulation and PDR Provide Additional Regulatory Mechanisms

Nucleic acids have roles as regulators of plant immunity and are critical targets of plant and pathogen degradation. Epigenetic changes can increase pathogen resistance through up-regulation of response proteins or down-regulation of response inhibitors. Pathogen exposure causes chromatin modification that affects expression of various plant defense response components [168]. DNA methylation, histone methylation/acetylation, RNA interference (RNAi), and recombination between homologous chromosomes influence plant defense [169]. Methylation of genes encoding resistance components has been shown to reduce pathogen resistance [66,170,171]. SA repression activity is facilitated by histone deacetylase HDA19 [172], which interacts with WRKY TFs in response to *P. syringae* [173]. SA analog treatment and *P. syringae* infection of *Arabidopsis* both result in histone acetylation and methylation of gene promoters [168,174]. Epigenetic regulation intensifies the host's reaction to likely threats and ensures that host resources are not wasted on resistance to unlikely threats. The physiological cost of defense and longevity of pathogen inoculum both give a selective advantage to plants that can repress and prime responses. Epigenetic factors serve as another layer of regulation of resistance responses by the plant.

In contrast to the ligand-receptor resistance mechanisms discussed earlier, pathogen-derived resistance (PDR) involves the use of pathogen components to confer resistance. Inoculation of a plant with a less virulent form of the pathogen can cause the plant to become more resistant to later infection by a more virulent race/pathotype [175]. Barley primed with chemical elicitors produced offspring with enhanced resistance to the fungi *Rhynchosporium commune* [176]. Interestingly, priming to increase resistance of

Fusarium graminearum by wheat caused an increase in *F. graminearum* mycotoxin production [177]. PDR to viruses often involves expression of viral coat proteins, replicases, and interfering RNAs [178,179]. Plants expressing the coat protein of a virus resist other viral strains [180] due to interference in viral disassembly that is necessary for viral replication. Viral sequences transcribed by the plant are used to generate microRNAs that interfere with viral transcription, which is necessary for viral replication within the host cell. RNAi is used as an anti-viral mechanism that degrades pathogenic RNA. Some pathogens possess effectors that suppress RNA silencing [181-183] and RNAi is the underlying mechanism in the development of transgenic wheat with resistance to the wheat streak mosaic virus [184]. RNAi has also been proposed for development of nematode resistance in soybean [185]. Plants regulate defense components using microRNAs, alternative splicing, and alternative polyadenylation [186-194]. Pathogens have evolved the ability to use small RNAs as effectors that move into plant cells and repress host defense machinery [195]. In order to facilitate infection, *Botrytis cinerea* uses RNA silencing of the host [196] and *Phytopthora sojae* suppresses RNA silencing [182,197]. Ribonucleases (RNases) in the apoplast have also been correlated with increased resistance to RNA viruses [198] and anti-fungal activity [199-201]. The wheat Wheatwin1 protein possesses both RNase and antifungal activity [202]. Corresponding to the diverse receptors and messengers that signal the presence of pests, plants possess a sophisticated array of defense tactics that restrict pathogens and pests from further infection, growth, herbivory, and reproduction.

4. DEFENSE RESPONSE

4.1. HR, ROS, and Cell Wall Modification Inhibit Pathogen Infection

HR is one of the most commonly used immune responses, causing planned cell death in the area surrounding an infection. This establishes a quarantine zone to stop the pathogen from spreading, an effective technique for pathogens that require living tissue (biotrophs). Pathogen infection triggers the production of peroxidases in order to generate ROS, which are used in multiple aspects of the resistance response [203,204]. NADPH oxidases are necessary for the production of superoxide, which peroxidases use to generate hydrogen peroxide (H_2O_2). One NADPH oxidase, RBOHD, associates with the PRRs EFR and FLS2, and is phosphorylated by BIK1, triggering ROS production [205]. ROS trigger programmed cell death and hydrogen peroxide moves to surrounding cells to initiate the production of compounds that prevent oxidative damage [206]. Transgenic plants that lack the ability to detoxify the ROS compounds were found to have more intense responses to pathogens that trigger HR [207]. Along with assisting in HR, ROS are used to create environments unsuitable for pathogen survival and reproduction, described as an oxidative burst [208]. Therefore, ROS are directly involved at levels of signal transduction and defense response, inhibiting fungal spore germination [208]. In addition to peroxidases and NADPH oxidase, other enzymes produce ROS, including amine and oxalate oxidases [209]. *Sclerotinia sclerotiorum*, a necrotrophic fungus with a broad host range that includes many crops, produces oxalic acid, which suppresses plant oxidative burst [210] in the early stages of infection, but increases ROS production after establishment [211]. Wheat and barley produce oxalate oxidase proteins, also known as germins [212,213], which break down oxalic acid, increasing their pathogen resistance [214]. Transgenic crops with wheat or

barley oxalate oxidase genes showed increased resistance to *S. sclerotiorum* [215-217] and other pests [218-221].

ROS mediate glycoprotein cross-linking, which strengthens cell walls [208,222]. Since fungi, bacteria, and nematodes need to penetrate the cell wall, this restricts pathogen movement and limits access to the nutrients necessary to complete reproduction. Bacterial pathogens, lacking some of the degradation enzymes that many fungi possess, use wounds and stomata in order to gain access to plant nutrients. Stomatal guard cells, recognizing bacterial PAMPs (i.e. flg22) and lipopolysaccharides, induce stomatal closure via SA and ABA signaling to prevent entry [137,223]. In response to this, *P. syringae* produces coronatine to initiate reopening of closed stomata [223] by interfering with hormone biosynthesis and mimicking phytohormones [28]. Once viewed as a passive way of pathogen entry into plant tissue, regulation of stomata has been demonstrated to include a complex defense mechanism, in addition to its response to abiotic stress. In addition to oxidase production, plants and pathogens generate many other factors that interfere with each other's carbohydrates, proteins, and lipids.

4.2. Enzymes and Enzyme Inhibitors Counter Pathogenic Effectors and Facilitate Detection

Fungi use enzymes like cellulases to degrade plant cell walls. Upon detection of these fungal proteins, plants respond by producing enzyme inhibitors and depositing callose and lignin to strengthen the cell wall [224]. In addition to cellulases, pathogens like *Fusarium graminearum* and *Fusarium culmorum* degrade plant carbohydrates with

pectinases and xylanases [225,226]. In a form of evolutionary retaliation, plants have evolved enzymes that degrade pathogen carbohydrates, including chitinases and β-1-3-glucanases [227,228]. This degradation of fungal cellular components not only inhibits microbial growth, but also makes PAMPs available to plant PRRs, thwarting the pest's attempt at evasion. Wheat chitinases, degrading a major component of fungal cell walls, inhibit fungal spore germination [229,230]. Recombinant wheat chitinases have been shown to possess antifungal activity against many different fungal species, not limited to wheat pathogenic fungi [231]. The chitin-binding Avr4 of *C. fulvum* protects chitin from plant chitinases [50]. Hevamine possesses chitinase and lysozyme activity and contains β barrel structural domains [232], otherwise associated with a multitude of functions generally associated with cellular metabolism [233]. Similar to chitinases, plant β-1,3-glucanases hydrolyze β-1,3-glucan in fungal cell walls, producing monomers that further stimulate plant defense responses [234,235]. This multifaceted approach aims at reducing the effectiveness of pathogenic components as well as strengthening plant defenses. Since the components of cell walls are diverse (i.e. cellulose, hemicellulose, pectin, lignin, etc.), pathogens must have a diverse set of proteins to infect the host, leading to an even more complex arrangement of plant receptors and defense proteins. Thus, the evolutionary battle over the ability to penetrate or reinforce the cell wall is a microcosm for the overall evolutionary coevolution of the plant-pathogen interaction.

Proteases released by both plants and pathogens evolved to reduce the efficacy of the catalytic proteins used by both groups (i.e. plant chitinase and fungal cellulase). Plants and pathogens also use protease inhibitors to impede the activity of these proteases [236]. Thomas and van der Hoorn (2018) discuss ten important types of proteases, grouping

them based on location: apoplastic, cytonuclear, vacuolar, endomembrane [237], showing how proteases act at multiple levels in plant-pathogen interaction. The *Fusarium verticillioides* Zn-metalloproteinase fungalysin cleaves defense chitinases [238], but wheat produces hevein-like antimicrobial proteins that inhibit fungalysin by binding to the enzyme without being cleaved [239], stopping the degradation of chitinases that are necessary to prevent infection. Barley protease inhibitors block the activity of *Fusarium* trypsin, chymotripsin, and α-amylase [240]. Wheat and barley α-amylase inhibitors interfere with pest starch digestion by interfering with α-amylase, used by insects and fungi to metabolize starches [241-245]. Both proteinaceous and non-proteinaceous α-amylase inhibitors are produced, the latter being organic compounds that mimic α-amylase substrates [245]. *Nicotiana attenuata* produces trypsin proteinase inhibitors along with nicotine to defend against *Spodoptera exigua* [246]. Plants may also detect pathogens through protease inhibitors. For example, *Cladosporium fulvum*, is percieved by tomato RLP Cf-2 detection of Avr2, which is initially produced to inhibit tomato proteases Rcr3 and PiP1 [247,248]. Oomycete *Phytophtora infestans* and nematode *Globodera rostochiensis* interfere with tomato protease activity [249,250], exemplifying a multilayered system of protein-protein interactions.

Like proteins, lipids participate in many cellular activities. Lipids form major barriers that separate a host plant from prospective pathogens. Pathogens initiate infection after perception of cutin or other compounds of the waxy cuticle. *Puccinia graminis* and *Blumeria graminis* initiate appressoria formation upon contact with surface wax of wheat [251] and barley [252], respectively. Cutinases are then utilized by fungi to hydrolyze cutin into cutin monomers to move through the cuticle [253]. Plant lipases inhibit fungal

infection, found in *Arabidopsis* [254] and wheat [255]. Lipid-transfer proteins also possess antimicrobial function linked to increased pathogen membrane permeability. However, details regarding how these proteins affect microbes remain to be elucidated [256]. Effectors that have activity within host cells, such as those of *P. syringae*, require lipid modification to move into the host cell [257], with NLRs receptors that detect these effectors located in the plasma membrane instead of the cytoplasm [258,259]. Lipids can also be direct targets of pathogens or plant perception such as the *Fusarium* toxin fumonisin interference with sphingolipid metabolism [260], or defense responses triggered by bacterial lipopolysaccharide PAMPs [261,262]. Phospholipases and lipoxygenases (LOXs) are involved in the break down of phospholipids/galactolipids into free fatty acids for the production of defense components [263] and can act in stomatal closure [264]. Phospholipase activity, which is involved with various hormone and stress responses [265], may generate products that are directly involved in defense response, such as phosphatidic acid [263,266]. In addition to this general enzyme activity related to cellular components, plants have evolved specialized proteins that defend them from pathogen infection.

4.3. Defensins and Thaumatin-like Proteins Offensively Inhibit Pathogen Infection

Defensins are a diverse class of small plant proteins that directly attack or inhibit invading microbes and parasitic plants [267]. Initially reported as barley [268] and wheat [269] γ-thionins, it was shown that they possess structural similarity to animal defensins [270,271]. Analogous to the action of many medical antibiotics, plant defensins interfere with pathogen protein synthesis and enzyme function. Barley defensins γ-hordothionin and ω-

hordothionin inhibit protein translation [268,272] and defensins can move into pathogen cytoplasm [273,274]. Wheat defensin TAD is expressed in the crown and possesses antifungal and antibacterial properties [275]. Defensins inhibit proteases [276], trigger pathogen ROS production [277], and block ion signaling [278]. Unlike animal defensins that inhibit bacterial growth, many plant defensins are antifungal and are especially active in seeds. They make up 0.5% of the total seed protein and a substantial amount of proteins that are released from the seed coats, at 30% [279]. Defensins have been found in many tissues [280] and may be induced during seasonal changes [275]. C-terminal hydrophobic and γ-core regions are critical for membrane interaction and antifungal activity, respectively [281,282]. Defensins contain scorpion toxin-like, knottin, and purothionin domains, with conserved cysteine residues that form a cysteine-knot structure composed of disulfide bridges. Scorpion toxins and some plant defensins both block potassium channels using similar protein domains [283,284]. Defensins cause an increase in pathogen membrane permeability that initiates necrosis [285]. Previous studies identified over 300 cysteine-rich defensin-like proteins in *Arabidopsis* [286] and *Medicago truncatula* [287]. Van Der Weerden and Anderson (2013) have proposed the classification of defensins into 18 groups based on species, structure, and function [288]. Plant defensins have potential as medical antibiotics, anti-tumor medication [289], and artificial sweeteners [290].

Thaumatin-like proteins, named after the protein thaumatin from *Thaumatococcus daniellii* [291], are also pathogenesis-related proteins. Barley thaumatin-like proteins bind to 1,3-β-D-glucans [292], associated with resistance to powdery mildew [293], *F. graminearum* [294], or general antifungal activity [295]. The antifungal thaumatin-like proteins osmotin (tobacco), zeamatin (maize), hordomatin (barley), avematin (oat), and trimatin (wheat)

are permatins that form transmembrane pores in fungal membranes [295-298]. Like defensins, permatins accumulate in seeds [299]. Research in barley and wheat shows thaumatin-like proteins PRHv-1 and PWIR2, respectively, expressed along with other genes during fungal resistance responses [300,301], with similar sequences in oat [302]. Thaumatin-like proteins have potential as artificial sweeteners and influence the malting quality of barley [303]. Thaumatin-like proteins make up one of the 17 categories of pathogenesis-related (PR) proteins: oxidases and oxidase-like (PR-9, 15, and 16), chitinases (PR-3, 4, 8, and 11), β-1,3-glucanases (PR-2), endoproteinases (PR-7), proteinase inhibitors (PR-6), lipid-transfer proteins (PR-14), ribonuclease-like (PR-10), defensins and thionins (PR-12 and 13, respectively), thaumatin-like (PR-5), and the less understood antifungal (Pr-1) and functionally uncategorized (Pr-17) groups [20,304-306]. Additional mechanisms that plants use to deter pests involve non-protein compounds and assistance from other species.

4.4. Phytoalexins and Beneficial Symbionts are Chemical and Biological Plant Weapons

Phytoalexins are organic compounds produced in response to invading pests to interfere with metabolism, development, and reproduction. Phytoalexins were initially investigated as defense compounds that protected potatoes from *Phytophthora infestans*. Several classes of plant chemicals have pesticide activity and those that are constitutively produced are described as phytoanticipins. A model phytoalexin used by *Arabidopsis*, camalexin is produced in response to many different types of microbial pathogens and pests [307]. Camalexin is regulated by MAPK cascades [308] and WRKY TFs [155,309,310]. Some

adaptive pathogens are able to detoxify this chemical [311]. Wheat produces benzoxazinoids (BXs), such as 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA), which defends against microbes and insects, such as reducing transmission of barley yellow dwarf virus through aphid feeding [312-314]. Wheat pathogens *Gaeumannomyces graminis* and *Fusarium culmorum* can detoxify BXs [315]. Wheat likely inherits genes required to convert 2,4-dihydroxy-1,4-benzoxazin-3-one (DIBOA) to DIMBOA through the progenitor of its B genome (chromosome 4B) since *Aegilops speltoides* accumulates DIMBOA, but *Triticum urartu* and *Aegilops tauschii* do not [316]. Additional phytoalexins in cereals include avenanthramides in oat and diterpenoids in rice [310]. As a phytoanticipin of the saponin group, avenacin A-1 is produced by oat root epidermis and forms pores in fungal membranes by interacting with fungal membrane sterols [317-319]. *G. graminis* var. *avenae* has evolved the ability to detoxify avenacin A-1 [320]. While phytoalexins have been found in many species [310] with classified allelopathic effects [321], some signaling mechanisms leading to production remain elusive. Phytoalexins may also show usefulness in medical applications [310,322]. Exogenous chemicals applied to crops, such as glyphosate-based herbicides, may increase crop susceptibility to disease [323].

In addition to producing pesticides, plants can recruit natural predators of herbivorous insects as a defense through the production of herbivore-induced plant volatiles. Caterpillars feeding on maize leaves induce the production of terpenoid compounds and indole, attracting parasitic wasps that feed on the caterpillars [324-326]. JA signaling is also activated, driven by volicitin from caterpillar oral secretions [128]. To protect themselves from herbivory, plants can also produce sticky substances that trap insects, such as resin and latex [327-329], along with increasing photosynthetic production

[330]. Morphological features, like trichomes, reduce insect herbivory, in addition to many cellular components, such as flavonoids, tannins, terpenoids, alkaloids, and phenolics [331]. Aside from providing the plant with access to nutrients, some symbionts assist their host in pathogen defense. Wheat rhizobacteria have activity against the soilborne pathogen *G. graminis* through the production of antibiotic substances [332] and rice arbuscular mycorrhiza trigger improved immune responses to protect the host [333]. Mycorrhiza in corn have the ability to enhance production of DIMBOA [334]. Symbionts have been shown to affect resistance responses through repression of JA-mediate defense [335] or interference with ROS and β-1,3-glucanse production [336,337]. These interactions demonstrate the multi-layered nature of the plant immune system.

5. CONCLUSIONS AND FUTURE DIRECTIONS

Knowledge of plant-pathogen interactions will undoubtedly continue to flourish in the 21st century, driven by new molecular techniques and greater computational power. Phytopathology, like other fields, will continue to grow as more details emerge regarding plant-pathogen interactions. Research will be driven by several factors, such as disease pressures associated with modern agricultural practices and climate change, increasing the need for durable pathogen resistance in crops [338]. In addition to improving knowledge of plant immunity, efforts will continue to alter crop genetics to develop better resistance. Continuing to alter the receptors necessary to initiate defense responses is likely the best route for development of resistance. NLRs may become a major topic of biotechnology, allowing the engineering of resistance to any pathogen, similar to the modified activity of the CRISPR/Cas9 system. The mechanism utilized in one of these approaches is the activation of *Arabidopsis* NLR RPS5 by *P. syringae* protease AvrPphB cleavage of PBS1

[339,340]. Kim et al. (2016) showed that the cleavage site of PBS1 can be replaced with a cleavage site for other pathogen proteases, allowing for defense responses to be triggered by other pathogens [341]. While this technology is currently limited, future studies will likely engineer crops with novel R-genes that were not directly transferred from other species. In order to trigger the most effective defense response, engineering novel resistance pathways to different pathogens will also need to pair the receptors with the appropriate method of signal transduction. PDR may also have applications in genetic engineering, by allowing plants to express pathogen genes that promote resistance [342]. Future studies will also focus on understanding quantitative resistance and gene pyramiding due to the durable resistance it holds [343], such as the multi-decade resistance found in barley cultivar NDB112 [344-347]. Additional mechanisms of resistance regulation and response will be uncovered in future years, having applications to agricultural systems. Understanding pathogen resistance and plant immunity will greatly enhance not only the modes of reducing crop loss, but also will contribute to the overall understanding of the molecular interactions and coevolution that underlies this field and numerous applications to other biological systems.

ACKNOWLEDGEMENTS

Support for this dissertation work came from the USDA-NIFA hatch Projects to M. Nepal (SD00H469-13 and SD00H659-18), South Dakota Agriculture Experiment Station, and Department of Biology & Microbiology at South Dakota State University.

REFERENCES

1. Whetzel HH. *An outline of the history of phytopathology.* WB Saunders; 1918.
2. Moutinho-Pereira J, Magalhaes N, Torres de Castro L, Manuela Chaves M, Torres-Pereira J. Physiological responses of grapevine leaves to Bordeaux mixture under light. *Vitis.* 2001;40(3):117-121.
3. Kutschera U, Hossfeld U. Physiological phytopathology: origin and evolution of a scientific discipline. *Journal of Applied Botany and Food Quality.* 2012;85(1):1.
4. Rhoades VH. The location of a gene for disease resistance in maize. *Proceedings of the National Academy of Sciences.* 1935;21(5):243-246.
5. Bushnell W. *The cereal rusts: origins, specificity, structure, and physiology.* Vol 1: Elsevier; 2012.
6. Flor HH. Inheritance of pathogenicity in Melampsora lini. *Phytopathology.* 1942;32:653-669.
7. Hammond-Kosack KE, Jones J. Resistance gene-dependent plant defense responses. *The Plant Cell.* 1996;8(10):1773.
8. Hammond-Kosack KE, Jones JD. Plant disease resistance genes. *Annual review of plant biology.* 1997;48(1):575-607.
9. Antolín-Llovera M, Ried MK, Binder A, Parniske M. Receptor kinase signaling pathways in plant-microbe interactions. *Annual review of phytopathology.* 2012;50:451-473.
10. Jones JD, Dangl JL. The plant immune system. *Nature.* 2006;444(7117):323-329.
11. Zipfel C. Plant pattern-recognition receptors. *Trends in immunology.* 2014;35(7):345-351.
12. Decreux A, Messiaen J. Wall-associated kinase WAK1 interacts with cell wall pectins in a calcium-induced conformation. *Plant and Cell Physiology.* 2005;46(2):268-278.

13. Dangl JL, Horvath DM, Staskawicz BJ. Pivoting the plant immune system from dissection to deployment. *Science.* 2013;341(6147):746-751.

14. Meng X, Zhang S. MAPK cascades in plant disease resistance signaling. *Annual Review of Phytopathology.* 2013;51:245-266.

15. Zhang H, Gao Z, Zheng X, Zhang Z. The role of G-proteins in plant immunity. *Plant signaling & behavior.* 2012;7(10):1284-1288.

16. Lecourieux D, Ranjeva R, Pugin A. Calcium in plant defence‐signalling pathways. *New Phytologist.* 2006;171(2):249-269.

17. Marino D, Peeters N, Rivas S. Ubiquitination during plant immune signaling. *Plant physiology.* 2012;160(1):15-27.

18. Robert-Seilaniantz A, Grant M, Jones JD. Hormone crosstalk in plant disease and defense: more than just jasmonate-salicylate antagonism. *Annual review of phytopathology.* 2011;49:317-343.

19. Juge N. Plant protein inhibitors of cell wall degrading enzymes. *Trends in plant science.* 2006;11(7):359-367.

20. van Loon LC, Rep M, Pieterse CM. Significance of inducible defense-related proteins in infected plants. *Annu Rev Phytopathol.* 2006;44:135-162.

21. Dangl JL, Jones JD. Plant pathogens and integrated defence responses to infection. *nature.* 2001;411(6839):826-833.

22. Gururani MA, Venkatesh J, Upadhyaya CP, Nookaraju A, Pandey SK, Park SW. Plant disease resistance genes: current status and future directions. *Physiological and Molecular Plant Pathology.* 2012;78:51-65.

23. Boch J, Bonas U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annual review of phytopathology.* 2010;48.

24. Gardiner DM, Upadhyaya NM, Stiller J, et al. Genomic analysis of Xanthomonas translucens pathogenic on wheat and barley reveals cross-kingdom gene transfer events and diverse protein delivery systems. *PLoS One.* 2014;9(1):e84995.

25. Friesen TL, Stukenbrock EH, Liu Z, et al. Emergence of a new disease as a result of interspecific virulence gene transfer. *Nature genetics.* 2006;38(8):953-956.

26. Glazebrook J. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu Rev Phytopathol.* 2005;43:205-227.

27. Vleeshouwers VG, Oliver RP. Effectors as tools in disease resistance breeding against biotrophic, hemibiotrophic, and necrotrophic plant pathogens. *Molecular plant-microbe interactions.* 2014;27(3):196-206.

28. Toruño TY, Stergiopoulos I, Coaker G. Plant-pathogen effectors: cellular probes interfering with plant defenses in spatial and temporal manners. *Annual review of phytopathology.* 2016;54:419-441.

29. de Jonge R, van Esse HP, Kombrink A, et al. Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. *Science.* 2010;329(5994):953-955.

30. Sun Y, Li L, Macho AP, et al. Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex. *Science.* 2013;342(6158):624-628.

31. Niehl A, Wyrsch I, Boller T, Heinlein M. Double‐stranded RNAs induce a pattern‐triggered immune signaling pathway in plants. *New Phytologist.* 2016;211(3):1008-1019.

32. Monaghan J, Zipfel C. Plant pattern recognition receptor complexes at the plasma membrane. *Current opinion in plant biology.* 2012;15(4):349-357.

33. Prince DC, Drurey C, Zipfel C, Hogenhout SA. The leucine-rich repeat receptor-like kinase BRASSINOSTEROID INSENSITIVE1-ASSOCIATED KINASE1 and the cytochrome P450

PHYTOALEXIN DEFICIENT3 contribute to innate immunity to aphids in Arabidopsis. *Plant physiology.* 2014;164(4):2207-2219.

34. Park C-J, Ronald PC. Cleavage and nuclear localization of the rice XA21 immune receptor. *Nature communications.* 2012;3:920.

35. Zipfel C, Robatzek S, Navarro L, Oakeley EJ. Bacterial disease resistance in Arabidopsis through flagellin perception. *Nature.* 2004;428(6984):764.

36. Kunze G, Zipfel C, Robatzek S, Niehaus K, Boller T, Felix G. The N terminus of bacterial elongation factor Tu elicits innate immunity in Arabidopsis plants. *The Plant Cell.* 2004;16(12):3496-3507.

37. Gust AA, Biswas R, Lenz HD, et al. Bacteria-derived peptidoglycans constitute pathogen-associated molecular patterns triggering innate immunity in Arabidopsis. *Journal of Biological Chemistry.* 2007;282(44):32338-32348.

38. Kaku H, Nishizawa Y, Ishii-Minami N, et al. Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. *Proceedings of the National Academy of Sciences.* 2006;103(29):11086-11091.

39. Ron M, Avni A. The receptor for the fungal elicitor ethylene-inducing xylanase is a member of a resistance-like gene family in tomato. *The Plant Cell.* 2004;16(6):1604-1615.

40. Fliegmann J, Mithöfer A, Wanner G, Ebel J. An ancient enzyme domain hidden in the putative β-glucan elicitor receptor of soybean may play an active part in the perception of pathogen-associated molecular patterns during broad host resistance. *Journal of Biological Chemistry.* 2004;279(2):1132-1140.

41. Du J, Verzaux E, Chaparro-Garcia A, et al. Elicitin recognition confers enhanced resistance to Phytophthora infestans in potato. *Nature Plants.* 2015;1:15034.

42. Feuillet C, Schachermayr G, Keller B. Molecular cloning of a new receptor‐like kinase gene encoded at the Lr10 disease resistance locus of wheat. *The Plant Journal.* 1997;11(1):45-52.

43. Zhou H, Li S, Deng Z, et al. Molecular analysis of three new receptor‐like kinase genes from hexaploid wheat and evidence for their participation in the wheat hypersensitive response to stripe rust fungus infection. *The Plant Journal.* 2007;52(3):420-434.

44. Jiang Z, Ge S, Xing L, et al. RLP1. 1, a novel wheat receptor-like protein gene, is involved in the defence response against Puccinia striiformis f. sp. tritici. *Journal of experimental botany.* 2013;64(12):3735-3746.

45. Brutus A, Sicilia F, Macone A, Cervone F, De Lorenzo G. A domain swap approach reveals a role of the plant wall-associated kinase 1 (WAK1) as a receptor of oligogalacturonides. *Proceedings of the National Academy of Sciences.* 2010;107(20):9452-9457.

46. Choi J, Tanaka K, Cao Y, et al. Identification of a plant receptor for extracellular ATP. *Science.* 2014;343(6168):290-294.

47. Lannoo N, Van Damme EJ. Lectin domains at the frontiers of plant defense. 2014.

48. Gust AA, Pruitt R, Nürnberger T. Sensing Danger: Key to Activating Plant Immunity. *Trends in plant science.* 2017;22(9):779-791.

49. Lo Presti L, Lanver D, Schweizer G, et al. Fungal effectors and plant susceptibility. *Annual review of plant biology.* 2015;66:513-545.

50. van den Burg HA, Harrison SJ, Joosten MH, Vervoort J, de Wit PJ. Cladosporium fulvum Avr4 protects fungal cell walls against hydrolysis by plant chitinases accumulating during infection. *Molecular plant-microbe interactions.* 2006;19(12):1420-1430.

51. Mentlak TA, Kombrink A, Shinya T, et al. Effector-mediated suppression of chitin-triggered immunity by Magnaporthe oryzae is necessary for rice blast disease. *The Plant Cell.* 2012;24(1):322-335.

52. Takken FL, Goverse A. How to build a pathogen detector: structural basis of NB-LRR function. *Current opinion in plant biology.* 2012;15(4):375-384.

53. Michelmore RW, Christopoulou M, Caldwell KS. Impacts of resistance gene genetics, function, and evolution on a durable future. *Annual review of phytopathology.* 2013;51:291-319.

54. Cui H, Tsuda K, Parker JE. Effector-triggered immunity: from pathogen perception to robust defense. *Annual review of plant biology.* 2015;66:487-511.

55. Jacob F, Vernaldi S, Maekawa T. Evolution and conservation of plant NLR functions. *Frontiers in immunology.* 2013;4.

56. Hayashi N, Inoue H, Kato T, et al. Durable panicle blast‐resistance gene Pb1 encodes an atypical CC‐NBS‐LRR protein and was generated by acquiring a promoter through local genome duplication. *The Plant Journal.* 2010;64(3):498-510.

57. Wu C-H, Abd-El-Haliem A, Bozkurt TO, et al. NLR network mediates immunity to diverse plant pathogens. *Proceedings of the National Academy of Sciences.* 2017:201702041.

58. Wu CH, Belhaj K, Bozkurt TO, Birk MS, Kamoun S. Helper NLR proteins NRC2a/b and NRC3 but not NRC1 are required for Pto‐mediated cell death and resistance in Nicotiana benthamiana. *New Phytologist.* 2016;209(4):1344-1352.

59. Baggs E, Dagdas G, Krasileva K. NLR diversity, helpers and integrated domains: making sense of the NLR IDentity. *Current opinion in plant biology.* 2017;38:59-67.

60. Tran DT, Chung E-H, Habring-Müller A, et al. Activation of a plant NLR complex through heteromeric association with an autoimmune risk variant of another NLR. *Current Biology.* 2017;27(8):1148-1160.

61. Horinouchi S, Kito M, Nishiyama M, et al. Primary structure of AfsR, a global regulatory protein for secondary metabolite formation in Streptomyces coelicolor A3 (2). *Gene.* 1990;95(1):49-56.

62. Hengartner MO, Horvitz HR. Programmed cell death in Caenorhabditis elegans. *Current opinion in genetics & development.* 1994;4(4):581-586.

63. Zou H, Henzel WJ, Liu X, Lutschg A, Wang X. Apaf-1, a human protein homologous to C. elegans CED-4, participates in cytochrome c–dependent activation of caspase-3. *Cell.* 1997;90(3):405-413.

64. Andersen EJ, Ali S, Reese RN, Yen Y, Neupane S, Nepal MP. Diversity and Evolution of Disease Resistance Genes in Barley (Hordeum vulgare L.). *Evolutionary bioinformatics online.* 2016;12:99.

65. Marone D, Russo MA, Laidò G, De Leonardis AM, Mastrangelo AM. Plant nucleotide binding site–leucine-rich repeat (NBS-LRR) genes: active guardians in host defense responses. *International journal of molecular sciences.* 2013;14(4):7302-7326.

66. Le T-N, Schumann U, Smith NA, et al. DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in Arabidopsis. *Genome biology.* 2014;15(9):458.

67. Seidl MF, Thomma BP. Transposable Elements Direct The Coevolution between Plants and Microbes. *Trends in Genetics.* 2017.

68. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of NBS-LRR–encoding genes in Arabidopsis. *The Plant Cell Online.* 2003;15(4):809-834.

69. Dodds PN, Lawrence GJ, Catanzariti A-M, et al. Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proceedings of the National Academy of Sciences.* 2006;103(23):8888-8893.

70. Van Der Biezen EA, Jones JD. Plant disease-resistance proteins and the gene-for-gene concept. *Trends in biochemical sciences.* 1998;23(12):454-456.

71. van der Hoorn RA, Kamoun S. From guard to decoy: a new model for perception of plant pathogen effectors. *The Plant Cell.* 2008;20(8):2009-2017.

72. Cesari S, Bernoux M, Moncuquet P, Kroj T, Dodds PN. A novel conserved mechanism for plant NLR protein pairs: the "integrated decoy" hypothesis. *Frontiers in plant science.* 2014;5.

73. Wu C-H, Krasileva KV, Banfield MJ, Terauchi R, Kamoun S. The "sensor domains" of plant NLR proteins: more than decoys? *Frontiers in plant science.* 2015;6.

74. Kroj T, Chanclud E, Michel‐Romiti C, Grand X, Morel JB. Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread. *New Phytologist.* 2016;210(2):618-626.

75. Marathe R, Dinesh-Kumar S. Plant defense: one post, multiple guards?! *Mol Cell.* 2003;11(2):284-286.

76. Axtell MJ, Staskawicz BJ. Initiation of RPS2-specified disease resistance in Arabidopsis is coupled to the AvrRpt2-directed elimination of RIN4. *Cell.* 2003;112(3):369-377.

77. Holt BF, Boyes DC, Ellerström M, et al. An evolutionarily conserved mediator of plant disease resistance gene function is required for normal Arabidopsis development. *Developmental cell.* 2002;2(6):807-817.

78. Noutoshi Y, Ito T, Seki M, et al. A single amino acid insertion in the WRKY domain of the Arabidopsis TIR‐NBS‐LRR‐WRKY‐type disease resistance protein SLH1 (sensitive to

low humidity 1) causes activation of defense responses and hypersensitive cell death. *The Plant Journal.* 2005;43(6):873-888.

79. Sarris PF, Duxbury Z, Huh SU, et al. A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell.* 2015;161(5):1089-1100.

80. Le Roux C, Huet G, Jauneau A, et al. A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. *Cell.* 2015;161(5):1074-1088.

81. Cesari S, Thilliez G, Ribot C, et al. The rice resistance protein pair RGA4/RGA5 recognizes the Magnaporthe oryzae effectors AVR-Pia and AVR1-CO39 by direct binding. *The Plant Cell.* 2013;25(4):1463-1481.

82. Césari S, Kanzaki H, Fujiwara T, et al. The NB‐LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. *The EMBO journal.* 2014;33(17):1941-1959.

83. Casey LW, Lavrencic P, Bentham AR, et al. The CC domain structure from the wheat stem rust resistance protein Sr33 challenges paradigms for dimerization in plant NLR proteins. *Proceedings of the National Academy of Sciences.* 2016;113(45):12856-12861.

84. Day B, Dahlbeck D, Staskawicz BJ. NDR1 interaction with RIN4 mediates the differential activation of multiple disease resistance pathways in Arabidopsis. *The Plant Cell.* 2006;18(10):2782-2791.

85. Knepper C, Savory EA, Day B. The role of NDR1 in pathogen perception and plant defense signaling. *Plant signaling & behavior.* 2011;6(8):1114-1116.

86. Wiermer M, Feys BJ, Parker JE. Plant immunity: the EDS1 regulatory node. *Current opinion in plant biology.* 2005;8(4):383-389.

87. Engelhardt S, Boevink PC, Armstrong MR, Ramos MB, Hein I, Birch PR. Relocalization of late blight resistance protein R3a to endosomal compartments is associated with

effector recognition and required for the immune response. *The Plant Cell.* 2012;24(12):5142-5158.

88. Shen Q-H, Saijo Y, Mauch S, et al. Nuclear activity of MLA immune receptors links isolate-specific and basal disease-resistance responses. *science.* 2007;315(5815):1098-1103.

89. Bozkurt TO, Schornack S, Win J, et al. Phytophthora infestans effector AVRblb2 prevents secretion of a plant immune protease at the haustorial interface. *Proceedings of the National Academy of Sciences.* 2011;108(51):20832-20837.

90. Nomura K, DebRoy S, Lee YH, Pumplin N, Jones J, He SY. A bacterial virulence protein suppresses host innate immunity to cause plant disease. *Science.* 2006;313(5784):220-223.

91. Chaparro-Garcia A, Schwizer S, Sklenar J, et al. Phytophthora infestans RXLR-WY effector AVR3a associates with dynamin-related protein 2 required for endocytosis of the plant pattern recognition receptor FLS2. *PLoS One.* 2015;10(9):e0137071.

92. Shi G, Zhang Z, Friesen TL, et al. The hijacking of a receptor kinase–driven pathway by a wheat fungal pathogen leads to disease. *Science advances.* 2016;2(10):e1600822.

93. Liu Z, Gao Y, Kim YM, et al. SnTox1, a Parastagonospora nodorum necrotrophic effector, is a dual‐function protein that facilitates infection while protecting from wheat‐produced chitinases. *New Phytologist.* 2016;211(3):1052-1064.

94. Ciuffetti LM, Manning VA, Pandelova I, Betts MF, Martinez JP. Host‐selective toxins, Ptr ToxA and Ptr ToxB, as necrotrophic effectors in the Pyrenophora tritici‐repentis‐wheat interaction. *New Phytologist.* 2010;187(4):911-919.

95. Liu Z, Zurn JD, Kariyawasam G, et al. Inverse gene-for-gene interactions contribute additively to tan spot susceptibility in wheat. *Theoretical and Applied Genetics.* 2017;130(6):1267-1276.

96. Faris JD, Zhang Z, Lu H, et al. A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. *Proceedings of the National Academy of Sciences.* 2010;107(30):13544-13549.

97. Manning VA, Chu AL, Steeves JE, Wolpert TJ, Ciuffetti LM. A host-selective toxin of Pyrenophora tritici-repentis, Ptr ToxA, induces photosystem changes and reactive oxygen species accumulation in sensitive wheat. *Molecular Plant-Microbe Interactions.* 2009;22(6):665-676.

98. Ichimura K, Shinozaki K, Tena G, et al. Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends in plant science.* 2002;7(7):301-308.

99. Zipfel C, Kunze G, Chinchilla D, et al. Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation. *Cell.* 2006;125(4):749-760.

100. Chinchilla D, Zipfel C, Robatzek S, et al. A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature.* 2007;448(7152):497.

101. Galletti R, Ferrari S, De Lorenzo G. Arabidopsis MPK3 and MPK6 play different roles in basal and oligogalacturonide-or flagellin-induced resistance against Botrytis cinerea. *Plant Physiology.* 2011;157(2):804-814.

102. Kohorn BD, Kohorn SL, Todorova T, Baptiste G, Stansky K, McCullough M. A dominant allele of Arabidopsis pectin-binding wall-associated kinase induces a stress response suppressed by MPK6 but not MPK3 mutations. *Molecular plant.* 2012;5(4):841-851.

103. Melech‐Bonfil S, Sessa G. Tomato MAPKKK ε is a positive regulator of cell‐death signaling networks associated with plant immunity. *The Plant Journal.* 2010;64(3):379-391.

104. del Pozo O, Pedley KF, Martin GB. MAPKKKα is a positive regulator of cell death associated with both plant immunity and disease. *The EMBO Journal.* 2004;23(15):3072-3082.

105. Oh C-S, Martin GB. Effector-triggered immunity mediated by the Pto kinase. *Trends in plant science.* 2011;16(3):132-140.

106. Frye CA, Tang D, Innes RW. Negative regulation of defense responses in plants by a conserved MAPKK kinase. *Proceedings of the National Academy of Sciences.* 2001;98(1):373-378.

107. Zhang J, Shao F, Li Y, et al. A Pseudomonas syringae effector inactivates MAPKs to suppress PAMP-induced immunity in plants. *Cell host & microbe.* 2007;1(3):175-185.

108. Temple BR, Jones AM. The plant heterotrimeric G-protein complex. *Annu Rev Plant Biol.* 2007;58:249-266.

109. Trusov Y, Botella JR. Plant G-proteins come of age: breaking the bond with animal models. *Frontiers in chemistry.* 2016;4.

110. Temple BR, Jones CD, Jones AM. Evolution of a signaling nexus constrained by protein interfaces and conformational states. *PLoS computational biology.* 2010;6(10):e1000962.

111. Jones AM, Assmann SM. Plants: the latest model system for G‐protein research. *EMBO reports.* 2004;5(6):572-578.

112. Park HJ, Kim W-Y, Park HC, Lee SY, Bohnert HJ, Yun D-J. SUMO and SUMOylation in plants. *Molecules and cells.* 2011;32(4):305-316.

113. Poovaiah B, Du L, Wang H, Yang T. Recent advances in calcium/calmodulin-mediated signaling with an emphasis on plant-microbe interactions. *Plant physiology.* 2013;163(2):531-542.

114. Reddy AS, Ali GS, Celesnik H, Day IS. Coping with stresses: roles of calcium-and calcium/calmodulin-regulated gene expression. *The Plant Cell.* 2011;23(6):2010-2032.

115. Takahashi F, Mizoguchi T, Yoshida R, Ichimura K, Shinozaki K. Calmodulin-dependent activation of MAP kinase for ROS homeostasis in Arabidopsis. *Molecular cell.* 2011;41(6):649-660.

116. Gao X, Chen X, Lin W, et al. Bifurcation of Arabidopsis NLR immune signaling via Ca2+-dependent protein kinases. *PLoS pathogens.* 2013;9(1):e1003127.

117. Tsuda K, Mine A, Bethke G, et al. Dual regulation of gene expression mediated by extended MAPK activation and salicylic acid contributes to robust innate immunity in Arabidopsis thaliana. *PLoS Genetics.* 2013;9(12):e1004015.

118. Vlot AC, Dempsey DMA, Klessig DF. Salicylic acid, a multifaceted hormone to combat disease. *Annual review of phytopathology.* 2009;47:177-206.

119. Aarts N, Metz M, Holub E, Staskawicz BJ, Daniels MJ, Parker JE. Different requirements for EDS1 and NDR1 by disease resistance genes define at least two R gene-mediated signaling pathways in Arabidopsis. *Proceedings of the National Academy of Sciences.* 1998;95(17):10306-10311.

120. Century KS, Shapiro AD, Repetti PP, Dahlbeck D, Holub E, Staskawicz BJ. NDR1, a pathogen-induced component required for Arabidopsis disease resistance. *Science.* 1997;278(5345):1963-1965.

121. Tada Y, Spoel SH, Pajerowska-Mukhtar K, et al. Plant immunity requires conformational charges of NPR1 via S-nitrosylation and thioredoxins. *Science.* 2008;321(5891):952-956.

122. Lorang J, Kidarsa T, Bradford C, et al. Tricking the guard: exploiting plant defense for disease susceptibility. *Science.* 2012;338(6107):659-662.

123. Geraats BP, Bakker PA, Lawrence CB, Achuo EA, Höfte M, Van Loon L. Ethylene-insensitive tobacco shows differentially altered susceptibility to different pathogens. *Phytopathology.* 2003;93(7):813-821.

124. Vijayan P, Shockey J, Lévesque CA, Cook RJ. A role for jasmonate in pathogen defense of Arabidopsis. *Proceedings of the National Academy of Sciences.* 1998;95(12):7209-7214.

125. Thomma BP, Penninckx IA, Cammue BP, Broekaert WF. The complexity of disease signaling in Arabidopsis. *Current opinion in immunology.* 2001;13(1):63-68.

126. McConn M, Creelman RA, Bell E, Mullet JE. Jasmonate is essential for insect defense in Arabidopsis. *Proceedings of the National Academy of Sciences.* 1997;94(10):5473-5477.

127. Ruther J, Kleier S. Plant–plant signaling: ethylene synergizes volatile emission in Zea mays induced by exposure to (Z)-3-hexen-1-ol. *Journal of chemical ecology.* 2005;31(9):2217-2222.

128. Schmelz EA, Alborn HT, Tumlinson JH. Synergistic interactions between volicitin, jasmonic acid and ethylene mediate insect‐induced volatile emission in Zea mays. *Physiologia Plantarum.* 2003;117(3):403-412.

129. Chung SH, Rosa C, Scully ED, et al. Herbivore exploits orally secreted bacteria to suppress plant defenses. *Proceedings of the National Academy of Sciences.* 2013;110(39):15728-15733.

130. Denoux C, Galletti R, Mammarella N, et al. Activation of defense response pathways by OGs and Flg22 elicitors in Arabidopsis seedlings. *Molecular plant.* 2008;1(3):423-445.

131. Tsuda K, Somssich IE. Transcriptional networks in plant immunity. *New Phytologist.* 2015;206(3):932-947.

132. Kim J-G, Stork W, Mudgett MB. Xanthomonas type III effector XopD desumoylates tomato transcription factor SlERF4 to suppress ethylene responses and promote pathogen growth. *Cell host & microbe.* 2013;13(2):143-154.

133. Pieterse CM, Leon-Reyes A, Van der Ent S, Van Wees SC. Networking by small-molecule hormones in plant immunity. *Nature chemical biology.* 2009;5(5):308-316.

134. Ton J, Flors V, Mauch-Mani B. The multifaceted role of ABA in disease resistance. *Trends in plant science.* 2009;14(6):310-317.

135. Asselbergh B, De Vleesschauwer D, Höfte M. Global switches and fine-tuning—ABA modulates plant pathogen defense. *Molecular Plant-Microbe Interactions.* 2008;21(6):709-719.

136. Lind C, Dreyer I, López-Sanjurjo EJ, et al. Stomatal guard cells co-opted an ancient ABA-dependent desiccation survival system to regulate stomatal closure. *Current Biology.* 2015;25(7):928-935.

137. Melotto M, Underwood W, He SY. Role of stomata in plant innate immunity and foliar bacterial diseases. *Annu Rev Phytopathol.* 2008;46:101-122.

138. Barth C, Moeder W, Klessig DF, Conklin PL. The timing of senescence and response to pathogens is altered in the ascorbate-deficient Arabidopsis mutant vitamin c-1. *Plant Physiology.* 2004;134(4):1784-1792.

139. Pavet V, Olmos E, Kiddle G, et al. Ascorbic acid deficiency activates cell death and disease resistance responses in Arabidopsis. *Plant Physiology.* 2005;139(3):1291-1303.

140. Coppola M, Cascone P, Madonna V, et al. Plant-to-plant communication triggered by systemin primes anti-herbivore resistance in tomato. *Scientific Reports.* 2017;7(1):15522.

141. Qiu JL, Fiil BK, Petersen K, et al. Arabidopsis MAP kinase 4 regulates gene expression through transcription factor release in the nucleus. *The EMBO journal.* 2008;27(16):2214-2221.

142. Lu J, Ju H, Zhou G, et al. An EAR‐motif‐containing ERF transcription factor affects herbivore‐induced signaling, defense and resistance in rice. *The Plant Journal.* 2011;68(4):583-596.

143. Zhao Y, Wei T, Yin KQ, et al. Arabidopsis RAP2. 2 plays an important role in plant resistance to Botrytis cinerea and ethylene responses. *New Phytologist.* 2012;195(2):450-460.

144. Moffat CS, Ingle RA, Wathugala DL, Saunders NJ, Knight H, Knight MR. ERF5 and ERF6 play redundant roles as positive regulators of JA/Et-mediated defense against Botrytis cinerea in Arabidopsis. *PloS one.* 2012;7(4):e35995.

145. Meng X, Xu J, He Y, et al. Phosphorylation of an ERF transcription factor by Arabidopsis MPK3/MPK6 regulates plant defense gene induction and fungal resistance. *The Plant Cell.* 2013;25(3):1126-1142.

146. Woldemariam MG, Dinh ST, Oh Y, Gaquerel E, Baldwin IT, Galis I. NaMYC2 transcription factor regulates a subset of plant defense responses in Nicotiana attenuata. *BMC plant biology.* 2013;13(1):73.

147. Okada A, Okada K, Miyamoto K, et al. OsTGAP1, a bZIP transcription factor, coordinately regulates the inductive production of diterpenoid phytoalexins in rice. *Journal of Biological Chemistry.* 2009;284(39):26510-26518.

148. Seo PJ, Kim MJ, Park JY, et al. Cold activation of a plasma membrane‐tethered NAC transcription factor induces a pathogen resistance response in Arabidopsis. *The Plant Journal.* 2010;61(4):661-671.

149. Wu Y, Deng Z, Lai J, et al. Dual function of Arabidopsis ATAF1 in abiotic and biotic stress responses. *Cell research.* 2009;19(11):1279.

150. Lorenzo O, Chico JM, Sánchez-Serrano JJ, Solano R. JASMONATE-INSENSITIVE1 encodes a MYC transcription factor essential to discriminate between different jasmonate-regulated defense responses in Arabidopsis. *The Plant Cell.* 2004;16(7):1938-1950.

151. Jensen MK, Rung JH, Gregersen PL, et al. The HvNAC6 transcription factor: a positive regulator of penetration resistance in barley and Arabidopsis. *Plant molecular biology.* 2007;65(1-2):137-150.

152. Rinerson CI, Rabara RC, Tripathi P, Shen QJ, Rushton PJ. The evolution of WRKY transcription factors. *BMC plant biology.* 2015;15(1):66.

153. Chen L, Zhang L, Li D, Wang F, Yu D. WRKY8 transcription factor functions in the TMV-cg defense response by mediating both abscisic acid and ethylene signaling in Arabidopsis. *Proceedings of the National Academy of Sciences.* 2013;110(21):E1963-E1971.

154. Li G, Meng X, Wang R, et al. Dual-level regulation of ACC synthase activity by MPK3/MPK6 cascade and its downstream WRKY transcription factor during ethylene induction in Arabidopsis. *PLoS genetics.* 2012;8(6):e1002767.

155. Mao G, Meng X, Liu Y, Zheng Z, Chen Z, Zhang S. Phosphorylation of a WRKY transcription factor by two pathogen-responsive MAPKs drives phytoalexin biosynthesis in Arabidopsis. *The Plant Cell.* 2011;23(4):1639-1653.

156. Chang C, Yu D, Jiao J, Jing S, Schulze-Lefert P, Shen Q-H. Barley MLA immune receptors directly interfere with antagonistically acting transcription factors to initiate disease resistance signaling. *The Plant Cell.* 2013;25(3):1158-1173.

157. Inoue H, Hayashi N, Matsushita A, et al. Blast resistance of CC-NB-LRR protein Pb1 is mediated by WRKY45 through protein–protein interaction. *Proceedings of the National Academy of Sciences.* 2013;110(23):9577-9582.

158. Peng Y, Bartley LE, Chen X, et al. OsWRKY62 is a negative regulator of basal and Xa21-mediated defense against Xanthomonas oryzae pv. oryzae in rice. *Molecular plant.* 2008;1(3):446-458.

159. Tao Z, Liu H, Qiu D, et al. A pair of allelic WRKY genes play opposite roles in rice-bacteria interactions. *Plant Physiology.* 2009;151(2):936-948.

160. Lozano-Durán R, Macho AP, Boutrot F, Segonzac C, Somssich IE, Zipfel C. The transcriptional regulator BZR1 mediates trade-off between plant innate immunity and growth. *Elife.* 2013;2:e00983.

161. Wang J, Tao F, Tian W, et al. The wheat WRKY transcription factors TaWRKY49 and TaWRKY62 confer differential high-temperature seedling-plant resistance to Puccinia striiformis f. sp. tritici. *PloS one.* 2017;12(7):e0181963.

162. McLellan H, Boevink PC, Armstrong MR, et al. An RxLR effector from Phytophthora infestans prevents re-localisation of two plant NAC transcription factors from the endoplasmic reticulum to the nucleus. *PLoS pathogens.* 2013;9(10):e1003670.

163. Block A, Toruño TY, Elowsky CG, et al. The Pseudomonas syringae type III effector HopD1 suppresses effector‐triggered immunity, localizes to the endoplasmic reticulum, and targets the Arabidopsis transcription factor NTL9. *New Phytologist.* 2014;201(4):1358-1370.

164. Wang X, Goregaoker SP, Culver JN. Interaction of the Tobacco mosaic virus replicase protein with a NAC domain transcription factor is associated with the suppression of systemic host defenses. *Journal of virology.* 2009;83(19):9720-9730.

165. Schornack S, Moscou MJ, Ward ER, Horvath DM. Engineering plant disease resistance based on TAL effectors. *Annual review of phytopathology.* 2013;51:383-406.

166. Tian D, Wang J, Zeng X, et al. The rice TAL effector–dependent resistance protein XA10 triggers cell death and calcium depletion in the endoplasmic reticulum. *The Plant Cell.* 2014;26(1):497-515.

167. Römer P, Hahn S, Jordan T, Strauß T, Bonas U, Lahaye T. Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science.* 2007;318(5850):645-648.

168. Jaskiewicz M, Conrath U, Peterhänsel C. Chromatin modification acts as a memory for systemic acquired resistance in the plant stress response. *EMBO reports.* 2011;12(1):50-55.

169. Holeski LM, Jander G, Agrawal AA. Transgenerational defense induction and epigenetic inheritance in plants. *Trends in ecology & evolution.* 2012;27(11):618-626.

170. Dowen RH, Pelizzola M, Schmitz RJ, et al. Widespread dynamic DNA methylation in response to biotic stress. *Proceedings of the National Academy of Sciences.* 2012;109(32):E2183-E2191.

171. López Sánchez A, Stassen JH, Furci L, Smith LM, Ton J. The role of DNA (de) methylation in immune responsiveness of Arabidopsis. *The Plant Journal.* 2016;88(3):361-374.

172. Choi SM, Song HR, Han SK, et al. HDA19 is required for the repression of salicylic acid biosynthesis and salicylic acid‐mediated defense responses in Arabidopsis. *The Plant Journal.* 2012;71(1):135-146.

173. Kim K-C, Lai Z, Fan B, Chen Z. Arabidopsis WRKY38 and WRKY62 transcription factors interact with histone deacetylase 19 in basal defense. *The Plant Cell.* 2008;20(9):2357-2371.

174. Mosher RA, Durrant WE, Wang D, Song J, Dong X. A comprehensive structure–function analysis of Arabidopsis SNI1 defines essential regions and transcriptional repressor activity. *The Plant Cell.* 2006;18(7):1750-1765.

175. Burketova L, Trda L, Ott PG, Valentova O. Bio-based resistance inducers for sustainable plant protection against pathogens. *Biotechnology advances.* 2015;33(6):994-1004.

176. Walters DR, Paterson L. Parents lend a helping hand to their offspring in plant defence. *Biology letters.* 2012;8(5):871-873.

177. Ameye M, Audenaert K, De Zutter N, et al. Priming of wheat with the green leaf volatile Z-3-hexenyl acetate enhances defense against Fusarium graminearum but boosts deoxynivalenol production. *Plant physiology.* 2015;167(4):1671-1684.

178. Beachy RN. Mechanisms and applications of pathogen-derived resistance in transgenic plants. *Current Opinion in Biotechnology.* 1997;8(2):215-220.

179. Abel PP, Nelson RS, De B, et al. Delay of disease development in transgenic plants that express the tobacco mosaic virus coat protein gene. *Science.* 1986;232:738-744.

180. Namba S, Ling K, Gonsalves C, Slightom JL, Gonsalves D. Protection of transgenic plants expressing the coat protein gene of watermelon mosaic virus II or zucchini yellow mosaic virus against six potyviruses. *Phytopathology.* 1992;82(9):940-946.

181. Navarro L, Jay F, Nomura K, He SY, Voinnet O. Suppression of the microRNA pathway by bacterial effector proteins. *Science.* 2008;321(5891):964-967.

182. Qiao Y, Liu L, Xiong Q, et al. Oomycete pathogens encode RNA silencing suppressors. *Nature genetics.* 2013;45(3):330-333.

183. Pumplin N, Voinnet O. RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. *Nature Reviews Microbiology.* 2013;11(11):745.

184. Cruz LF, Rupp JLS, Trick HN, Fellers JP. Stable resistance to Wheat streak mosaic virus in wheat mediated by RNAi. *In Vitro Cellular & Developmental Biology-Plant.* 2014;50(6):665-672.

185. Tian B, Li J, Oakley TR, Todd TC, Trick HN. Host-derived artificial MicroRNA as an alternative method to improve soybean resistance to soybean cyst nematode. *Genes.* 2016;7(12):122.

186. Zhai J, Jeong D-H, De Paoli E, et al. MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes & development.* 2011;25(23):2540-2553.

187. Li F, Pignatta D, Bendix C, et al. MicroRNA regulation of plant innate immune receptors. *Proceedings of the National Academy of Sciences.* 2012;109(5):1790-1795.

188. Li Y, Zhao S-L, Li J-L, et al. Osa-miR169 negatively regulates rice immunity against the blast fungus Magnaporthe oryzae. *Frontiers in plant science.* 2017;8.

189. Yang S, Tang F, Zhu H. Alternative splicing in plant immunity. *International journal of molecular sciences.* 2014;15(6):10424-10445.

190. Tsuchiya T, Eulgem T. An alternative polyadenylation mechanism coopted to the Arabidopsis RPP7 gene through intronic retrotransposon domestication. *Proceedings of the National Academy of Sciences.* 2013;110(37):E3535-E3543.

191. Sela H, Spiridon LN, PETRESCU AJ, et al. Ancient diversity of splicing motifs and protein surfaces in the wild emmer wheat (Triticum dicoccoides) LR10 coiled coil (CC) and leucine‐rich repeat (LRR) domains. *Molecular plant pathology.* 2012;13(3):276-287.

192. Saintenac C, Zhang W, Salcedo A, et al. Identification of wheat gene Sr35 that confers resistance to Ug99 stem rust race group. *Science.* 2013;341(6147):783-786.

193. Halterman DA, Wei F, Wise RP. Powdery mildew-induced Mla mRNAs are alternatively spliced and contain multiple upstream open reading frames. *Plant physiology.* 2003;131(2):558-567.

194. Halterman D, Zhou F, Wei F, Wise RP, Schulze‐Lefert P. The MLA6 coiled‐coil, NBS‐LRR protein confers AvrMla6‐dependent resistance specificity to Blumeria graminis f. sp. hordei in barley and wheat. *The Plant Journal.* 2001;25(3):335-348.

195. Wang M, Weiberg A, Dellota Jr E, Yamane D, Jin H. Botrytis small RNA Bc-siR37 suppresses plant defense genes by cross-kingdom RNAi. *RNA biology.* 2017;14(4):421-428.

196. Weiberg A, Wang M, Lin F-M, et al. Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. *Science.* 2013;342(6154):118-123.

197. Qiao Y, Shi J, Zhai Y, Hou Y, Ma W. Phytophthora effector targets a novel component of small RNA pathway in plants to promote infection. *Proceedings of the National Academy of Sciences.* 2015;112(18):5850-5855.

198. Sugawara T, Trifonova EA, Kochetov AV, Kanayama Y. Expression of an extracellular ribonuclease gene increases resistance to Cucumber mosaic virus in tobacco. *BMC plant biology.* 2016;16(3):246.

199. Galiana E, Bonnet P, Conrod S, et al. RNase activity prevents the growth of a fungal pathogen in tobacco leaves and increases upon induction of systemic acquired resistance with elicitin. *Plant Physiology.* 1997;115(4):1557-1567.

200. Hugot K, Ponchet M, Marais A, Ricci P, Galiana E. A tobacco S-like RNase inhibits hyphal elongation of plant pathogens. *Molecular plant-microbe interactions.* 2002;15(3):243-250.

201. Filipenko E, Kochetov A, Kanayama Y, Malinovsky V, Shumny V. PR-proteins with ribonuclease activity and plant resistance against pathogenic fungi. *Russian Journal of Genetics: Applied Research.* 2013;3(6):474-480.

202. Caporale C, Di Berardino I, Leonardi L, et al. Wheat pathogenesis‐related proteins of class 4 have ribonuclease activity. *Febs Letters.* 2004;575(1-3):71-76.

203. Passardi F, Cosio C, Penel C, Dunand C. Peroxidases have more functions than a Swiss army knife. *Plant cell reports.* 2005;24(5):255-265.

204. Almagro L, Gómez Ros L, Belchi-Navarro S, Bru R, Ros Barceló A, Pedreno M. Class III peroxidases in plant defence reactions. *Journal of experimental botany.* 2008;60(2):377-390.

205. Kadota Y, Sklenar J, Derbyshire P, et al. Direct regulation of the NADPH oxidase RBOHD by the PRR-associated kinase BIK1 during plant immunity. *Mol Cell.* 2014;54(1):43-55.

206. Levine A, Tenhaken R, Dixon R, Lamb C. H 2 O 2 from the oxidative burst orchestrates the plant hypersensitive disease resistance response. *Cell.* 1994;79(4):583-593.

207. Mittler R, Herr EH, Orvar BL, et al. Transgenic tobacco plants with reduced capability to detoxify reactive oxygen intermediates are hyperresponsive to pathogen infection. *Proceedings of the National Academy of Sciences.* 1999;96(24):14165-14170.

208. Lamb C, Dixon RA. The oxidative burst in plant disease resistance. *Annual review of plant biology.* 1997;48(1):251-275.

209. Kärkönen A, Kuchitsu K. Reactive oxygen species in cell wall metabolism and development in plants. *Phytochemistry.* 2015;112:22-32.

210. Cessna SG, Sears VE, Dickman MB, Low PS. Oxalic acid, a pathogenicity factor for Sclerotinia sclerotiorum, suppresses the oxidative burst of the host plant. *The Plant Cell.* 2000;12(11):2191-2199.

211. Williams B, Kabbage M, Kim H-J, Britt R, Dickman MB. Tipping the balance: Sclerotinia sclerotiorum secreted oxalic acid suppresses host defenses by manipulating the host redox environment. *PLoS pathogens.* 2011;7(6):e1002107.

212. Lane B, Dunwell JM, Ray J, Schmitt M, Cuming A. Germin, a protein marker of early plant development, is an oxalate oxidase. *Journal of Biological Chemistry.* 1993;268(17):12239-12242.

213. Dumas B, Sailland A, Cheviet J, Freyssinet G, Pallett K. Identification of barley oxalate oxidase as a germin-like protein. *Comptes rendus de l'Academie des sciences Serie III, Sciences de la vie.* 1993;316(8):793-798.

214. Dumas B, Freyssinet G, Pallett KE. Tissue-specific expression of germin-like oxalate oxidase during development and fungal infection of barley seedlings. *Plant Physiology.* 1995;107(4):1091-1096.

215. Dong X, Ji R, Guo X, et al. Expressing a gene encoding wheat oxalate oxidase enhances resistance to Sclerotinia sclerotiorum in oilseed rape (Brassica napus). *Planta.* 2008;228(2):331.

216. Donaldson PA, Anderson T, Lane BG, Davidson AL, Simmonds DH. Soybean plants expressing an active oligomeric oxalate oxidase from the wheat gf-2.8 (germin) gene are resistant to the oxalate-secreting pathogen Sclerotina sclerotiorum. *Physiological and Molecular Plant Pathology.* 2001;59(6):297-307.

217. Liu F, Wang M, Wen J, et al. Overexpression of barley oxalate oxidase gene induces partial leaf resistance to Sclerotinia sclerotiorum in transgenic oilseed rape. *Plant Pathology.* 2015;64(6):1407-1416.

218. Liang H, Maynard CA, Allen RD, Powell WA. Increased Septoria musiva resistance in transgenic hybrid poplar leaves expressing a wheat oxalate oxidase gene. *Plant molecular biology.* 2001;45(6):619-629.

219. Ramputh A, Arnason J, Cass L, Simmonds J. Reduced herbivory of the European corn borer (Ostrinia nubilalis) on corn transformed with germin, a wheat oxalate oxidase gene. *Plant Science.* 2002;162(3):431-440.

220. Wei Y, Zhang Z, Andersen CH, et al. An epidermis/papilla-specific oxalate oxidase-like protein in the defence response of barley attacked by the powdery mildew fungus. *Plant molecular biology.* 1998;36(1):101-112.

221. Zimmermann G, Bäumlein H, Mock H-P, Himmelbach A, Schweizer P. The multigene family encoding germin-like proteins of barley. Regulation and function in basal host resistance. *Plant Physiology.* 2006;142(1):181-192.

222. Bradley DJ, Kjellbom P, Lamb CJ. Elicitor-and wound-induced oxidative cross-linking of a proline-rich plant cell wall protein: a novel, rapid defense response. *Cell.* 1992;70(1):21-30.

223. Melotto M, Underwood W, Koczan J, Nomura K, He SY. Plant stomata function in innate immunity against bacterial invasion. *Cell.* 2006;126(5):969-980.

224. Bellincampi D, Cervone F, Lionetti V. Plant cell wall dynamics and wall-related susceptibility in plant–pathogen interactions. *Frontiers in plant science.* 2014;5.

225. Kang Z, Buchenauer H. Ultrastructural and cytochemical studies on cellulose, xylan and pectin degradation in wheat spikes infected by Fusarium culmorum. *Journal of Phytopathology.* 2000;148(5):263-275.

226. Wanjiru WM, Zhensheng K, Buchenauer H. Importance of cell wall degrading enzymes produced by Fusarium graminearum during infection of wheat heads. *European Journal of Plant Pathology.* 2002;108(8):803-810.

227. Kong L, Anderson JM, Ohm HW. Induction of wheat defense and stress-related genes in response to Fusarium graminearum. *Genome.* 2005;48(1):29-40.

228. Anguelova‐Merhar V, Westhuizen A, Pretorius Z. β‐1, 3‐Glucanase and Chitinase Activities and the Resistance Response of Wheat to Leaf Rust. *Journal of Phytopathology.* 2001;149(7‐8):381-384.

229. Collinge DB, Kragh KM, Mikkelsen JD, Nielsen KK, Rasmussen U, Vad K. Plant chitinases. *The Plant Journal.* 1993;3(1):31-40.

230. Broekaert W, Van Parijs J, Allen A, Peumans W. Comparison of some molecular, enzymatic and antifungal properties of chitinases from thorn-apple, tobacco and wheat. *Physiological and molecular plant pathology.* 1988;33(3):319-331.

231. Singh A, Kirubakaran SI, Sakthivel N. Heterologous expression of new antifungal chitinase from wheat. *Protein expression and purification.* 2007;56(1):100-109.

232. van Scheltinga ACT, Kalk KH, Beintema JJ, Dijkstra BW. Crystal structures of hevamine, a plant defence protein with chitinase and lysozyme activity, and its complex with an inhibitor. *Structure.* 1994;2(12):1181-1189.

233. Nagano N, Orengo CA, Thornton JM. One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *Journal of molecular biology.* 2002;321(5):741-765.

234. Leubner-Metzger G, Meins Jr F. 3 Functions and Regulation of Plant β-(PR-2). *Pathogenesis-related proteins in plants.* 1999.

235. Doxey AC, Yaish MW, Moffatt BA, Griffith M, McConkey BJ. Functional divergence in the Arabidopsis β-1, 3-glucanase gene family inferred by phylogenetic reconstruction of expression states. *Molecular biology and evolution.* 2007;24(4):1045-1055.

236. Jashni MK, Mehrabi R, Collemare J, Mesarich CH, De Wit PJ. The battle in the apoplast: further insights into the roles of proteases and their inhibitors in plant–pathogen interactions. *Frontiers in plant science.* 2015;6.

237. Thomas EL, van der Hoorn RA. Ten prominent host proteases in plant-pathogen interactions. *International journal of molecular sciences.* 2018;19(2):639.

238. Naumann TA, Wicklow DT, Price NP. Identification of a chitinase-modifying protein from Fusarium verticillioides truncation of a host resistance protein by a fungalysin metalloprotease. *Journal of biological chemistry.* 2011;286(41):35358-35366.

239. Slavokhotova AA, Naumann TA, Price NP, et al. Novel mode of action of plant defense peptides‐hevein‐like antimicrobial peptides from wheat inhibit fungal metalloproteases. *The FEBS journal.* 2014;281(20):4754-4764.

240. Pekkarinen AI, Longstaff C, Jones BL. Kinetics of the inhibition of Fusarium serine proteinases by barley (Hordeum vulgare L.) inhibitors. *Journal of agricultural and food chemistry.* 2007;55(7):2736-2742.

241. Franco OL, Rigden DJ, R Melo F, Bloch C, Silva CP, Grossi de Sá MF. Activity of wheat α‐amylase inhibitors towards bruchid α‐amylases and structural explanation of observed specificities. *The FEBS Journal.* 2000;267(8):2166-2173.

242. Feng GH, Richardson M, Chen MS, Kramer KJ, Morgan TD, Reeck GR. α-Amylase inhibitors from wheat: amino acid sequences and patterns of inhibition of insect and human α-amylases. *Insect biochemistry and molecular biology.* 1996;26(5):419-426.

243. Mundy J, Svendsen I, Hejgaard J. Barley α-amylase/subtilisin inhibitor. I. Isolation and characterization. *Carlsberg Research Communications.* 1983;48(2):81-90.

244. Petrucci T, Rab A, Tomasi M, Silano V. Further characterization studies of the α-amylase protein inhibitor of gel electrophoretic mobility 0.19 from the wheat kernel. *Biochimica et Biophysica Acta (BBA)-Protein Structure.* 1976;420(2):288-297.

245. Franco OL, Rigden DJ, Melo FR, Grossi‐de‐Sá MF. Plant α‐amylase inhibitors and their interaction with insect α‐amylases. *The FEBS Journal.* 2002;269(2):397-412.

246. Steppuhn A, Baldwin IT. Resistance management in a native plant: nicotine prevents herbivores from compensating for plant protease inhibitors. *Ecology Letters.* 2007;10(6):499-511.

247. Shabab M, Shindo T, Gu C, et al. Fungal effector protein AVR2 targets diversifying defense-related cys proteases of tomato. *The Plant Cell.* 2008;20(4):1169-1183.

248. Rooney HC, van't Klooster JW, van der Hoorn RA, Joosten MH, Jones JD, de Wit PJ. Cladosporium Avr2 inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance. *Science.* 2005;308(5729):1783-1786.

249. Song J, Win J, Tian M, et al. Apoplastic effectors secreted by two unrelated eukaryotic plant pathogens target the tomato defense protease Rcr3. *Proceedings of the National Academy of Sciences.* 2009;106(5):1654-1659.

250. Lozano-Torres JL, Wilbers RH, Gawronski P, et al. Dual disease resistance mediated by the immune receptor Cf-2 in tomato requires a common virulence target of a fungus and a nematode. *Proceedings of the National Academy of Sciences.* 2012;109(25):10119-10124.

251. Reisige K, Gorzelanny C, Daniels U, Moerschbacher BM. The C28 aldehyde octacosanal is a morphogenetically active component involved in host plant recognition and infection

structure differentiation in the wheat stem rust fungus. *Physiological and molecular plant pathology.* 2006;68(1):33-40.

252. Tsuba M, Katagiri C, Takeuchi Y, Takada Y, Yamaoka N. Chemical factors of the leaf surface involved in the morphogenesis of Blumeria graminis. *Physiological and Molecular Plant Pathology.* 2002;60(2):51-57.

253. Woloshuk CP, Kolattukudy P. Mechanism by which contact with plant cuticle triggers cutinase gene expression in the spores of Fusarium solani f. sp. pisi. *Proceedings of the National Academy of Sciences.* 1986;83(6):1704-1708.

254. Oh IS, Park AR, Bae MS, et al. Secretome analysis reveals an Arabidopsis lipase involved in defense against Alternaria brassicicola. *The Plant Cell.* 2005;17(10):2832-2847.

255. Gottwald S, Samans B, Lück S, Friedt W. Jasmonate and ethylene dependent defence gene expression and suppression of fungal virulence factors: two essential mechanisms of Fusarium head blight resistance in wheat? *BMC genomics.* 2012;13(1):369.

256. Finkina E, Melnikova D, Bogdanov I. Lipid transfer proteins as components of the plant innate immune system: structure, functions, and applications. *Acta Naturae (англоязычная версия).* 2016;8(2 (29)).

257. Nimchuk Z, Marois E, Kjemtrup S, Leister RT, Katagiri F, Dangl JL. Eukaryotic fatty acylation drives plasma membrane targeting and enhances function of several type III effector proteins from Pseudomonas syringae. *Cell.* 2000;101(4):353-363.

258. Boyes DC, Nam J, Dangl JL. The Arabidopsis thaliana RPM1 disease resistance gene product is a peripheral plasma membrane protein that is degraded coincident with the hypersensitive response. *Proceedings of the National Academy of Sciences.* 1998;95(26):15849-15854.

259. Leister RT, Ausubel FM, Katagiri F. Molecular recognition of pathogen attack occurs inside of plant cells in plant disease resistance specified by the Arabidopsis genes RPS2 and RPM1. *Proceedings of the National Academy of Sciences.* 1996;93(26):15497-15502.

260. Abbas HK, Tanaka T, Duke SO, et al. Fumonisin-and AAL-toxin-induced disruption of sphingolipid metabolism with accumulation of free sphingoid bases. *Plant Physiology.* 1994;106(3):1085-1093.

261. Desaki Y, Miya A, Venkatesh B, et al. Bacterial lipopolysaccharides induce defense responses associated with programmed cell death in rice cells. *Plant and cell physiology.* 2006;47(11):1530-1540.

262. Zeidler D, Zähringer U, Gerber I, et al. Innate immunity in Arabidopsis thaliana: lipopolysaccharides activate nitric oxide synthase (NOS) and induce defense genes. *Proceedings of the National Academy of Sciences of the United States of America.* 2004;101(44):15811-15816.

263. Canonne J, Froidure-Nicolas S, Rivas S. Phospholipases in action during plant defense signaling. *Plant signaling & behavior.* 2011;6(1):13-18.

264. Montillet J-L, Leonhardt N, Mondy S, et al. An abscisic acid-independent oxylipin pathway controls stomatal closure and immune defense in Arabidopsis. *PLoS biology.* 2013;11(3):e1001513.

265. Zhao J. Phospholipase D and phosphatidic acid in plant defence response: from protein–protein and lipid–protein interactions to hormone signalling. *Journal of Experimental Botany.* 2015;66(7):1721-1736.

266. Rentel MC, Lecourieux D, Ouaked F, Usher SL. OXI1 kinase is necessary for oxidative burst-mediated signalling in Arabidopsis. *Nature.* 2004;427(6977):858.

267. de Zélicourt A, Letousey P, Thoiron S, et al. Ha-DEF1, a sunflower defensin, induces cell death in Orobanche parasitic plants. *Planta.* 2007;226(3):591-600.

268. Mendez E, MORENO A, COLILLA F, et al. Primary structure and inhibition of protein synthesis in eukaryotic cell‐free system of a novel thionin, γ‐hordothionin, from barley endosperm. *The FEBS Journal.* 1990;194(2):533-539.

269. Colilla FJ, Rocher A, Mendez E. γ-Purothionins: amino acid sequence of two polypeptides of a new family of thionins from wheat endosperm. *FEBS letters.* 1990;270(1-2):191-194.

270. Bruix M, Jimenez M, Santoro J, et al. Solution structure of. gamma. 1-H and. gamma. 1-P thionins from barley and wheat endosperm determined by proton NMR: a structural motif common to toxic arthropod proteins. *Biochemistry.* 1993;32(2):715-724.

271. Terras FRG, Eggermont K, Kovaleva V, et al. Small cysteine-rich antifungal proteins from radish: their role in host defense. *The Plant Cell.* 1995;7(5):573-588.

272. Méndez E, Rocher A, Calero M, Girbés T, Citores L, Soriano F. Primary Structure of ω‐Hordothionin, a Member of a Novel Family of Thionins from Barley Endosperm, and Its Inhibition of Protein Synthesis in Eukaryotic and Prokaryotic Cell‐Free Systems. *The FEBS Journal.* 1996;239(1):67-73.

273. Van Der Weerden NL, Lay FT, Anderson MA. The plant defensin, NaD1, enters the cytoplasm of Fusarium oxysporum hyphae. *Journal of Biological Chemistry.* 2008;283(21):14445-14452.

274. van der Weerden NL, Hancock RE, Anderson MA. Permeabilization of fungal hyphae by the plant defensin NaD1 occurs through a cell wall-dependent process. *Journal of Biological Chemistry.* 2010;285(48):37513-37520.

275. Koike M, Okamoto T, Tsuda S, Imai R. A novel plant defensin-like gene of winter wheat is specifically induced during cold acclimation. *Biochemical and biophysical research communications.* 2002;298(1):46-53.

276. Wijaya R, Neumann GM, Condron R, Hughes AB, Polya GM. Defense proteins from seed of Cassia fistula include a lipid transfer protein homologue and a protease inhibitory plant defensin. *Plant Science.* 2000;159(2):243-255.

277. Aerts AM, François IE, Meert EM, Li Q-T, Cammue BP, Thevissen K. The antifungal activity of RsAFP2, a plant defensin from Raphanus sativus, involves the induction of reactive oxygen species in Candida albicans. *Journal of molecular microbiology and biotechnology.* 2007;13(4):243.

278. Spelbrink RG, Dilmac N, Allen A, Smith TJ, Shah DM, Hockerman GH. Differential antifungal and calcium channel-blocking activity among structurally related plant defensins. *Plant Physiology.* 2004;135(4):2055-2067.

279. Stotz HU, Thomson J, Wang Y. Plant defensins: defense, development and application. *Plant signaling & behavior.* 2009;4(11):1010-1012.

280. García‐Olmedo F, Molina A, Alamillo JM, Rodríguez‐Palenzuéla P. Plant defense peptides. *Peptide Science.* 1998;47(6):479-491.

281. De Paula VS, Razzera G, Barreto-Bergter E, Almeida FC, Valente AP. Portrayal of complex dynamic properties of sugarcane defensin 5 by NMR: multiple motions associated with membrane interaction. *Structure.* 2011;19(1):26-36.

282. Lacerda AF, Vasconcelos ÉA, Pelegrini PB, de Sa MFG. Antifungal defensins and their role in plant defense. *Frontiers in microbiology.* 2014;5.

283. Vriens K, Peigneur S, De Coninck B, Tytgat J, Cammue BP, Thevissen K. The antifungal plant defensin AtPDF2. 3 from Arabidopsis thaliana blocks potassium channels. *Scientific Reports.* 2016;6.

284. Almeida MS, Cabral KM, Kurtenbach E, Almeida FC, Valente AP. Solution structure of Pisum sativum defensin 1 by high resolution NMR: plant defensins, identical backbone with different mechanisms of action. *Journal of molecular biology.* 2002;315(4):749-757.

285. Brogden KA. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nature reviews Microbiology.* 2005;3(3):238.

286. Silverstein KA, Graham MA, Paape TD, VandenBosch KA. Genome organization of more than 300 defensin-like genes in Arabidopsis. *Plant physiology.* 2005;138(2):600-610.

287. Graham MA, Silverstein KA, Cannon SB, VandenBosch KA. Computational identification and characterization of novel genes from legumes. *Plant physiology.* 2004;135(3):1179-1197.

288. Van der Weerden NL, Anderson MA. Plant defensins: common fold, multiple functions. *Fungal Biology Reviews.* 2013;26(4):121-131.

289. Wang S, Rao P, Ye X. Isolation and biochemical characterization of a novel leguminous defense peptide with antifungal and antiproliferative potency. *Applied microbiology and biotechnology.* 2009;82(1):79-86.

290. Ming D, Hellekant G. Brazzein, a new high-potency thermostable sweet protein from Pentadiplandra brazzeana B. *FEBS letters.* 1994;355(1):106-108.

291. Wel H, Loeve K. Isolation and characterization of thaumatin I and II, the sweet‐tasting proteins from Thaumatococcus daniellii Benth. *The FEBS Journal.* 1972;31(2):221-225.

292. Osmond RI, Hrmova M, Fontaine F, Imberty A, Fincher GB. Binding interactions between barley thaumatin‐like proteins and (1, 3)‐β‐D‐glucans. *The FEBS Journal.* 2001;268(15):4190-4199.

293. Bryngelsson T, Green B. Characterization of a pathogenesis-related, thaumatin-like protein isolated from barley challenged with an incompatible race of mildew. *Physiological and molecular plant pathology.* 1989;35(1):45-52.

294. Trümper C, Paffenholz K, Smit I, et al. Identification of regulated proteins in naked barley grains (Hordeum vulgare nudum) after Fusarium graminearum infection at different grain ripening stages. *Journal of proteomics.* 2016;133:86-92.

295. Hejgaard J, Jacobsen S, Svendsen I. Two antifungal thaumatin‐like proteins from barley grain. *FEBS letters.* 1991;291(1):127-131.

296. Vigers AJ, Roberts WK, Selitrennikoff CP. A new family of plant antifungal proteins. *Mol Plant Microbe Interact.* 1991;4(4):315-323.

297. Roberts WK, Selitrennikoff CP. Zeamatin, an antifungal protein from maize with membrane-permeabilizing activity. *Microbiology.* 1990;136(9):1771-1778.

298. Abad LR, D'Urzo MP, Liu D, et al. Antifungal activity of tobacco osmotin has specificity and involves plasma membrane permeabilization. *Plant Science.* 1996;118(1):11-23.

299. Skadsen R, Sathish P, Kaeppler H. Expression of thaumatin-like permatin PR-5 genes switches from the ovary wall to the aleurone in developing barley and oat seeds. *Plant Science.* 2000;156(1):11-22.

300. Hahn M, Jüngling S, Knogge W. Cultivar-specific elicitation of barley defense reactions by the phytotoxic peptide NIP1 from Rhynchosporium secalis. *MPMI-Molecular Plant Microbe Interactions.* 1993;6(6):745-754.

301. Rebmann G, Mauch F, Dudler R. Sequence of a wheat cDNA encoding a pathogen-induced thaumatin-like protein. *Plant molecular biology.* 1991;17(2):283-285.

302. Lin K-C, Bushnell WR, Szabo LJ, Smith AG. Isolation and expression of a host response gene family encoding thaumatin-like proteins in incompatible oat-stem rust fungus interactions. *Molecular plant-microbe interactions: MPMI.* 1996;9(6):511-522.

303. Singh S, Tripathi RK, Lemaux PG, Buchanan BB, Singh J. Redox-dependent interaction between thaumatin-like protein and β-glucan influences malting quality of barley. *Proceedings of the National Academy of Sciences.* 2017:201701824.

304. Van Loon L, Van Strien E. The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. *Physiological and molecular plant pathology.* 1999;55(2):85-97.

305. Ebrahim S, Usha K, Singh B. Pathogenesis related (PR) proteins in plant defense mechanism. *Sci Against Microb Pathog.* 2011;2:1043-1054.

306. Sinha M, Singh RP, Kushwaha GS, et al. Current overview of allergens of plant pathogenesis related protein families. *The Scientific World Journal.* 2014;2014.

307. Kettles GJ, Drurey C, Schoonbeek Hj, Maule AJ, Hogenhout SA. Resistance of Arabidopsis thaliana to the green peach aphid, Myzus persicae, involves camalexin and is regulated by microRNAs. *New Phytologist.* 2013;198(4):1178-1190.

308. Xu J, Li Y, Wang Y, et al. Activation of MAPK kinase 9 induces ethylene and camalexin biosynthesis and enhances sensitivity to salt stress in Arabidopsis. *Journal of Biological Chemistry.* 2008;283(40):26996-27006.

309. Pandey SP, Roccaro M, Schön M, Logemann E, Somssich IE. Transcriptional reprogramming regulated by WRKY18 and WRKY40 facilitates powdery mildew infection of Arabidopsis. *The Plant Journal.* 2010;64(6):912-923.

310. Ahuja I, Kissen R, Bones AM. Phytoalexins in defense against pathogens. *Trends in plant science.* 2012;17(2):73-90.

311. Pedras MSC, Hossain S, Snitynsky RB. Detoxification of cruciferous phytoalexins in Botrytis cinerea: Spontaneous dimerization of a camalexin metabolite. *Phytochemistry.* 2011;72(2):199-206.

312. Weibull J, Niemeyer H. Changes in dihydroxymethoxybenzoxazinone glycoside content in wheat plants infected by three plant pathogenic fungi. *Physiological and molecular plant pathology.* 1995;47(3):201-212.

313. Wouters FC, Blanchette B, Gershenzon J, Vassão DG. Plant defense and herbivore counter-defense: benzoxazinoids and insect herbivores. *Phytochemistry Reviews.* 2016;15(6):1127-1151.

314. Givovich A, Niemeyer H. Hydroxamic acids affecting barley yellow dwarf virus transmission by the aphid Rhopalosiphum padi. *Entomologia experimentalis et applicata.* 1991;59(1):79-85.

315. Friebe A, Vilich V, Hennig L, Kluge M, Sicker D. Detoxification of Benzoxazolinone Allelochemicals from Wheat byGaeumannomyces graminis var. tritici, G. graminis var. graminis, G. graminis var. avenae, and Fusarium culmorum. *Applied and environmental microbiology.* 1998;64(7):2386-2391.

316. Niemeyer HM, Jerez JM. Chromosomal location of genes for hydroxamic acid accumulation in Triticum aestivum L.(wheat) using wheat aneuploids and wheat substitution lines. *Heredity.* 1997;79(1).

317. Osbourn A, Clarke B, Lunness P, Scott P, Daniels M. An oat species lacking avenacin is susceptible to infection by Gaeumannomyces graminis var. tritici. *Physiological and Molecular Plant Pathology.* 1994;45(6):457-467.

318. Morrissey JP, Osbourn AE. Fungal resistance to plant antibiotics as a mechanism of pathogenesis. *Microbiology and Molecular Biology Reviews.* 1999;63(3):708-724.

319. González-Lamothe R, Mitchell G, Gattuso M, Diarra MS, Malouin F, Bouarab K. Plant antimicrobial agents and their effects on plant and human pathogens. *International Journal of Molecular Sciences.* 2009;10(8):3400-3419.

320. Osbourn A, Clarke B, Dow J, Daniels M. Partial characterization of avenacinase from Gaeumannomyces graminis var. avenae. *Physiological and molecular plant pathology.* 1991;38(4):301-312.

321. Wu H, Pratley J, Lemerle D, Haig T. Allelopathy in wheat (Triticum aestivum). *Annals of Applied Biology.* 2001;139(1):1-9.

322. Adhikari KB, Tanwir F, Gregersen PL, et al. Benzoxazinoids: Cereal phytochemicals with putative therapeutic and health‐protecting properties. *Molecular nutrition & food research.* 2015;59(7):1324-1338.

323. Martinez DA, Loening UE, Graham MC. Impacts of glyphosate-based herbicides on disease resistance and health of crops: a review. *Environmental Sciences Europe.* 2018;30(1):2.

324. Turlings TC, Tumlinson JH, Lewis WJ. Exploitation of herbivore-induced plant odors by host-seeking parasitic wasps. *Science.* 1990;250(4985):1251.

325. Turlings TC, Tumlinson JH, Heath RR, Proveaux AT, Doolittle RE. Isolation and identification of allelochemicals that attract the larval parasitoid, Cotesia marginiventris (Cresson), to the microhabitat of one of its hosts. *Journal of chemical ecology.* 1991;17(11):2235-2251.

326. Turlings TC, McCall PJ, Alborn HT, Tumlinson JH. An elicitor in caterpillar oral secretions that induces corn seedlings to emit chemical signals attractive to parasitic wasps. *Journal of Chemical Ecology.* 1993;19(3):411-425.

327. McKay SAB, Hunter WL, Godard K-A, et al. Insect attack and wounding induce traumatic resin duct development and gene expression of (—)-pinene synthase in Sitka spruce. *Plant Physiology.* 2003;133(1):368-378.

328. Agrawal AA, Konno K. Latex: a model for understanding mechanisms, ecology, and evolution of plant defense against herbivory. *Annu Rev Ecol Evol Syst.* 2009;40:311-331.

329. Becerra JX. Squirt‐Gun Defense in Bursera and the Chrysomelid Counterploy. *Ecology.* 1994;75(7):1991-1996.

330. Thomson VP, Cunningham SA, Ball MC, Nicotra AB. Compensation for herbivory by Cucumis sativus through increased photosynthetic capacity and efficiency. *Oecologia.* 2003;134(2):167-175.

331. War AR, Paulraj MG, Ahmad T, et al. Mechanisms of plant defense against insect herbivores. *Plant signaling & behavior.* 2012;7(10):1306-1320.

332. Cook RJ, Thomashow LS, Weller DM, et al. Molecular mechanisms of defense by rhizobacteria against root disease. *Proceedings of the National Academy of Sciences.* 1995;92(10):4197-4201.

333. Campos‐Soriano L, José García‐Martínez, and Blanca San Segundo. The arbuscular mycorrhizal symbiosis promotes the systemic induction of regulatory defence‐related genes in rice leaves and confers resistance to pathogen infection. *Molecular plant pathology.* 2012;13(6):579-592.

334. Song YY, Cao M, Xie LJ, et al. Induction of DIMBOA accumulation and systemic defense responses as a mechanism of enhanced resistance of mycorrhizal corn (Zea mays L.) to sheath blight. *Mycorrhiza.* 2011;21(8):721-731.

335. Plett JM, Daguerre Y, Wittulsky S, et al. Effector MiSSP7 of the mutualistic fungus Laccaria bicolor stabilizes the Populus JAZ6 protein and represses jasmonic acid (JA) responsive genes. *Proceedings of the National Academy of Sciences.* 2014;111(22):8299-8304.

336. Cárdenas L, Martínez A, Sánchez F, Quinto C. Fast, transient and specific intracellular ROS changes in living root hair cells responding to Nod factors (NFs). *The Plant Journal.* 2008;56(5):802-813.

337. Mitra RM, Long SR. Plant and bacterial symbiotic mutants define three transcriptionally distinct stages in the development of the Medicago truncatula/Sinorhizobium meliloti symbiosis. *Plant physiology.* 2004;134(2):595-604.

338. Zhan J, Thrall PH, Papaïx J, Xie L, Burdon JJ. Playing on a pathogen's weakness: using evolution to guide sustainable plant disease control strategies. *Annual review of phytopathology.* 2015;53:19-43.

339. Shao F, Golstein C, Ade J, Stoutemyer M, Dixon JE, Innes RW. Cleavage of Arabidopsis PBS1 by a bacterial type III effector. *Science.* 2003;301(5637):1230-1233.

340. Qi D, Dubiella U, Kim SH, et al. Recognition of the protein kinase AVRPPHB SUSCEPTIBLE1 by the disease resistance protein RESISTANCE TO PSEUDOMONAS SYRINGAE5 is dependent on s-acylation and an exposed loop in AVRPPHB SUSCEPTIBLE1. *Plant physiology.* 2014;164(1):340-351.

341. Kim SH, Qi D, Ashfield T, Helm M, Innes RW. Using decoys to expand the recognition specificity of a plant disease resistance protein. *Science.* 2016;351(6274):684-687.

342. Sarma BK, Singh HB, Fernando D, Silva RN, Gupta VK. Enhancing Plant disease resistance without r genes. *Trends in biotechnology.* 2016;34(7):523-525.

343. De Wit PJ. Host resistance to a fungal tomato pathogen lost by a single base-pair change in an avirulence gene. *Nature.* 1994;367.

344. Zhou H, Steffenson B. Genome-wide association mapping reveals genetic architecture of durable spot blotch resistance in US barley breeding germplasm. *Molecular breeding.* 2013;32(1):139-154.

345. Wilcoxson R, Rasmusson D, Miles M. Development of barley resistant to spot blotch and genetics of resistance. *Plant Disease.* 1990;74(3):207-210.

346. Haas M, Menke J, Chao S, Steffenson BJ. Mapping quantitative trait loci conferring resistance to a widely virulent isolate of Cochliobolus sativus. *Theoretical and Applied Genetics.* 2016:1-12.

347. Valjavec-Gratian M, Steffenson B. Pathotypes of Cochliobolus sativus on barley in North Dakota. *Plant Disease.* 1997;81(11):1275-1278.

CHAPTER 2: DIVERSITY AND EVOLUTION OF DISEASE RESISTANCE GENES IN BARLEY (*HORDEUM VULGARE* L.)

This chapter is published in the journal *Evolutionary Bioinformatics*:

Andersen, E.J., S. Ali, R.N. Reese, Y. Yen, S. Neupane, and M.P. Nepal. 2016. Diversity and Evolution of Disease Resistance Genes in Barley (*Hordeum vulgare* L.). *Evolutionary Bioinformatics*. 12:99.

ABSTRACT

Plant disease resistance genes (R-genes) play a critical role in the defense response to pathogens. Barley is one of the most important cereal crops, having a genome recently made available, for which the diversity and evolution of R-genes are not well-understood. The main objectives of this research were to conduct a genome-wide identification of barley Coiled-coil, Nucleotide Binding Site, and Leucine Rich Repeat (CNL) genes and elucidate their evolutionary history. A Hidden Markov Model was employed using 52 *Arabidopsis thaliana* CNL reference sequences and analyzed for phylogenetic relationships, structural variation, and gene clustering. A total of 175 barley CNL genes were identified, nested into three clades, showing a) evidence of an expansion of the CNL-C clade, primarily due to tandem duplications, b) very few members of clade CNL-A and CNL-B, and c) a complete absence of CNL-D clade. The results also showed that several of previously identified mildew locus A (*MLA*) genes may be allelic variants of two barley CNL genes, MLOC_66581 and MLOC_10425, which respond to powdery mildew. Approximately 23% of the barley CNL genes formed 15 gene clusters located in the extra-pericentromeric regions on six of the seven chromosomes; over half of the

clusters were located on chromosomes 1H and 7H. Higher average numbers of exons and multiple splice variants in barley relative to that in *Arabidopsis* and rice may have contributed to a diversification of the CNL-C members. These results will help us understand the evolution of R-genes with potential implications for developing durable resistance in barley cultivars.

INTRODUCTION

Plants have evolved complex signaling pathways for pathogen detection and defense response [1]. Lacking an adaptive immunity and cell-transporting circulatory system, plant resistance to pathogens depends upon innate immunity that utilizes molecular signaling to initiate local and systemic responses [2]. Resistance genes (R-genes) encode proteins that detect pathogens [3,4]. Plant immunity can be divided into two types: pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI) and effector-triggered immunity (ETI) [2,5]. PAMPs are pathogen structural molecules, such as bacterial flagellin, peptidoglycan, or fungal chitin, that the plant's immune system perceives through membrane-localized, receptor-like kinases called pattern recognition receptors (PRRs), which elicit a response [6,7]. In contrast, ETI involves the interaction between specific pathogen effectors and NBS-LRR receptors within the cell[5]. Resistance responses vary widely and act in limiting the spread and effectiveness of the pathogen [3] including: 1) causing localized death of infected tissue through hypersensitive response [8], 2) promoting hostile conditions for pathogens such as hydrogen peroxide production in an oxidative burst[9], and 3) fortifying cell walls to strengthen the physical barrier between

pathogens and the plant protoplasm [10]. Resistance responses are expensive for the cell [11], therefore, in the absence of a pathogen, diverse down-regulating control factors are mobilized [12], including salicylic acid production for localized and systemic resistance [13,14], WRKY transcription factors[15], and silencing through micro-RNA [16].

Several models have been proposed to describe the mechanism of host-pathogen relationship. The Gene-for-Gene Model involves direct interaction between a single pathogen avirulence-gene and a plant R-gene [17]. Additionally, there is evidence of indirect interaction as described in the Guard Model, where R-proteins bind with or guard, particular target proteins, activating a response when the 'guarded' protein is cleaved or modified by a pathogen [18,19]. Similar to the Guard Model, the Decoy Model describes specific decoy proteins that mimic unguarded pathogen effector targets, forming a complex with effectors that is perceived by NBS-LRR R-proteins[20]. With increasing understanding of molecular interactions between the pathogen and host, the Zig-Zag Model was proposed to describe co-evolution of plant R-genes and pathogen effectors[2]. In this model, the pathogen evolves effectors to reduce the effectiveness of the plant's PTI response, and the plant responds to these newly-evolved effectors by developing receptors that initiate ETI [2]. Intense selection pressures from pathogens cause R-genes to evolve rapidly through several mechanisms, including recombination and transposable elements [4,21,22]. However, R-genes can also be removed from the genome through loss of lineages and deficient duplications [23].

R-genes have been recently classified into eight major groups: 1) Toll Interleukin Receptor, Nucleotide Binding Site, Leucine Rich Repeat (TIR-NBS-LRR or TNL); 2) Coiled-Coil, NBS, LRR (CC-NBS-LRR or CNL), 3) LRR Transmembrane domain

(LRR-TrD); 4) LRR-TrD-Kinase; 5) LRR TrD Protein degradation domain proline-glycine-serine-threonine (LRR-TrD-PEST); 6) TrD-CC; 7) TNL-nuclear localization signal amino acid domain (TNL-NLS-WRKY); and 8) Enzymatic genes [24]. Among these groups, the NBS-LRR (TNL and CNL) genes form the largest group and respond to various pests and pathogens. NB-ARC-containing genes originated before grasses, with many plant ancestors possessing them (see **Appendix II**). The NBS-LRR genes are highly variable [25,26], but their NBS region contains several conserved motifs that can be traced back to early land plant groups [27]. The N-terminal region of the protein contains either a TIR or CC region, the former being restricted to only dicot species [26]. The NBS contains a highly conserved Nucleotide Binding Domain shared by Apaf-1, Resistance gene products, and CED-4 (NB-ARC) [28], whereas the C-terminal LRR is a highly variable region that can bind to many different molecules [7,29]. The CNL genes have been identified in the genome of many plant species: 52 in *Arabidopsis* [26], 159 in rice [30,31], 188 in soybean [30,32], 203 in grape [33], 65 in potato [34], 94 in common bean [35], 177 in alfalfa [36], six in papaya [37], and 18 in cucumber [38]. Recent studies have shown that CNL genes are effective at resistance to the devastating Ug99 stem rust strain in wheat [39,40]. In the present study, the recently available barley genome[41] was explored to understand the diversity and evolution of CNL genes.

Cultivated barley (*Hordeum vulgare* L.) is a grass family (Poaceae) member that was domesticated approximately 10,000 years ago [42] and is now a major cereal crop [43]. Even before genomic information was available, the use of barley cultivars with resistance gene *Rpg1* in 1942 greatly reduced the loss of barley yield due to stem rust, *Puccinia graminis*, in the Midwestern United States and Canada [44,45]. Additionally, barley

cultivars containing the gene *Rph20* are resistant to barley leaf rust (pathogen: *Puccinia hordei*), which otherwise causes up to 62% crop loss [46,47]. It has been shown that the recessive barley *mlo* mutant allele confers broad-spectrum resistance to powdery mildew (pathogen: *Erysiphe graminis f. sp. hordei*) [48,49], but presence of these *mlo* mutant alleles also increases susceptibility to *Ramularia* Leaf Spot (RLS) [22]. Genes within the mildew locus A (*MLA*), some of which are CNL, also play a role in resistance to powdery mildew and were formed through duplication, inversion, and insertion over a period of greater than seven million years [50]. It has been hypothesized that many variants of *MLA* are different alleles rather than separate genes[51]. In a recent study, higher nucleotide diversity was found in wild barley samples relative to that in the cultivated samples [52].

The objectives of this research project were to identify CNL resistance genes in the barley genome and elucidate their evolutionary relationships. This *in silico* analysis aims at comparing barley CNL genes with their orthologs in rice and *Arabidopsis*. With barley and related species making up a significant portion of the staple food supply, analyses that would potentially lead to more pathogen-resistant cultivars make a significant contribution to agriculture. Wheat, another member of the same family, may contain many similar R-gene pathways and barley resistance may be conferrable to the wheat cultivars.

MATERIALS AND METHODS

CNL gene identification. Barley CNL gene identification followed methods used in *Arabidopsis* [26] and soybean [53]. Barley protein sequences were accessed through the

Ensembl Genomes database [54]. *Arabidopsis* CNL genes, as identified and classified by Meyers et al. (2003) [26], were obtained from Phytozome [55] and their orthologs in rice were obtained, as confirmed in Benson (2014) [31]. Fifty two *Arabidopsis* CNL genes were used as reference sequences to explore orthologs in the barley genome (62,236 analyzed protein sequences), by aligning the sequences in the program ClustalW[56] and constructing a Hidden Markov Model (HMM) using HMMER version 3.1b2 [57] at a stringency of 0.05. Further selection involved identification of NB-ARCs using the database Pfam [58], accessed through Interproscan [59]. Genes containing NB-ARCs were then aligned using ClustalW, integrated within the program Geneious [60]. A second HMM profile was constructed to use these barley NB-ARC-containing proteins to perform a reiterative search of the genome with a stringency of 0.001. Interproscan [59] was then used to identify the protein sequences with both an NB-ARC and a "DiseaseResist" region. MEME analysis [61], set to display the 20 most prevalent motifs, was used to identify protein sequences with P-loop, Kinase-2, and GLPL regions, the diagnostic motifs of the CNL genes.

Phylogenetic analysis. The NB-ARCs were extracted from the protein sequences identified by the MEME search. These sequences were aligned using ClustalW integrated within the program Geneious. The protein sequences were imported along with the original *Arabidopsis* genes and their orthologs in rice for phylogenic comparison. An evolutionary model for the CNL amino acid sequences was determined using a Maximum Likelihood Model Test function in the program MEGA 6.0 [62], which identified JTT+G+I as the best substitution model. This model was used to construct a Maximum Likelihood tree with 100 bootstrap replicates.

Gene structural variation, clustering, and K_s analysis. Information on location and exon size was obtained from Ensembl Genomes, which was uploaded into the program Fancygene v1.4 [63] to generate an exon map. Entire chromosome sequences were accessed through Ensembl Genomes and imported into the program Geneious. A genomic map to visualize gene clustering was generated by matching gene locations with their respective chromosomes, along with centromere locations [41]. Nucleotide intervals between genes on each chromosome were determined in order to quantify any clustering following Jupe et al. (2012) [64]. Accessions were grouped into clades according to their nesting pattern. Coding sequences were downloaded from Ensembl Genomes to estimate the nonsynonymous substitutions per nonsynonymous site (K_a) and synonymous substitutions per synonymous site (K_s) values, and K_a/K_s ratios were calculated using the program DnaSP 5.10.1 [65]. Average K_s values were used to infer relative time of duplication events.

RESULTS

Identification of CNL genes. A total of 175 CNL genes were identified in the barley genome (**Figure 2.1**, **Supplementary Figure 2.1** for the complete phylogenetic tree, and **Supplementary Table 2.1** for all identified accessions and clade information). Initial Hidden Markov Model (HMM) analysis of the 62,236 barley protein sequences resulted in 982 orthologous sequences when using the 52 *Arabidopsis* CNL reference sequences and a stringency of 0.05. Interproscan integrated into the program Geneious was used to identify 908 sequences with Nucleotide Binding Domain shared by Apaf-1,

Resistance gene products, and CED-4 regions (NB-ARCs). Using these 908 putative barley sequences, the reiterative HMM analysis against the genome at a stringency of 0.001 yielded 950 protein sequences in barley. Using Interproscan, 654 of the 950 barley genes were identified as containing both NB-ARCs and DiseaseResist regions. All splice variants were removed, yielding 233 unique NBS-LRR genes, 175 of which contained the signature motifs: P-loop, Kinase-2, and GLPL (**Supplementary Figure 2.2**).

Phylogenetic relationships. Phylogenetic relationships of barley CNL genes and their orthologs in *Arabidopsis* are shown in Fig. 1 (also in **Supplementary Figure 2.1**, and **Supplementary Table 2.1**). Among the four clades previously reported in dicot species[26,31], CNL-D is completely absent in barley. The vast majority of the barley CNL genes (168 of the 175 members) belong to the clade CNL-C. Very few members of the CNL-A (2 members) and CNL-B clades (5 members), as well as the large amount of the CNL-C genes in barley were consistent to those in rice, but diverse from *Arabidopsis* (**Figure 2.1**). The orthologs in rice and barley show a high degree of interspecific nesting with a diversified CNL-C clade with complete absence of CNL-D members. Basal support for CNL-C is weak but leaf branches with specific gene relationships are strongly supported (BS >90%). Identification of *MLA* genes using BLAST within the Ensembl Genomes database showed that MLOC_10425 and MLOC_66581 are the likely accession names for many *MLA* sequences (**Figure 2.2**).

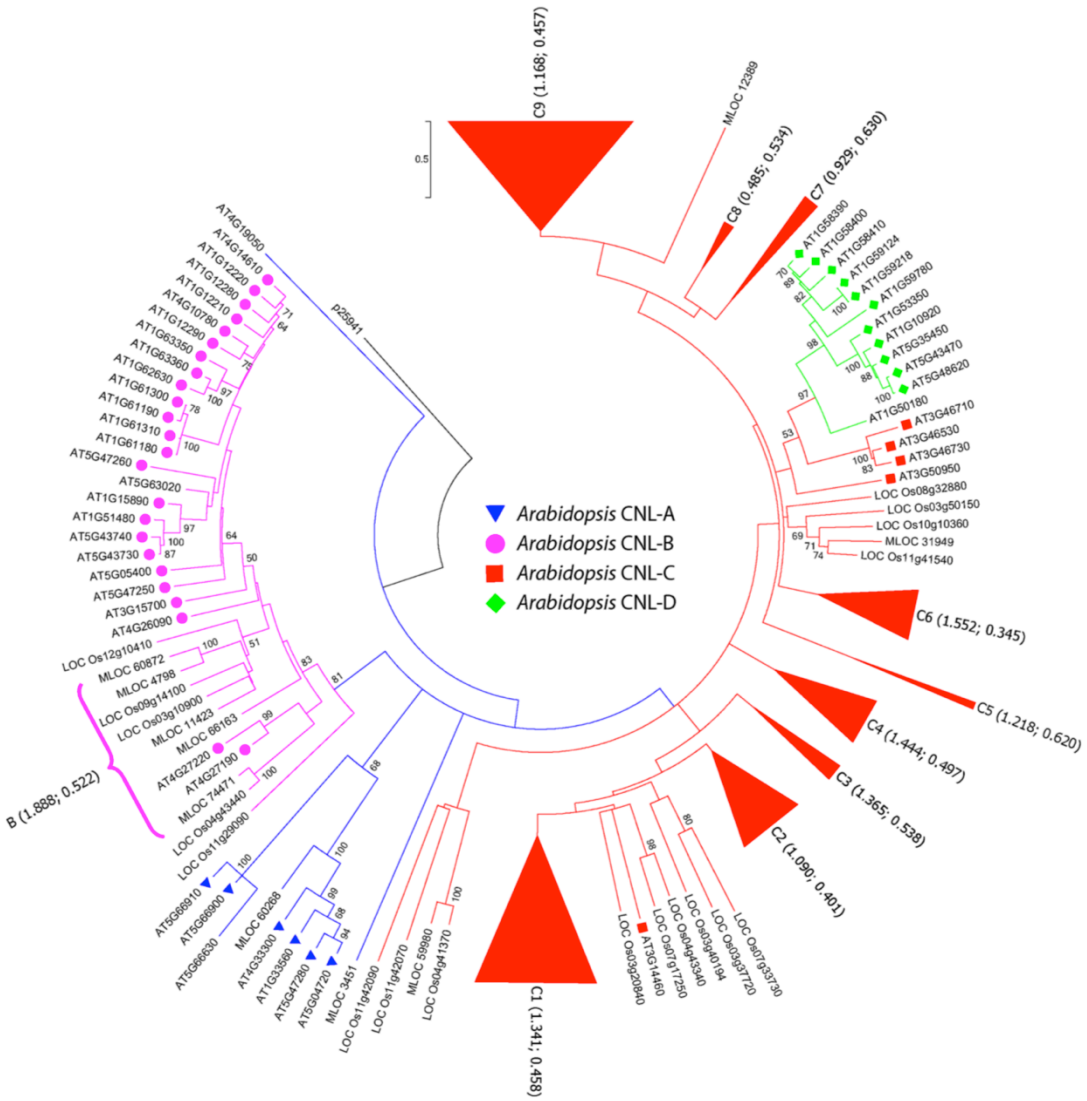


Figure 2.1 Phylogenetic analysis of the CNL genes from *H. vulgare* (MLOC), *Arabidopsis thaliana* (AT), and *Oryza sativa* (LOC). The Maximum Likelihood tree was constructed using the JTT+G+I model with 100 bootstrap replicates. *Arabidopsis* CNL-A, CNL-B, CNL-C, and CNL-D groups are represented as blue triangles, pink circles, red squares, and green diamonds, respectively. The tree was rooted using outgroup p25941 as used in *Arabidopsis*[26]. CNL-C clades were collapsed to increase readability (for the complete tree see Supplementary Figure 1) and lists of genes can be found in Supplementary Table 1. The K_s values and K_a/K_s

ratios are shown in parentheses following the clade name, first K_s and then K_a/K_s ratio. The collapsed clades contain only barley genes with the exception of clades C2 and C6, containing Arabidopsis orthologs AT3G14470 and AT3G07040, respectively.

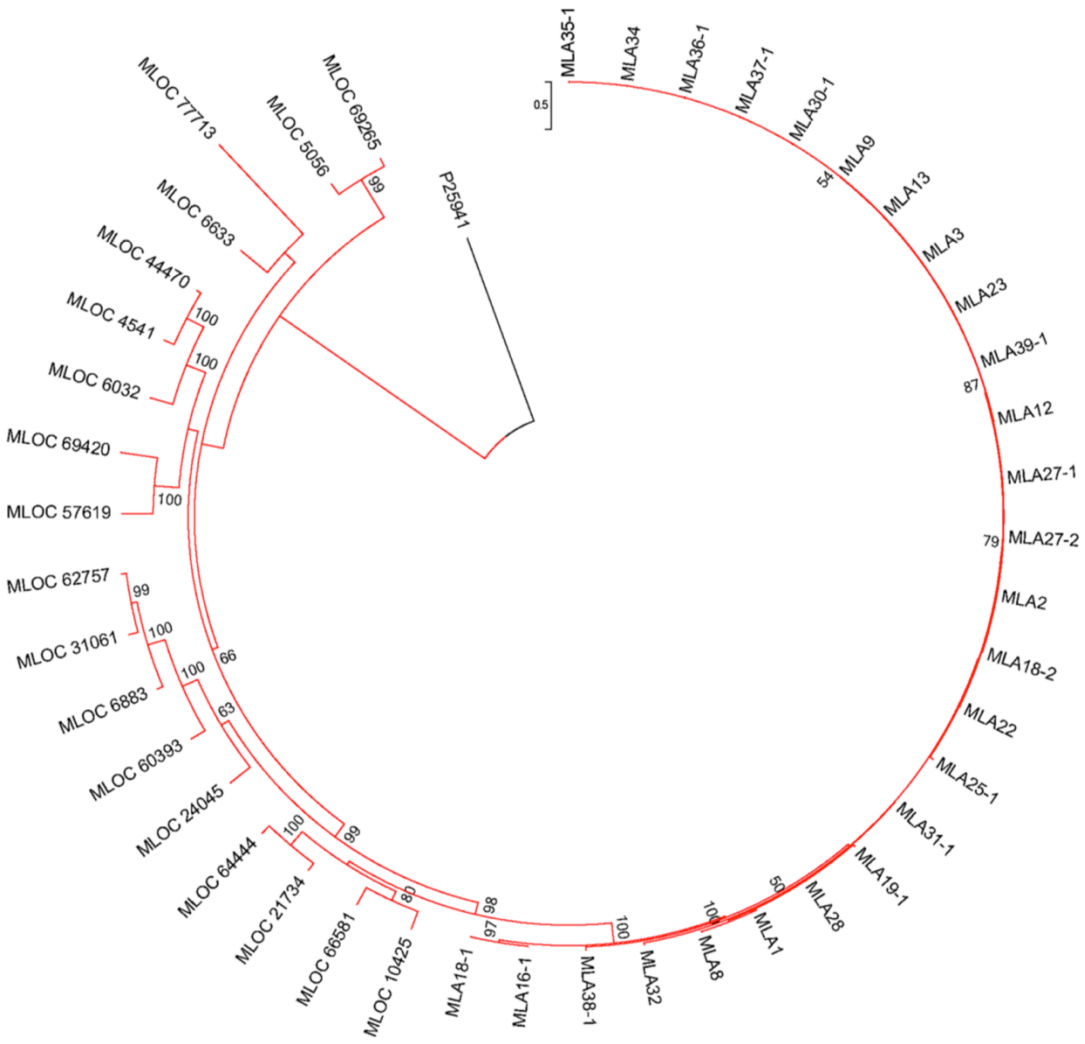


Figure 2.2 Maximum Likelihood phylogenetic analysis of *MLA* accessions and selected barley CNL-C9 gene members using the JTT+G+I model with 100 bootstrap replicates. The tree was rooted using outgroup, p25941as previously used in *Arabidopsis* [26].

MEME analysis, gene clustering, and structural variation. Conserved motifs visualized through MEME analysis show structural differences between the NB-ARC regions of various barley CNL clades (**Supplementary Figure 2.2**). The P-loop, Kinase-2, and GLPL motifs are present in all genes and RNBS A, B, and C motifs are present in 165, 172, and 151 members, respectively (**Supplementary Figure 2.2** and **Supplementary Table 2.2**). Exon-intron analysis shows that CNL genes are composed of an average of 3.34 exons, ranging from one exon in accession MLOC_6570 to twelve exons in MLOC_10066 (see **Supplementary Figure 2.3**). Of the 175 genes, 30, 46, 26, and 35 had one, two, three, and four exons, respectively; thus over 78% of the genes were found to contain one to four exons.

Gene locations on each chromosome were visualized to show CNL gene clustering (Fig. 3), which is defined as: 1) genes within a 200 Kb sliding window and 2) fewer than eight other genes between the beginning and end of the cluster. Using these criteria, 15 gene clusters were identified (**Table 2.1**). Genes tended to be located in the extra-pericentromeric regions of chromosomes (**Figure 2.3**). Each chromosome except chromosome 4H contained at least one cluster, and ten of the fifteen clusters were composed of only two genes, as shown in **Table 2.1**.

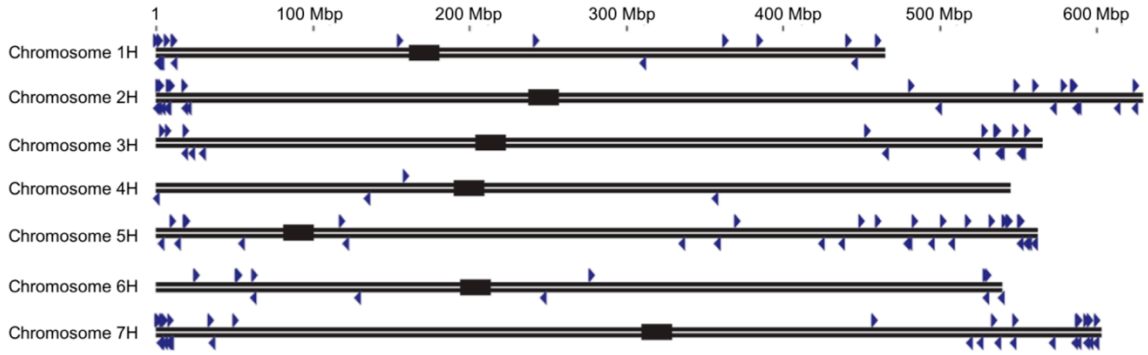


Figure 2.3 Distribution of the CNL genes on the chromosomes of barley (N=7). The black lines and blue arrows represent chromosomal length and gene location/orientation, respectively. Black rectangles indicate the centromere positions on each chromosome.

Table 2.1 CNL gene clusters in barley genome: 15 clusters containing 39 genes were identified using a sliding window of 200 Kb and eight open reading frames (ORFs). CNL clades for each gene are included in parentheses and K_s values are included for each individual cluster.

Cluster	Clustered Genes					K_s value
1_1	MLOC_66596 (C2)	MLOC_5818 (C6)	MLOC_3117 (C2)			1.435
1_2	MLOC_70559 (C1)	MLOC_73882 (C7)				1.753
1_3	MLOC_767 (C9)	MLOC_53251 (C9)	MLOC_70910 (C9)	MLOC_69663 (C6)		1.229
2_1	MLOC_44743 (C1)	MLOC_24729 (C1)				0.194
2_2	MLOC_66581 (C9)	MLOC_10425 (C9)				0.520
2_3	MLOC_4541 (C9)	MLOC_65574 (C6)				1.443
2_4	MLOC_76088 (C1)	MLOC_5583 (C1)				0.579
3_1	MLOC_56904 (C4)	MLOC_56905 (C4)				0.217
5_1	MLOC_12201 (C1)	MLOC_64708 (C9)	MLOC_64709 (C9)			0.880
6_1	MLOC_38183 (C9)	MLOC_76360 (C1)				2.622
6_2	MLOC_11605 (C1)	MLOC_10242 (C1)				0.392
6_3	MLOC_79526 (C9)	MLOC_67477 (C9)				0.293
7_1	MLOC_57007 (C9)	MLOC_78491 (C2)	MLOC_4344 (C9)	MLOC_4343 (C9)	MLOC_10643 (C1)	1.249
7_2	MLOC_11112 (C8)	MLOC_75786 (C9)	MLOC_30912 (C8)	MLOC_72805 (C6)		1.353
7_3	MLOC_6883 (C9)	MLOC_31061 (C9)				0.163

K_s values. Synonymous substitutions per synonymous site (K_s values) are often used as a proxy for inferring duplication events, so K_s values were used in inferring relative age of the CNL gene clusters (**Table 2.1**). Average K_s values were highest for CNL-B members and lowest for the CNL-C8 members (**Figure 2.1**). All average K_a/K_s ratios were less than 1, indicating a prevalence of purifying selection. Functional homologs for the identified barley genes were compiled and compared with results from the phylogenetic analysis (**Table 2.2**). Using this information, instances of genomic expansions as well as reductions were inferred.

Table 2.2 CNL orthologs of *Arabidopsis*, rice, and barley with associated pathogens.

Barley Accession	Rice Homolog	*Arabidopsis* Homolog	Synonym	Pathogen
MLOC_55575, MLOC_56324, MLOC_67526, MLOC_51950, MLOC_6570, MLOC_20874, MLOC_34944, MLOC_5818, MLOC_69663, MLOC_77773, MLOC_1192, MLOC_72805, MLOC_65574, MLOC_1818, MLOC_64033, MLOC_16581, and MLOC_56093	LOC_Os06g22460, LOC_Os06g30430, LOC_Os08g09430, LOC_Os12g31620, LOC_Os07g08890, LOC_Os08g16070, LOC_Os02g09790, LOC_Os11g35580, LOC_Os08g16120, LOC_Os11g12000, and LOC_Os11g12340	AT3G07040	*RPM1*	*Pseudomonas syringae*[66]
MLOC_31949	LOC_Os08g32880, LOC_Os03g50150, LOC_Os10g10360, and LOC_Os11g41540	AT3G50950	*ZAR1*	*Pseudomonas syringae*[67]

MLOC_74471, MLOC_66163, MLOC_11423, MLOC_60872, and MLOC_4798	LOC_Os12g10410, LOC_Os11g29090, LOC_Os09g14100, LOC_Os03g10900, and LOC_Os04g43440	AT1G12210 AT4G26090 AT1G12220 AT1G12280	*RFL1* *RPS2* *RPS5* *SUMM2*	*Pseudomonas syringae*[68,69,19,70]
MLOC_60268	-----	AT1G33560 AT4G33300 AT5G04720 AT5G47280	*ADR1* *ADR1-L1* *ADR1-L2* *ADR1-L3*	*Peronospora parasitica* and *Erysiphe cichoracearum*[71]
MLOC_31949	LOC_Os08g32880, LOC_Os03g50150, LOC_Os10g10360, and LOC_Os11g41540	AT3G46530 AT3G46710 AT3G46730	*RPP13* *RPP13*-like *RPP13*-like	*Peronospora parasitica*[72,73]
MLOC_57619, MLOC_69420, MLOC_58258, and MLOC_1443	LOC_Os07g19320, LOC_Os11g15500, and LOC_Os11g37740	-----	*Yr10*	*Puccinia striiformis*[74]
MLOC_44141	LOC_Os05g34230 and LOC_Os04g02110	-----	*Rga3*	*Magnaporthe oryzae*[75]
MLOC_66596	LOC_Os01g25740, LOC_Os01g25810, and LOC_Os03g63150	-----	*Pm3*	*Blumeria graminis*[76]
MLOC_4581	LOC_Os02g16270 and LOC_Os02g16330	-----	*Xa1*	*Xanthomonas oryzae*[77]
MLOC_67378 and MLOC_10643	LOC_Os01g57310	-----	*Rp1*	*Puccinia sorghi*[78]

DISCUSSION

Phylogenetic analysis and evidence of duplications. Phylogenetic analysis of the CNL protein sequences from barley and *Arabidopsis* showed a high level of tandem duplications within each species. Barley R-genes were nested as expected within the CNL-A, CNL-B, and CNL-C clades with their orthologs in *Arabidopsis*, concurring with the previous findings in rice [31] and *Aegilops* [79]. Fewer members of CNL-A, CNL-B, and complete absence of CNL-D were observed in barley relative to that in *Arabidopsis*.

Using comprehensive phylogeny of flowering plants [80] as a reference, it can be inferred that *Arabidopsis* has experienced a reduction in CNL-C and expansions in CNL-A, CNL-B, and CNL-D. In a recent analysis of CNL genes in soybean (*Glycine max*) [53], a similar expansion in the CNL-C clade was observed. In contrast to CNL genes in soybean, there was a sharp reduction in CNL-A and CNL-B, and absence of CNL-D, in both barley and rice, which may be common in other grass species as well. Phylogenetic analysis of CNL genes of barley with rice (a model monocot[81] with a more recent common ancestor [82]) showed more interspecific nesting patterns than with *Arabidopsis* (**Figure 2.1**). Existing differences in R-gene diversity, structure, and evolutionary rates across these species may reflect phylogenetic constraints and species-specific evolutionary history[83].

Closely related genes within the same gene cluster in the phylogenetic tree (**Figure 2.1** and **Table 2.1**) show strong evidence of gene duplication events. Despite the huge genome size (5.1 Gb) of barley, there are numerous closely located CNL genes and their clusters that diversified through tandem duplications. One of the most striking examples of tandem duplication involves MLOC_24729 and MLOC_44743 genes which are only 113 bases apart and are 69.5% identical (528 out of 760 sites). The gene accessions MLOC_19475, MLOC_58383, MLOC_44175, and MLOC_12318 are closely related and form their own clade (**Figure 2.1**), with three of these genes located within a 2.24 Mb segment of chromosome 7H, another instance of tandem duplication. The fourth gene in the same clade, MLOC_12318, is located on chromosome 2H, indicating that it resulted from segmental duplication. Similar duplication events have been reported in other plant genomes [84]. Overall variation within R-genes is attributed to duplications, recombination, and diversifying selection [25], with whole genome duplications lessening

selective pressures and allowing for diversification, as seen in the soybean genome[85]. Increased diversity of R-genes may provide barley with a selective advantage even though maintenance of R-genes during low pathogen exposure might prove very costly as suggested in literature [86]. While not residing within a technically defined cluster in barley, many genes are likely formed by gene duplication events, the origin of which could be traced to a common ancestor gene. The genes MLOC_11112, MLOC_30912, and MLOC_15443 form their own clade, with MLOC_30912 basal to the other two. MLOC_11112 and MLOC_30912 are clustered on chromosome 7H, likely formed by tandem duplication. The third gene, MLOC_15443, is approximately one Mb upstream of the other two, a possible instance of segmental duplication. Another example is a five-gene sub-clade (MLOC_66610, MLOC_66596, MLOC_19284, MLOC_68128, and MLOC_3117; BS 78%) in which all five genes are located within a 2.1 Mb section of chromosome 1H, likely to have arisen through gene duplication. It has been shown that R-genes can cluster in larger regions that do not fall within the defined criteria (i.e. with the narrow sliding window) of a cluster [87]. In *Medicago*, superclusters have been identified in which a single chromosome arm contains a large percentage of the genome's resistance genes[36]. Zhou et al. (2004)[30] suggests that duplications of diversely clustered R-genes could explain the frequent and dissimilar duplications.

K_s values have been used to infer the history of duplication events within a genome, especially when analyzing genome duplications or polyploidy [88,89]. The barley CNL-B clade has a higher average K_s value than any CNL-C subclade, suggesting recent expansion of CNL-C members in grasses (see **Figure 2.1**). While average K_a/K_s values for each CNL-C clade were <1 indicating purifying selection, 23 individual pairwise

values were >1, 15 of those being from CNL-C9. This indicates that while the majority of the identified genes are undergoing purifying selection, a few genes are undergoing positive selection. These K_s values can also give insight into the clustered genes that arise from duplications. For instance, cluster 3_1, composed of MLOC_56904 and MLOC_56905, has a very low K_s value of 0.217, indicating a recent duplication event. Since rice only has one paralog to these two sequences, LOC_Os01g05620 (Fig. 1), the duplication event likely happened after the split of rice and barley lineages. A similar case is shown by MLOC_44743 and MLOC_24729 (cluster 2_1) which have the K_s value of 0.194 and do not have a close paralog in rice, suggesting more recent evolution after rice and barley split. The same happens with cluster 7_3 (MLOC_6883 and MLOC_31061) with a low K_s value of 0.163. From this information, it can be concluded that cluster 3_1 formed first, followed by cluster 2_1, and finally 7_3.

***Arabidopsis* and rice homologs in barley**. Looking more closely at the gene duplications and expansions within the barley genome, a species-specific history of pathogen load can be inferred. *Arabidopsis* gene AT3G07040 is functionally known as *RPM1*, an NBS-LRR gene that recognizes either the AvrRpm1 or AvrB type III effectors of *Pseudomonas syringae,* conferring resistance through a hypersensitive response[66]. As shown in Fig. 1 and summarized in Table 2, barley contains nine homologs (clade CNL-C6) of *RPM1*, what can be inferred to be a large expansion. It is possible that monocots faced a heavy *P. syringae* load during their evolutionary history, perhaps both before and after barley and rice diverged, since rice contains only five *RPM1* homologs (Table 2). Another possibility is that *Arabidopsis* experienced a reduction through pseudogenization. In some other cases, barley genome contains fewer R-genes than

Arabidopsis. The *Arabidopsis ADR1* genes (AT1G33560, AT4G33300, AT5G04720, and AT5G47280) are involved in the resistance response to *Peronospora parasitica* and *Erysiphe cichoracearum* [71]. Barley genome contains only one homolog (i.e. MLOC_60268) for these four genes in *Arabidopsis*. The same occurs with the *RPP8* and *RPP13* genes where many *Arabidopsis* gene members do not have any homologs in barley. Barley and rice appear to differ in the number of *ZAR1*, *RPP13*, and *ADR1* homologs, with barley's single *ADR1* homolog not being represented in the rice genome. There are also no barley homologs for AT1G10920 (*LOV1* – CNL-D), which causes susceptibility to *Cochliobolus victoriae* [90].

The *MLA* locus in barley confers resistance to powdery mildew (*Blumeria graminis* f. sp. *hordei*). Many variants of *MLA* have been identified in this analysis (see **Figure 2.2**). Two CNL-C9 gene members, MLOC_66581 and MLOC_10425, are highly similar to many different *MLA* sequences, with MLOC_66581 being a gene that most likely responds to powdery mildew. A BLAST search using MLOC_66581 and MLOC_10425 within the Ensembl Genomes database reveals that these two genes have the highest sequence identity to all *MLA* sequences. Seeholzer et al. (2010)[91] identified two functional *MLA* genes, *MLA27* and *MLA18*, that both correspond to MLOC_66581 and MLOC_10425 accessions, respectively. As shown in Fig. 2, these genes nested close to the *MLA* sequences, along with MLOC_64444 and MLOC_21734, which would also be closely related to the *MLA* genes. Thus, the results support the previous predictions by Shen et al. (2003)[51] and Seeholzer et al. (2010)[91] that many *MLA* variant sequences are alleles rather than separate genes [51,91].

MLOC_60268 and MLOC_3451 are the only barley genes that nest with *Arabidopsis* CNL-A, with high bootstrap support. This shows that these two genes represent current CNL-A members in barley and are likely to have existed before the evolutionary split between monocot and dicot plants, between 200 and 140 million years ago [82,92]. Accession MLOC_3451 shows most homology to the Apoptotic Protease-Activating Factor 1 (*APAF1*) from *Triticum uratu*, contributor of wheat's A-genome [93]. The similarity is not partial; entire protein sequence alignment shows the sequences are 96.3% similar (1002 identical sites out of 1040). The presence of *APAF1* would be expected since hypersensitive response involves an apoptosis-like cell death to prevent the spread of a pathogen. Therefore, CNL-A members in barley are predicted to contribute in hypersensitive response.

Gene structure and genomic content. Since there is no strict correlation between CNL gene content and genome size, a reasonable prediction of barley's CNL gene content could range from a few dozen members to several hundred. Two earlier studies in barley reported 50 CNL genes [45] and 191 NBS-LRR genes [41]. While the rice and barley genomes have vastly different sizes, 420 Mb and 5.1 Gb, respectively[41,94], the genome wide CNL diversity is rather similar, 159 and 175 genes, respectively. The P-loop, Kinase-2, and GLPL motifs are highly conserved in both species[30] and the Resistance Nucleotide Binding Site (RNBS) A, B, and C motifs (**Supplementary Figure 2.2** and **Supplementary Table 2.2**) are also prevalent and conserved within CNL genes[26,30].

The CNL genes in barley showed a higher number of exons (3.34 exons per gene; **Supplementary Fig. 3**) than *Arabidopsis* and rice, with *Arabidopsis* genes generally

consisting of one exon each [26] and rice averaging 2.1 exons per gene [31]. The higher number of exons per gene in barley could enable a more variable response to the pathogens through multiple splice variation. Since many of the 982 initially identified protein sequences were variants of the same genes, it is possible that barley has used multiple splicing patterns to vary its pathogen-response proteins. It has been shown that NBS-LRR genes go through alternative splicing in *Arabidopsis* [95], and ratios of different transcripts are required for a resistance response [96].

While number of exons per gene is higher than other species, the amount of CNL gene clustering is lower in barley, where only 39 of 175 CNL genes form 15 gene clusters (**Figure 2.3** and **Table 2.1**). In *Arabidopsis*, 109 of the 149 NBS-LRR genes formed 43 clusters [26], but it was predicted that larger genomes may have a more complex distribution of CNL genes and that unclustered CNL genes are not unusual [26]. Barley genes that are highly clustered, such as those on chromosome 7H, allow for higher recombination rates and faster evolution [26,97]. Resistance genes show varying speed of evolution, with Type I genes evolving relatively faster than Type II genes [98]. The expansion of CNL-C indicates that many of the CNL genes in barley are of the Type I class, suggesting a potential expansion in all grass species. Combining the evidence of duplications and clustering with K_a/K_s ratios, the majority of barley CNL genes are currently undergoing purifying selection, which has been reported to be a common phenomenon among duplicated genes, especially in crop species [99]. The reduction in nucleotide diversity that took place during the cultivation of barley also likely impacted evolution of R-genes [52].

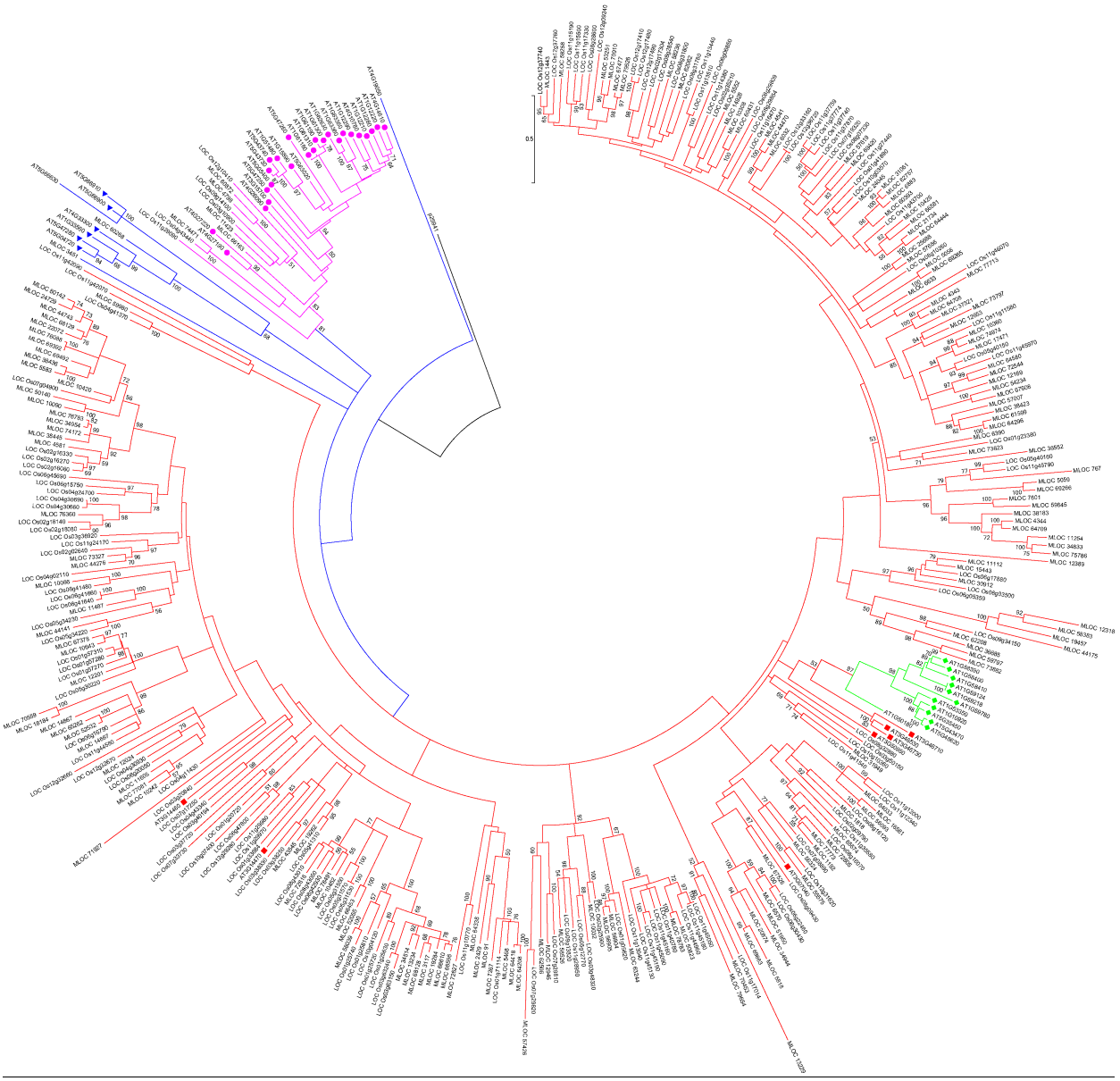
Current challenges in the development of durable resistance and future directions. Understanding of disease resistance has expanded greatly due to advances in molecular techniques and computational ability. Challenges regarding how efficiently researchers utilize genomic data to develop more durable resistance continue to exist and can be overcome through the development and utilization of transcriptomic and metabolomics data. Additional genomic annotations are also needed as some chromosomal locations could not be accessed to determine clustering, and standardization of nomenclature is necessary. Specifically in the case of barley, current proteomic information is not complete and additional data would allow us to assess functionality. This, along with expression data upon pathogen exposure, and biochemical assays of signaling pathways are major areas that require continued research. Also, cultivar-specific genome sequences would be useful to determine variation and educate breeders about how variation across cultivars is related to crop yield. This would allow for the development of barley cultivars that can better combat pathogens, and may indirectly uncover directions for developing durable resistance in wheat and other closely related species.

CONCLUSIONS

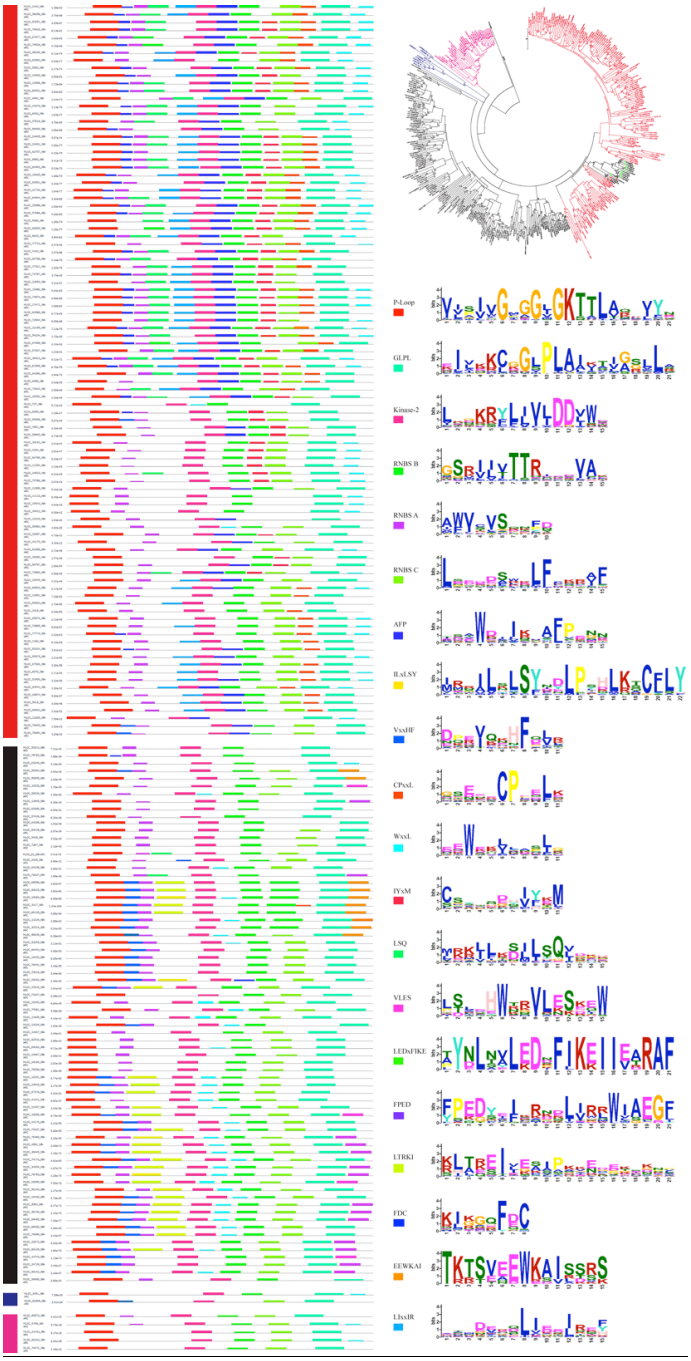
In this study findings on the diversity and evolution of CNL genes in barley have been presented. The 175 identified barley R-genes show evidence of gene duplications as well as expansions and contractions of the NBS-LRR clades. The CNL gene diversity in barley is slightly higher than in rice and more than three times that in *Arabidopsis*. Many

RPM1 homologs could be identified, indicating substantial exposure to pathogens such as *P. syringae* in barley's evolutionary history. The results also indicated that several previously identified *MLA* sequences are the allelic variants of two CNL genes (MLOC_66581 and MLOC_10425). Many splice variants and multiple exons per gene may have allowed rapid diversification of R-genes in barley, especially the members of the CNL-C clade. As expected, several gene clusters were found, especially in the extra-pericentromeric regions of chromosomes, a location that experiences high rate of recombination needed for rapid gene diversification. Further research should aim to measure expression levels of these genes upon pathogen exposure and assess if some of these CNL genes could be used in developing cultivars with durable resistance.

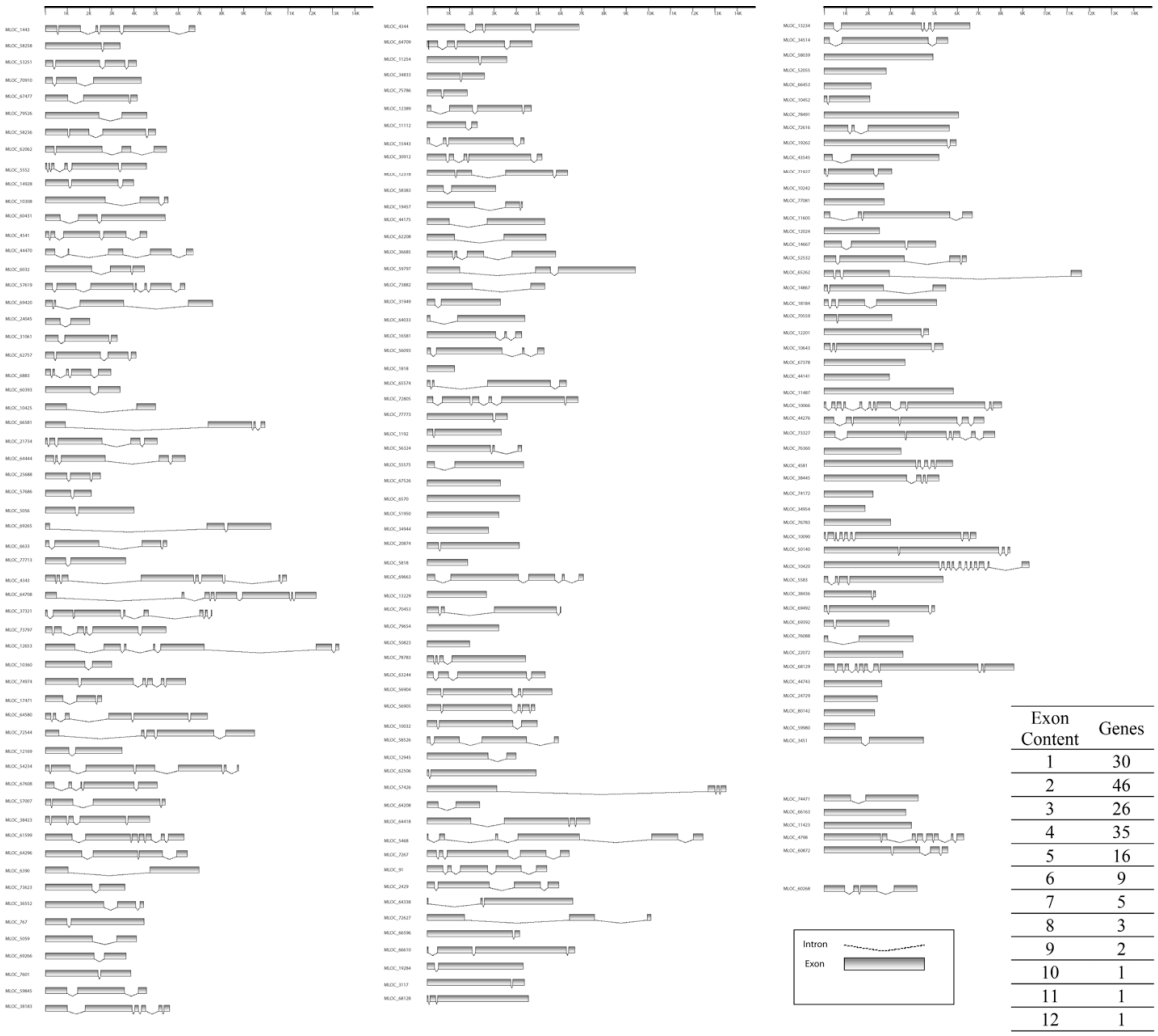
SUPPLEMENTARY DATA (SEE NEXT PAGE)



Supplementary Figure 2.1 Phylogenetic analysis of the CNL genes from *H. vulgare* (MLOC), *Arabidopsis* (AT), and *Oryza sativa* (LOC). The Maximum Likelihood tree was constructed using the JTT+G+I model with 100 bootstrap replicates. *Arabidopsis* CNL-A, CNL-B, CNL-C, and CNL-D groups are represented as blue triangles, pink circles, red squares, and green diamonds, respectively. The tree was rooted using outgroup p25941 as previously used in *Arabidopsis*[26].



Supplementary Figure 2.2 Motif structure of the 175 *H. vulgare* CNL genes based on MEME analysis. The CNL-A, B, and C clades are in blue, pink, and red respectively. The six characteristic motifs P-loop, Kinase2, GLPL, RNBS-B, RNBS-A, and RNBS-C are specifically named and the following fourteen motifs are named based upon their amino acid residues.



Exon Content	Genes
1	30
2	46
3	26
4	35
5	16
6	9
7	5
8	3
9	2
10	1
11	1
12	1

Supplementary Figure 2.3 Exon-intron variation across 175 CNL R-genes in barley. This illustration was generated using the program Fancygene 1.4 after input from Ensembl Genomes transcript information. Genes are presented by clade. Thick gray bars and dashed lines represent exons and introns, respectively. On the lower right corner is the summary information on the abundance of exons.

Supplementary Table 2.1 List of identified CNL genes and their corresponding clades.

Name	Clade	Name	Clade	Name	Clade	Name	Clade	Name	Clade
MLOC_91	C3	MLOC_10643	C1	MLOC_34833	C9	MLOC_58383	C7	MLOC_68128	C2
MLOC_767	C9	MLOC_11112	C8	MLOC_34944	C6	MLOC_58526	C4	MLOC_68129	C1
MLOC_1192	C6	MLOC_11254	C9	MLOC_34954	C1	MLOC_59797	C7	MLOC_69265	C9
MLOC_1443	C9	MLOC_11423	B	MLOC_36552	C9	MLOC_59845	C9	MLOC_69266	C9
MLOC_1818	C6	MLOC_11487	C1	MLOC_36685	C7	MLOC_59980	C	MLOC_69392	C1
MLOC_2429	C3	MLOC_11605	C1	MLOC_37321	C9	MLOC_60268	A	MLOC_69420	C9
MLOC_3117	C2	MLOC_12024	C1	MLOC_38183	C9	MLOC_60393	C9	MLOC_69492	C1
MLOC_3451	A	MLOC_12169	C9	MLOC_38423	C9	MLOC_60431	C9	MLOC_69663	C6
MLOC_4343	C9	MLOC_12201	C1	MLOC_38436	C1	MLOC_60872	B	MLOC_70453	C5
MLOC_4344	C9	MLOC_12318	C7	MLOC_38445	C1	MLOC_61599	C9	MLOC_70559	C1
MLOC_4541	C9	MLOC_12389	C	MLOC_43545	C2	MLOC_62062	C9	MLOC_70910	C9
MLOC_4581	C1	MLOC_12653	C9	MLOC_44141	C1	MLOC_62208	C7	MLOC_71927	C
MLOC_4798	B	MLOC_12945	C4	MLOC_44175	C7	MLOC_62506	C4	MLOC_72544	C9
MLOC_5056	C9	MLOC_13229	C5	MLOC_44276	C1	MLOC_62757	C9	MLOC_72616	C2
MLOC_5059	C9	MLOC_13234	C2	MLOC_44470	C9	MLOC_63244	C4	MLOC_72627	C3
MLOC_5468	C3	MLOC_14667	C1	MLOC_44743	C1	MLOC_64033	C6	MLOC_72805	C6
MLOC_5552	C9	MLOC_14867	C1	MLOC_50140	C1	MLOC_64208	C3	MLOC_73327	C1
MLOC_5583	C1	MLOC_14928	C9	MLOC_50823	C4	MLOC_64296	C9	MLOC_73623	C9
MLOC_5818	C6	MLOC_15443	C8	MLOC_51950	C6	MLOC_64338	C3	MLOC_73797	C9
MLOC_6032	C9	MLOC_16581	C6	MLOC_52055	C2	MLOC_64418	C3	MLOC_73882	C7
MLOC_6390	C9	MLOC_17471	C9	MLOC_52532	C1	MLOC_64444	C9	MLOC_74172	C1
MLOC_6570	C6	MLOC_18184	C1	MLOC_53251	C9	MLOC_64580	C9	MLOC_74471	B
MLOC_6633	C9	MLOC_19262	C2	MLOC_54234	C9	MLOC_64708	C9	MLOC_74974	C9
MLOC_6883	C9	MLOC_19284	C2	MLOC_55575	C6	MLOC_64709	C9	MLOC_75786	C9
MLOC_7267	C3	MLOC_19457	C7	MLOC_56093	C6	MLOC_65262	C1	MLOC_76088	C1
MLOC_7601	C9	MLOC_20874	C6	MLOC_56324	C6	MLOC_65574	C6	MLOC_76360	C1
MLOC_10032	C4	MLOC_21734	C9	MLOC_56904	C4	MLOC_66163	B	MLOC_76783	C1
MLOC_10066	C1	MLOC_22072	C1	MLOC_56905	C4	MLOC_66453	C2	MLOC_77081	C1
MLOC_10090	C1	MLOC_24045	C9	MLOC_57007	C9	MLOC_66581	C9	MLOC_77713	C9
MLOC_10242	C1	MLOC_24729	C1	MLOC_57426	C4	MLOC_66596	C2	MLOC_77773	C6
MLOC_10308	C9	MLOC_25688	C9	MLOC_57619	C9	MLOC_66610	C2	MLOC_78491	C2
MLOC_10360	C9	MLOC_30912	C8	MLOC_57686	C9	MLOC_67378	C1	MLOC_78783	C4
MLOC_10420	C1	MLOC_31061	C9	MLOC_58039	C2	MLOC_67477	C9	MLOC_79526	C9
MLOC_10425	C9	MLOC_31949	C	MLOC_58236	C9	MLOC_67526	C6	MLOC_79654	C5
MLOC_10452	C2	MLOC_34514	C2	MLOC_58258	C9	MLOC_67608	C9	MLOC_80142	C1

Supplementary Table 2.2 Sequence information with the conserved motifs as identified by MEME analysis.

Accession	P-loop	Kinase-2	GLPL	RNBS-A	RNBS-B	RNBS-C	Protein Length
MLOC_91	IIPIVGPTGLGKTSLAQLVFN	LNKYSCLIVLDSLWG	EIVRRCGGTPLVANFLGSVVN	IWVHVSMGNV	ESKIVVTTHSYKVAE	LSEDDCLNIFSQRAM	1022
MLOC_767	VVAIVGFGGSGKTTLAKAVYD	YHQDRYLIVVDDIKD	KLLKKCGGLPLALYSVACQLS		NGRIIVTTATQSVAN		726
MLOC_1192	LITLTGMGGLGKTTLAANVYR	LEELKYLIILDDVWT	EILSKCKGLPLAIVSIGSLLH	AWVSISQTYS	GSRLIITTRQADVAA	LPEDKAWDLFCKKSF	923
MLOC_1443	VVSIVGSGGLGKTTLANQVYQ	LKHKRYFIVIDDIWG	EILKKCGGLPLAIITVASLLA	AFVSVSQKPD	SSRILTTTRIITVAK	LSATHSKSLFFKRAF	952
MLOC_1818	IIALWGMGGIGKSTLVSNVFR	LKKKRYLIILDDVWT	LIVGRCDGLPLALVAIGSSLS	AWVSVSQSYK	GSRIIITTRSEEVAS	LEEEDAWRLFCRKAF	404
MLOC_2429	IIPIVGLVGLGKTTLARLIFH	LHDKRCLIVLDGLCS	QIVKRCDGIPSLAHFLGSVVH	IWIDLNRKFD	WIKVLVTTSSDITAE	LSEDDCWTIFFEKAF	1305
MLOC_3117	VVPIVGMGGLGKTTLAQLIYN	LSGKRYLLVLDDVWN	EIVKRCRGSPLAATALGSVLR	LWVCVSDTFD	GSAVLTTTRDKQVAE		1288
MLOC_3451	VVLIVGLSGIGKSCLARQIAS	LVGKSMLILLDDVWE	DLLDSCGHHPLTVAVLGKALR		DCRYLVTTRDEAVYE		1042
MLOC_4343	VVSVLGSGCMGKTTLANTLYS	LQDKRYLVVIDDLLD	EIVNMCGGLPLATISIASHLA	AFIRMSKKPD	GSRIIITTQIEDVAL	LDDDHSRKLFFNRLF	551
MLOC_4344	VISIVGFDGIGKTLLARNVYD	LGSKRFFIVIDDMRT	EALKKCDGLPLALVTTARYLQ	AWVSTTEQGG	SGRVLVTTAIQSVAN	LTDEHSRQLFLKEAF	1087
MLOC_4541	VVSIVGLGGLGKTTLANSVYE	LGDKRYLIVIDDIWD	KILDRCGGVPLAIITIGSLLS	AFVSVSLNPT	GSRVIVTTRILDVAQ	LSVADSRKLFYQRIY	928
MLOC_4581	VLPIVGPGGIGKTTLTQHLYN	LKSKRFLIVLDDVWK	KISQKLKCSPLAAKTVARLLK	VWVCVSTDFD	GSMVIVTTRFPKIEQ	LDPDEFFEFFLACVF	1264
MLOC_4798	VIGIWGPGGVGKTHLLNNIKN	LKKRSFLVLLDDLWD	ELVMEMKGLPLALITIAKAMY	VQVTASRGCS	RRKVVITTRSRKVCG	LQEEESWQLFQDKVG	1170
MLOC_5056	MVSIVGCRGIGKTTLAVEVYQ	LKTRRYFIVIEDIWD	DILKKCGGIPLAIITIASLLA		GSRILVTTCIISIAS	LSDAEARKLFFSRIF	992
MLOC_5059	VVSILGPRGIGKTTLARVVYD	QANTRYLIVIDDIQT	SILKKCEGLPLGIVNIANYLK	AWIVASEHRG	NGRIIVTTTMHSIAN	LSKEDSKALLLKIVL	896
MLOC_5468	IVPIVGLGGLGKSTLAESVFV	LADRRYLIVLDDLWE	KIAEKCGGVPLVAIALGQVMS	AWVHVSKEFD	GSRVIVTTRNQSVVN	LSNDDCWEVMKQTAF	1487
MLOC_5552	VAAVVGSGGMGKTTLAHQVYT	LRDKRYLIVLDDIWA	EILKKCDGVPLAILSVASYLA	AFVTVSQNAS	GSRVITTTRNEDVAK	LNDLDSRRLFYKRVF	546
MLOC_5583	ILPIVGPGGIGKTTLVQHIYH	LKSKRFLFILDDIWK	KIAEKLKGSPLAAKTVGRLLR	VWICVSQNFI	GNVILVTTRFPAIAE	LEHKEFRNFFRACIF	1493
MLOC_5818	VTVVWGMPGVGKTTLVDHVYN	LQGKRYILVLDDVWA	KFIAKCQGLPIAIACISRLLS	AWVTVSEGYR	TSRVIITSRKQAVLA	LEEHYSWLLFCKGAF	567
MLOC_6032	IVSIVGFGGLGKTTLANAVYQ	LLNKRYFIVIDDIWN	NILKRCGGVPLAIMTTASMLA	AFVSVSLNPN	GSRIITTTRIFDVAK	LSPIDSRKLFHQRIF	920
MLOC_6390	VIFIVGFGGLGKTTLAMEVCR	LQNNRYLILIDDVWS	VILKKCGGLPLAIISIASLLS	AKVSVSQTFG	GSRIIVTTRTNTLAK	LDRVESKQLFMSKTF	964
MLOC_6570	IITVWGMGGVGKTTLVAHVYN	LEKKRYVLVLDDVWS	KFVEKCNGLPIAIVCIGRLLS	AWITVSHSYE	LGRIIFTSRIYEVAL	LQNHYAWDLFCKEAF	897
MLOC_6633	VVSIVGVGGLGKTTLANVVYQ	LRDKRYLIIVDDIWD	EIIRKCDGVPLAIITIASLLS	AFVTVSLKPD	SSRVITTTRIAGVAA	LDHDDSKKLFYERIF	983
MLOC_6883	TISIVGVGGLGKTTLTKAVFE	LVDKRYLVVIDDIWE	EILKKCGGVPLAIITIASLLV	AFVPVGQNPD	HSRVITTTRNVSVSQ	LSDDDSLILFHRRIF	465
MLOC_7267	IIPIVGLGGIGKTTLAKSVYA	LAGKKYLIVLDDFWE	EIAGKCGGVPLVANAFGQVMS	AWVHVSEKFD	GSKIIVTTRNQSVVK	LPTKDCWELIKIRAF	1166
MLOC_7601	VIAVVGFCGLGKTALAAEVYN	LVKKRYFIVIDDIRT	EVLKKCDGQPLALVTIGEFLQ	AWVYAALRSP	SSRILVTTTIQSVAN	LDKICSKQLFTKKAC	989
MLOC_10032	VFAIVGAVGIGKTTLAREIFT	ALSKRFLIVLDDLDS	KIAEKCKGHPLAIKVIAGVLR	VWVKMSKDLS	RGRILVTTRDEEVAT	MDADNSWALLREQVF	1093
MLOC_10066	VIFIIGVGGVGKTSLARLAFG	IAGKKFLLILDDVWY	QICRRLNGLPLAAKIIGRLLN	MWVSVSDTYD	GSKILVTARADKACS	LDKDDYWLLFRNHAF	1176
MLOC_10090	VLPIVGPGGVGKTTFTQHIFE	LRSKRFLLVLDDVWK	KIVENLKGFPLAAKTVGRLLR	IWICVSLNFT	GNMVIVTTRIPGVAN	LGPQDIMSFFEACVF	1569
MLOC_10242	IVSVVGHGGMGKTALAQSICE	LKSMNYLLVLDDVWE	QIAKKLGGCPLVTKVTSGHLR	IWVTVSTTFD	GSKILLTTRMQSVSN	LEEDKNLELFNHHVF	689
MLOC_10308	VVSIVGFGGLGKTTLANEVYR	LEEKSYLVVIDDLWD	EILKKCGGMPLAIISISSLLA	AFVPVSQKPN	GSRLIVTTRIKNVAE	LTAEDSRKLFFDRIF	905
MLOC_10360	VVCLVGSGGIGKTTLAKELYR	LQDKRYLIVIDDVWA	NIIRKCDGLPLAVVTVAGLLV	AFVRTSRKPD	CSGILMTTEVDDVAL	LGHDDSSKLFFSTAF	872
MLOC_10420	VIPIVGPGGIGKTTLTQYIYN	LKNKRFLFVLDDIWK	KIVEKLKGSPLAAKTVGRLLR	VWICVSLDFS	GDTILVTTRFLEVAE	LEPKEYWTLFLACVF	1550
MLOC_10425	IISVVGFGGMGKTTLVKTVYE	LKNKRYLIVIDDIWD	DVLKKCGGVPLAIITIASLLA	AFVPVGRKAE	GSRLITTTRIVSVCE	LSDEDSERLFYKRIF	561
MLOC_10452	VLPIFGMGGLGKTTLAKMVYN	IGRKRFLLVLDDVWN	KIVNKCGGLPLALKTMGGLLS	MWHCVSDNFD	GSVILVTCRSKQVAS	LGEEDSWELFSKKAF	599
MLOC_10643	VLAIVGAGGMGKSTLAQLVYN	QESGKFLLVLDDVWF	KIAKRLGQSPLAAKIVGSQLK	MWVSISRKLD	GSRVLVTTRQNSFPA	MEDAQFLALFKHHAF	1289
MLOC_11112	VAFVVGMGGLGKTTLARKIFE	LEDKRYFIVLDDVWS	KIVNKCGCLPLAILTIGGMLA	AWITVSQTFS	GSRIMITTRNIGLHQ		612
MLOC_11254	VIIIVGFHGMGKTLLANHVYR	IGKKRFFIVIDDLRT	EALKKCDGLPLALVTTARFLQ	AWVLGGRRTA	TGRFLVTTAIQPVAN	LTNQDSRTLFFKEAF	931
MLOC_11423	IVGIWGMGGVGKTTLLRKINN	LRRKKFLLLLDDLWN	EVAEECGGLPLALVTIGRAMS		KQKVVMATRYESVCG	LDPEKAWKLFKQKAT	966
MLOC_11487	VHTIVGVGGIGKTTLAQLIYN	LRCKRFLLVLDDVWY	QIVQKLKGSPLALKVVGGHLN	MWVSVSNNFD	GSKILVTTRTELVAK	LGRDDSWSLFRRCAF	1257
MLOC_11605	ILSVVGHGGMGKTALAQRICE	NLHTKFLLVLDDVWE	KIAKKLGGCPMLTKVVSGHLQ	IWVHVSDSFD	GSKILLTTRMQSVAD	LEEGENLELFNHHVF	1293
MLOC_12024	VFSLVGLGGIGKTSLAQDICS	LISKTFLLVLDDVWN	KMVKKLSGSPLAAKVLGGLLN	IWVCASHDFD	ESKILLTTRMQSVAV		651
MLOC_12169	VACIDGPAGVGKTTLAKQLYY	LQDKRYFIVIDDLWE	GIIRKCGGLPLATICISGLLA	AFVRASRKSD	YSRIITTAETDSVAL	LGSHASAELFFSIVF	1022
MLOC_12201	GVAIVGLGGMGKSTLAQHVYN	QKSNRYLLVLDDVWF	KLAKRIGTSPLAAKTVGSQLS	MWVCISRRLD	GSKILITSRSNILPA	MEDNDILALFKDHAF	1283
MLOC_12318	VMSVWGIAGVGKSALVMNMFC	LKKHRCLVVIDGLQS	ELTSLCGGLPKVIVEVAVVFA	GWVDVSHPFN			1207
MLOC_12389	VLCIVGLAGIGKTTIARSLYQ	LEKYSYLLVLDDVWS	AHWEMCMSLPLAIVIMAGYVA	AMVTVSHNSD	GSRVIVTSRFQAVAS	LSENKSEELFMDIMA	999
MLOC_12653	VVSIVGVGGIGKTTLAKQLWL	LQNKRYFLIIDDLWA	EIVRRCAGLPLAIISVASVLA	AFVRTAKKPD	CSRIVTTTEIEDVAR	LNVSNSKELFTNGVL	1545
MLOC_12945	VVTIVGCCGIGKTTLAQKVYD	LEGKRFLLVLDDISS	TIIQRCNKVPLAIKIIGGVLA	IWVDGSKDTT	GSRILITTRDQRVAA		1004
MLOC_13229	VISVMGKSGVGKTKLVKERYE	LENKKYLLVIDAEVS	FIFDVTGGLPLAVVLLSGLLR	AWVTCAPNLN			732
MLOC_13234	VFPIVGVGGLGKTTLAQLIYN	LSGRRYLLVLDDVWN	EIVKRCRGSPLAATALGSVLR	IWVCVSDSFD	GSVVLTTTRDKGVAE		744
MLOC_14667	IIGIHGIGGSGKSTLSQLVCA	LPGKRFLLVLDDVWY	DIAKKLKGSPLAAKIVGGQLR	MWVHVSQDFS	GSKILVTSRTKDALL	LDEDVFLNLFKHYAF	985
MLOC_14867	VIGIHGVSGSGKSTLAQFVCA	LDGKRFLLVLDDVWS	EIAEKLKGSPLAARTVGGNLR	MWVHVSQNFS	GSKILVTTRSKHALT	VDETAFFELFMHYAL	960
MLOC_14928	VVSIVGFGGLGKTTLANQVYH	LREKRYFFVIDDIWD	EILKKCGGLPLAIVSIASLLV	VFVSISQRPN	GSKVITTTRIEMVAK	LDDQNSRKLFFSRVG	981
MLOC_15443	VICVVGMGGLGKTTLARKVYE	LEDMRYFIVLDDLWT	KMVKKCGGLPLAVLTIGGMLA	AWITVSQSFS	DSRILITTRDVGLAA		968
MLOC_16581	VITVSGMGGLGKSTLVSNVYE	LENRRYLIVLDDVWE	SIVEKCHGLPLAIVTIGSMLS	AWIVVSQIYT	GSRIIITTRKDHVAG	LSNPDAFKLFCRRAF	903
MLOC_17471	VVSVVGSGGLGKTTLVKELYR	LQHKRYLIVIDDAWA	SIIRKCAGLPLAIVAVASLLS	AFVRSSRKPD	CSGILITTEINDVAL	LIDDDSSKLFFSTAY	557
MLOC_18184	VIGIYGVAGSGKTTFAGYIRD	LCGKRFFLILDDLWV	EIAKKLHRSPIAAVTVAGRLV	MCIHVSETFS	GSKILVTARTKEAAG	LDEDQYLKMFMHYAL	1163
MLOC_19262	ILPIVGMGGLGKTTLTQLVYN	LKGKRFLLVLDDVWN	EIVKKLKGLPLAAKAIGSLLC	VWLCVSENFD	GSRIVVTTRNKNVGK	LSDNDCWFLFRSYAF	1131
MLOC_19284	VVPIVGMGGLGKTTLTQLIYN	VSGQRYLLVLDDVWD	EIVKRCSGSPLAATALGSVLY	LWVCVSDTFD	GSVVLTTTRDKRVAE		1294
MLOC_19457	VMSVWGIAQVGKSALVKKLFY	LKKDWCLIVIDDLRS	DLISKCGGLPKVISAIAGLLA	YWVDVSHPFN	KSVVIVITTDPTIAE	LEATHALDLFKKKVH	846
MLOC_20874	ISTVWGMGGVGKTTLVAHVYN	LRGKKYLLVLDDVWG	RFVDKCNGLPIAIACIGRLLS	AWITVSKAYQ	TSRFVITSRIHDVAL	LEEHHSWELFCKEAF	921
MLOC_21734	IVSVVGFGGLGKTTLVRAVYD	LEDKRYLVIIDDIWD	DILKKCGGVPLAILTIASLLS	AFVPVGRNPD	GSRLITTTRIFTVSE	LSSNDSARLFYKRIF	1010
MLOC_22072	VLPVVGPGGMGKTTLIQHIYH	LKSKRFLLVLDDIWQ	KIMEKLKGSPLAAKTVGRLLS	IWICVSFNFN	GSMVLVTTRFQAIAQ	LKSEEFRKLFLAFIF	1081
MLOC_24045	TVSIVGFGGLGKTTLARAAYD	LHEKRYLIIIDDIWD	DILKKCGGIPLAIITIASLLA	AFVSVSQNPN	GSRLITTTRVISVSE	LSNDVSRTLFCKRVF	489
MLOC_24729	VLPVIGPGGMGKTTLIQHIYN	LKHQRFLLVLDDIWQ	EIMGKLKGSPLAAKTVGRLLR	IWICVSFNFN	GSVILVTTRQKEIAD	LERGEFRKLFLVYVF	754
MLOC_25688	VVAIVGGGGLGKTTLAMEVYR	LTRKRYLIVIDDVWK	RILRKCGGLPLAIITIASLLA	ASVSVSRTLD	GSRIIATTRITKVAK	LNVVDSQRLFFKRIF	406
MLOC_30912	VVCVVGMGGLGKTTLTRKVYE	LLEKRYFVVLDDLWN	KIVKKSGLLPLAILTIGGMLA	AWITVSQSFV	GSRIMVTTRDVGLAE		998
MLOC_31061	TISIVGVGGLGKTTLTKAVYE	LVDKRYLIVIDDIWE	DILKKCGGVPLAIITIASLLA	AFVPVGQKPD	QSRVITTTRNVSVSE	LSDEDSQILFHKRIF	797
MLOC_31949	VVSLVGAGGAGKSTIARKVYN	LKDKRYLVVLDDVWR	ILALKCNGLPLAIVVMGGFLS	AWICISQQFT	GSRIVLTTRNSAVSN	LNNEESVELFNRKAF	880
MLOC_34514	VVPIVGVGGLGKTTLAQLIYN	VSGRRYLLVLDDVWN	EIVKRCCGSPLAATALGSVLR	IWVCVSDSFD	GSVVVTTTRDEGVAK		1271
MLOC_34833	VITIVGFHGMGKTLLANHVYK	IGTKRFFIVIDDLRK	AVLKKCDGLPLALVTTARFLQ	AWVPPRKLGE	SGRFLVTTTIQRVAN	LADQHSRLLFFKEAF	625
MLOC_34944	IAAIWGMGGVGKTTLVHHLYK	LQGSKYLIVLDDVWH	QFLGKCGGLPIAIACVGRLLS	VWITVSSSCQ	TGRFVITTRIQEVAL	LQRRYAWQLFCNEAF	744
MLOC_34954	VLPIVGPGGIGKTTFTTHLYN	LKSKRFLIVLDDIWK	KIANKLKGSPLAARTVGRLLY	VWVCVSTDFD	GSMLLVTTRFPKVAD	LESNDFITFFEACIF	585
MLOC_36552	VVSILGMEGLGKTTLGNEIYS	LATKRYFILVDGLWS	NMFEMCGGMPLALVLAAGLLA	AFVTLGRRPS	GSRVLTTTCISAVAQ	LNEEMSARLYLNLIR	1009
MLOC_36685	VISVWGMGGLGKTTLVRSIYK	LKMRSCFVVLDDISS	LTLQKCDGLPLAISTIGGFLA		AGRIIVTTRQKNIAK	LKHAAALDLFIKKVF	1237
MLOC_37321	VLSVLGSACLGKTTLARVLYN	LQDKRYLIIIDDLWA	EIVEICDGLPLATVSIASVLA	AFIRVSKKPD	GSRIIITTQMEDVAL	LDDDHSRKLFFNSLF	433
MLOC_38183	VISLVGFGGIGKTTLLAKHVYN	LGTKRFFIVIDDMRR	EVIRKCDGLPLALITTARLLH	AWVCAAEKGS	NSRVIVTTATQTIAI	LAEQHSKQLFCEKAS	1164
MLOC_38423	VIPIIGCAGVGKTTAARTLYH	LQGKGYFIVLDDLWT	EIIRKCEGLPLAIVSIASLLA	AFVSVSPNPD	RSRVITTTQVHDVAL	LDEHESRELFFRMVF	1168
MLOC_38436	ILPIVGPGGIGKTTLIQHIYN	LNSKSLLLIDDIWK	RIADKLKGSPLAAKTVGRLLR	IWICVSQNFI	GNVILVTTRFPAVAE	LEHGEFRKFFRACIF	701
MLOC_38445	VLPIVGPGGIGKTTFVRHLYN	LKSKRFLIVLDDIWK	RIANKLKGSPLAAKTVGRLLQ	VWVCVSTDFD	GSVLLVTTRFPKLAE	LEPNDFLTFFEACVF	1146
MLOC_43545	VIPVVGMGGLGKTTLMQMVYN	LRGKRYLLVLDDVWN	QIVKKLKGLPLASKALGSLLF	IWIYVSECFD	GSKIVVTTRNDNVGR	LSDDDSWSVFKNHAF	1183
MLOC_44141	VAAIIGLGGMGKTALARVLLH	VSSKRFLLVLDNVWN	RLVHKLKGLPLAAKVVGGMLK	MWVCPAAAYH	GSKIMVTTRKKIVAT	LAFDDIWSLFTRIAF	908
MLOC_44175	VTSVWGIAGIGKSYLTRRFFY	MQQHRCFVVIDGLRS	LITPMCGGLPQVIAAISKYCR	GWVDVPHPFD	QGYIIVVTNEESIAR	LNDNEALHLFNKISR	1012
MLOC_44276	VLPVVGTGGIGKTTLARYVYH	IRNKRFLLVLDDMWE	QIAKALKGCPLAAQSVGALLS	MWICVSTNFD	GCMVLATTRTKSVAK	LDEKDFWLFFKACAF	950
MLOC_44470	VVSIVGFGGLGKTTLANAVYQ	LRNKRYLIVIDDIWD	KILQRCGGVPLAIITIGSLLS	AFISVSLNPN	ESRVITTTRILDVAQ	LSVVDSRKLFHQRIY	865
MLOC_44743	VLPVVGPGGMGKTTLIQHIYH	LKYKRFLLVLDDIWQ	KIMEKLKGSPLAAKTVGRLLR	IWICVSFNFN	GSMILVTTRFQEIAK	LEPEEFRKLFLVYVF	855
MLOC_50140	VVPLVGPGGIGKTTLTQHIYG	LKSKRLLLVLDDMWT	KIVTNLKGFPLAAKTVGRLLR	VWVCVSLDFN	GNVIIVTTRIPEVAS	LGYKDIMSFFEACVF	530
MLOC_50823	KLAIVGTGGVGKTTLAQKIFN	IEGKSFFLVLDDVWH	EIVRKCGRLPLSIKVTSSALA	AWICVSRDYN	SGVILITTRDDQIAV		570
MLOC_51950	VISVWGMGGVGKTTLVTHVYN	LHSRRYLILLDDLWD	KIVESCDGLPIAIVCIGRLLS	AFITVSQHCR	SSKVVLTSRIHDVAS	LESQHSWDLFCKEAF	919
MLOC_52055	LLPIIGMGGIGKTTLAQLVHN	LGKKRYLLVLDDVWN	SIVLKCKGLPLAIKTIAALLR	IWVCVSDKFV	GSAIIVTSRSNQVAS	LNDYQSWKLFRRNAF	839
MLOC_52532	VIGIHGIAGSGKSTLAQYVYE	LRGKQILLVLDDVWY	DIAKKLKGSPLAARTVGSRLR	MWIHVSQKFD	GSKILVTSRTEAALV	LDEKVFLEMFMHYAL	991
MLOC_53251	VLSIVGFGGLGKTTLANEVYL	LQDKRYLVIIDDIWS	TILKKCGGLPLAIISISGLLA	AFVSVSQKPD	SSRIITTTRILEVAR	LSDLLSERLFCNRIF	931
MLOC_54234	VASIVGTSGVGKTTLARQVYH	LQDKTYIIVIDNLQA	EIIKRSGGLPLAIIILASLLV	AFVQASPKQD	CSRILTTTEVDAIAQ	LSEDESRELFLGTVF	1146
MLOC_55575	IIMIQGMGGSGKSTLARSIYR	LQNRKYLIVLDDLWD	KILEKCQGLPLAIAAIGSLLS	AWISTSRNRQ	GSIVIITTRNEAVAS	LSEEESWDLFSKKAF	900
MLOC_56093	VITVSGMGGLGKSTLVSNVYE	LKNRKYLIVLDDVWE	SIVDRCHGLPLAIVTIGGMLS	AWIVVSQVYT	GSRIIITTRKDHVAG	LSKSDAFDLFCRRAF	915
MLOC_56324	IMSILGMGGLGKTTIASSVYK	LQDKKYLIVLDDVWD	KIVKKCQGLPLAIVAIGSLLS	VWVTLSQNYL	GSRVLVTTRKKDVAS	LPYAESWHLFCQKAF	923
MLOC_56904	VFGIQGMGGIGKTTLAQKIYN	IKGKSVLLVLDDVWK	EIVKKCDGLPLAIKVVAGVLS	IWLCISQSYT	NFHVLVTTRDLDVLA	MNYHDGLELLMKKSF	989
MLOC_56905	VLGIQGMGGIGKTTLAQKIYN	IKGKSVFLVLDDVWK	DIVKKCDGLPLAIKVVAGVLS	IWLCITQSYT	NFHILVTTRDLCVLS	MNYSDGLELLMKKSF	990
MLOC_57007	VISIFGFAGVGKTTLIRTFYR	LQCKRYLIVIDDLWA	EIIRKCGGLPLAIENMASMLA	AFVRVSRNPD	CSRIIVTTEIKDVAL	LNDHQSRKLFLSRVF	1237
MLOC_57426	LFAIVGTIGVGKSTLARKIYH	GSTRNLLLVIDNVSK	SIVQKCSGVPMAIRTIGWNLR	LWVNISNDSR	GCKVIVTTRDENVAR		1051
MLOC_57619	ILSIFGFGGLGKTTLAKAVYD	LHNKRYLIVIDDIWD	KILHKCGGIPLAIITIASLLA	AFVPVGRNPS	GSRIITTTRILEVAT	LSLDLSKDLFHRRLS	844
MLOC_57686	VVAIVGGGGLGKTTLAMEAYR	LTGKRYLLVIDDVWK	RILRKCGGLPLAIITIAKLLS	ASVSVSRTLD	GSRIIVTTRIANVAK	LNYIDSRRLFFKRIF	469
MLOC_58039	VLPIVGMGGLGKTTFAQIIYN	VSGRRYLLVLDDVWN	EIVDRCCGSPLAAKALGSVLS	KWVCVLDDFD	GSAVLMTTRDERVAQ	MDTSDLLAIFEKRAF	1337
MLOC_58236	VVSMVGCGGLGKTTLARRVYD	LQDKRYFIIVDDVWD	KILQKCGGLPLAIIAISGLLI	AFVSVSRRPH	GSRIITTTRIHEVAK	LSAIDSERLFFKLVF	894
MLOC_58258	VVSVVGAGGLGKTTLANQVYR	LQDKRYLIVIDDIWS	EIVKKCGGLPLAIITMASLLT	SFVSLSQNPD	GSRIIVTTRIGTVAK	LSEDHSKMLFFRRIF	892
MLOC_58383	VMSVWGIAGVGKSALLKNLLC	LEQGRCLVVIDGLQS	ELISKCGGLPHVIVELAGLLG	AWVDVSYPFN	TNGIIVITTEASVAK	LQAEDAFALFEKEVS	830
MLOC_58526	VVAIVGQGGIGKSTLAKKILA	LSKNKFLLVMDDVWD	KIVKKCDGLPLAIKVMGGLLS	IWLSVTQQFT	GSRVLVTSRKAGVVR	LNDEDAWCLLKKQLP	1042
MLOC_59797	VISVWGMGGLGKTTLVRDVYQ	LEGKKYLIVLDDLSS	LILKKCDGLPLAIVTIGGFLA	ACVTIMHPFN	SSRIIVTTRSENIAR	LGYEDGLNLFMEKVF	1214
MLOC_59845	VISIVGCWGLGKTALAADVYR	LANKRYFIVIDDIQT	AILNKCYGQPLALITMGEFLK	AWVCAALKSP	SSRVVVTTTIQSVAS	LDEKCSKRLFSEKAC	958
MLOC_59980	ILPITGMAGIGKTTLAQLVFW	LSGRKLFLVLDDAWD	MIAETCAGVPSLLLSASNKLK	IWVSVPRNFN	GSKMIVTSRIPDVVK		407
MLOC_60268	VVSICGMGGSGKTTLAMEIYK	RDRGPVLVILDDVWS	QVAAECRGLPLALKVIGASLR			LGEEEALSVFCSAAF	847
MLOC_60393	TISIVGFGGLGKTTLAKAVYD	VDKSRYLIIVDDIWE	DILKKCGGVPLAIITVASLLV	AFVSVAQNPD	CSRIVTTTRKISVSE	LCDGDSQILFYRRIF	932
MLOC_60431	IVSVVGFGGLGKTTLANQVYH	LKNKRYLVIIDDLWD	EILKKCAGLPLAIISISSLLA	AFLSVSQKPN	CSRLIVTTRIRTVAG	LSEKDSRKLFFGRIF	900
MLOC_60872	MIGIWGPGGVGKTHLLKNINN	MKTKSFLVLLDDLWD	ELMKELKGLPLALITIGKAMY	LFVTASRGCS	NRKVVLTTRLREVCG	LQEHEAWHLFEENIG	1355
MLOC_61599	VIPIIGSAGIGKTMVARAIYH	LQGKSYFIVVDDLWT	RIIRKCSGLPLSIVNIASLLL	AFVRVSRNPD	WSRILTTTQVEDVAL	LNYGESRKLFFNNVF	1368
MLOC_62062	VLSIVGSGGLGKTTLARQVYN	LQDKRYFIIIDDIWD	DILKKCGGLPLAINAISSLLT	AFVSVSRNPN	GSAIITTTRNIDVAK	LGVDHSKKLFFKRIF	954
MLOC_62208	VVSVWGMGGMGKSSLVRMLYN	LEEKSCLVVVDDVSS	RILERCRGLPLAIATIGGLLA	AWVTVPHPLE	GSRVVVTTRRDDVAR	LQDEEARKLFCQKVY	947
MLOC_62506	VVAIVGPDGVGKTKLARMVYE	VANKRFLLVVDDVWY	RLVDKCGCIPLAIKAVAGVLR	SWVRLSRGYT	GGKVLVTARRGTTAR		1460
MLOC_62757	TISIVGVGGLGKTTLTKAIYE	LVDKRYLIVIDDIWE	DILRKCGGVPLAIITIASLLV	AFVPVGQNPD	QSRVITTTRSVSVSE	LSDEDSQILFHRRIF	928
MLOC_63244	VFGIEGMGGIGKTTLAQKIYW	IKRKTVFLVLDDVCL	AIVKKCDGLPLAIKAVAGVLS	IWLCISETYT	NACVLITTRDHKVLQ		989
MLOC_64033	VITVSGMGGLGKTTLVTNVYE	LKDRKCLIVLDDVWN	SIVDRCQGLPLAILSIGGLLS	AWMVVSQTYT	ASRVIITTRNNHVAA	LGNAHAFELFCRRVF	910
MLOC_64208	IIPIVGPGGQGKTTLAESVIA	LAGRRYLIVLDDLWE	QIVDKCGGLPLVVNALGQVMS	AWVHVSKEFD	GSRIIVTTRNQRVVD		479
MLOC_64296	VISISGSAGVGKTTVARTIYH	LQGKSYFIVVDDLWA	RIIRKCSGLPLSIVSIASLLL	AFVRVSRNPD	CSRILTTTQVEDVAL	LNDGEARKLFFNNVF	1293
MLOC_64338	VIPIVGFGGLGKTTLAQLVFN	FTGKKYLLVLDDVWS	GIVQKCGGVPLAAKALGSMLR	IWVSMSVDFS	GSKIMVTTRSRKVGM	LSDDDCWELFKGKAF	1286
MLOC_64418	IIPIVGLGGLGKTTLVQSVIG	LAGRRYLIVLDDLWE	QIVDKCGGLPLVVNALGQVMS	AWVHVSKEFD	GSRIIVTTRNQRVVD	LSRDDCWKMMKQRAL	1319
MLOC_64444	IVSVVGFGGLGKTTLVRAVFD	LQNKRYLVVIDDIWD	DILKKCGGVPLAILTIASLLT	AFVPVGQNPD	GSRLITTTRIFSVSI	LSDSDSTRLFCKRIF	967
MLOC_64580	MVAITGHAGVGKTTLAKQLSR	LQNKRYFIVIDDLWE	IIITRCAGLPLATICVSGLLT	AFVRASRKPD	CSRIIATTEIEEVAL	LGGEDSGKLLLNLVF	1344
MLOC_64708	VVSVLGSGCLGKTTLANVLYD	LQDKRYLIVIDDLWD	EIVDICGGLPLATINIASHLA	AFIRVSKKPD	GSRIIITTQIEDVAL	LNIGHSRELFFNRLF	1280
MLOC_64709	VISIVGFDGIGKTLLARHVYH	LGTKRFFIVIDDMQT	EALKKCDGLPLALVTTARYLQ	AWVCTAEQRG	SGRILVTTAIQSVAN	LPNEHSRQLFFKEVF	1110
MLOC_65262	VVGIHGIAGSGKSTLAQCVYH	LCGKRILLVLDDVWY	DISKKLKRSPLAARTVGSRLR	MWIHVSQKFD	GSIILVTSRTEAALV	LDDEVFLEMFMHYAL	900
MLOC_65574	VVVLSGMGGLGKTALAANVYK	FLAQKYLIVLDDVWK	QIVAKCKGLPLAIVSVGSLLF	AWISVSQTYS	GSSIVITTRSADVAR	LPDDKAQELFHRKAF	660
MLOC_66163	VVGIYGMAGIGKTALLNKFNN	LTKMNFVLLLDDLWE	GLAMKCGGLPLALITVGRAMA		KSKIIVATRIEDVCD	LEPQPAWDLFCEKVG	908
MLOC_66453	VLPIVGMGGIGKTTLAQLVHN	LVNKRCLLVLDDVWN	NIVPNCKGLPLAIKSISTLLC	LWVCVSDKFV	GSAIIVTTRSDKVAS	LNEYDSWELFQRNTF	700
MLOC_66581	IVSVVGFGGLGKTTLVKTVYD	LEEKRYLIVIDDIWE	GILKKCGGVPLAIITIASVLA	AFVSVGQKAD	GSRLITTTRIVDVSK	LSDYDSQRLFHKRIF	943
MLOC_66596	IVPIVGMGGLGKTTLAQLIYN	VSGQRYLLVLDDIWN	EIVKRCRGSPLAATALGSVLR	LWVCVSDTFD	GSAVLTTTRDKRVAE		1267
MLOC_66610	VVPIIGMGGLGKTTLAQLIYN	VSGRRYLLVLDDVWD	EIVKRCCGSPLAATALGSVLR	LWVCVSDTFD	GSAVLTTTRDKRVAK		1382
MLOC_67378	GLAIVGVGGMGKSTLAQLVYN	QESGKFLLVLDDVWF	KIAKRLGQSPLVAKVVGSQLK	MWVSISRKLD	GSKVLVTSRRDRFPS	MEDAEFLALFKHHAF	1177
MLOC_67477	VLSIVGFGGLGKTTLAKEVHR	LQDKRYLIVIDDIWS	KILKKCGGLPLAIISISSLLA	AFVSVSQKPN	SSRIIATTRIVDVAR	LSDLHSKRLFFKRIF	951
MLOC_67526	LVAVCGMGGVGKTTLVTSVYK	LAKKRYLLLLDDVWD	KILDRCCGLPLAIVSVGNLLA	AWVSVSKNFT	GSKIIITTRSQNVAS	LPKQEAWSLFCNTTF	925
MLOC_67608	VASIVGPGGYGKTTLAKQVYR	LQDKTYLIVIQDLWA	EIITASGGLPLAIIILASILA	AFVRTSPKAD	CSRILTTTQVEVIAQ	LTEYESRELFLSTAF	1125
MLOC_68128	VVPVVAMGGLGKTTLAQLIYN	VNGQRYILVLDDVWN	EIVKRCRGSPLAATALGSVLR	IWVCVSDTFD	GSAVLTTTRDKRVAE		1371
MLOC_68129	VLPVVGPGGIGKTTLMQHIYH	LKSKRFLLVLDDIWQ	KIMEKLKGSPLAAKTVGRLLS	FWICVSLNFN	GSMILVTTRFQAIAE	LEPEEFRKLFLAFIF	1471
MLOC_69265	LVSIVGSRGIGKTTLAMEVYQ	LKTRRYFIVIEDIWD	DILKKCGGIPLAIIIIASLLA		GSRILVTTCIVSIAS	LSDAEARKLFFRRIF	954
MLOC_69266	VISILGPRGIGKTTLARAIYD	QVKSRYLIVIDDVQT	SILRKCEGLPLGIVNIANYLK	AWVVASEHRG	NGRIIVTTSMQSIAN	LSDDDSKALLFKIVL	903
MLOC_69392	VLPLVGPGGIGKTTLTQYIHN	LKSKRLLLILDDMWK	KIIEKLKGSPLAAKTVGRLLR	VWTCVSLDFN	GSMILVTTRFQALAE	LEPQEFRKLFLAYIF	779
MLOC_69420	ILSIFGFGGLGKTTLAKAVYD	LTNKRYLIIIDDIWD	KILQKCGGIPLAIITIASLLA	AFYSVGQNPN	GSRVITTTRIFEVAK	LSPGRSKELFCMRLS	1033
MLOC_69492	VLPIVGPGGIGKTTLIQHIYH	LISKRFLLVLDDIWK	KIMEKLKGSPLAAKTVGTLLG	IWVCVSHSFN	GSVILLTTRFPAIAK	LESEEFRKLFLAFVF	1468
MLOC_69663	VTTVWGMPGVGKTTLVAHVYN	LQGKKYILVLDDVWT	KFIAKCQGLPIAIACIGRLLS	AWVTVSKSCH	TGRFVITSRKHEVSL	LQEHHSWLLFCKGAF	930
MLOC_70453	VISILGECGIGKKTLAKKLYN	MQNRRYLVVLDGLLE	KVHNISSGLPLAIIVLAGVLR	AWVCLPPHIR	GSRILLTTRLSVKEI	LETIHGQQLFCRRVF	1001
MLOC_70559	VIGIYGIAGSGKTTFAGYIQD	LRGKRFFLILDDLWV	EIAKKLHRSPIAAVTVAGRLG	ICIHVSETFT	GSKILVTARTREAAG	LAEDQYIKMFMHYAL	830
MLOC_70910	VLSIVGFGGLGKTTLANEVSR	LKEKRYLVIIDDIWS	AILKKCGGLPLAIISISGLLA	AFVSVSQKPD	SSRIVATTRILEVAS	LSDPHSERLFFRRIF	929
MLOC_71927	LLPIVGDIGVGKTTLVQHVCD	FRRKRFLMVFEDVDM	AIARKLNGSFFGAKIVGGVLK		GSKIILTTNIRHVAA	LPHPEYWFFFKAHAF	378
MLOC_72544	VVAIVGPAGVGKTTLAKQVYH	LQNKRYVVVIDDLWE	SIITKCAGLPLATIFVAGLLT	AFVRASRKPD	CSRIITTTESEGVAQ	LGREDSGKLLYNLIF	1021
MLOC_72616	VLPVFGMGGLGKTTLAKMVYN	IGWKRFLLVLDDVWN	RIVSKCKGDPLALSAMGGLMS	MWHCASDSFE	GSVIIVTTRSPRVAS	LTEEHSWELFSKKAF	1240
MLOC_72627	VLAIHGIGGLGKTTLAKMIYN	FTNKKILIILDDLWA	AIAAKCGGVALAAQSIGSILE	VWIYVSEKFD	KVVVIATTRNEEIAR		940
MLOC_72805	VIALHGMGGLGKTALAANVYK	LEQKKYLIILDDVWT	KIVSRCKGLPLVIVLVGSLLR	AWVSISQTYS	GSRLIFTTREHHVAT	LPEDKAWDLFCKKTF	915
MLOC_73327	VLPVVGIGGVGKTTLARFVCK	IRNKRFLLVLDDMWE	QIVKALKGCPLAAQSVGALLN	MWVCVSTDFN	GCMILAATRMDSVAK	LNEEEFWLLFKACAF	1491
MLOC_73623	VVSIVGMAGSGKTTLAREVYR	LQDKRYLVMIDDLWS	RIMRKCGGLPLAIVSVGGLLA	ALVSVGRSSD	GSRIITTTRIEAVAK	LDEANAETLFKQRTF	873
MLOC_73797	VVPIVGVGGIGKTTLAQKLWS	LRDKRYFIIIDDLWA	QIKRSCGGLPLAIIVAASLLT	AFVRTAQKPD	GSRIVTITETMEVAL	LGEDDSEKLLLQRIL	1224
MLOC_73882	VISVWGMGGLGKTTLVRDVYQ	LAGHKYLIVLDDLSS	LILKKCSGLPLAIVTIGGFLS	ACVTVLRPFD	SSRIIVTTRVKGIAI	LGHSSALDLFTEKVF	839
MLOC_74172	VLPIVGPGGIGKTTFLQHLYN	LKSKRFLIVLDDIWK	KIADKLKGSPLAAKTVGKLLH	VWVCVSTDFD	GSMLLVTTRFPKVAD	LESNAFFTFFEACIF	680
MLOC_74471	IIGVWGQGGVGKTTLLHVFNN	LARKRFVILLDDVRK	AIARSCGGLPLALNVIGTAVA		RSKLILTSRYQEVCF	LGNDASWELFLSKLS	920
MLOC_74974	VVTVVGSGGIGKSTLVKELYR	LQDKRYLIVIDDVWA	NIIRKCAGFPLAVVTIAGLLV	AFVRTSRKPN	YSGILITTEIDDVAL	LCHDDSRKLFFRTAF	1377
MLOC_75786	VITIVGFHGMGKTLLADHVYK	IGDKRFFIVIDGLRK	EALKKCDGLPLALVTTARFLQ	AWVPPAKLGV	RGRFLVTTAIQRVAN	LANQHSKLLFFKEIF	486
MLOC_76088	VLPLVGPGGIGKTTLTQYIHN	LKTKRLLLILDDMWK	KIIEKLKGSPLAAKTVGRLLR	VWTCVSVDFN	GSMILVTTRFQALAE	LEPQEFRKLFLSYIF	790
MLOC_76360	VLPIVGIGGVGKTTIAQHIYQ	VCQEKLLIVLDDLWD	GILPKLKGSPLAAKTLGRMLR	IWLSVSDNFD	GSMMLVTTRSQKVAD	LQDDVLWNIFKLCMF	1100
MLOC_76783	VLPIVGPGGIGKTTFTTHLYN	LKSKRFLIVLDDIWK	KIADKLKGSPLAAKTVGRLLH	VWVCVSTDFD	GSMLLVTTRFPKVAD	LEPNDFITFFEACIF	869
MLOC_77081	IVSLVGHGGMGKTTLAQSICE	MERLSFLLVLDDVWE	QIARKLGGCPLVTKVASGHLH	VWVTVSTTFD	RSKILLTTRMRSVAS	LKEDENLELFNHHVY	669
MLOC_77713	IVSIVGSGGLGKTALANLVFR	LQSERYLIIVDDLWR	EILAICSGIPSVIITIAELLN	AFVSVSSMPD	GSRVIVTTRVNDIAK	LDHLNSKSLFLRRCF	930
MLOC_77773	VIALLGMGGLGKTTLAANIYR	LEQKKYLIILDDVWT	EIVSKCKGLPLIIVSVGSLLR	AWVSISQTYS	GSRLIITTREGDVAA	LPEDKACDLFHKKAF	918
MLOC_78491	VLPIFGMGGLGKTTLAKMVYN	IGQNRFMLVLDDVWN	RIINKCRGLPLALKTMGGLLS	LWHCVSDNFD	GSVIVVTTRSQKVAS	LNEQDSWQLFAQKAY	1169
MLOC_78783	KLAIVGTGGVGKTTLAQKIYN	IEGKSFFLVLDDLWH	EIVRKCGHLPLAIKVTSSALA	AWICVSRDYN	TEVILVTTRDDQIAL		1045
MLOC_79526	VLSIVGFGGLGKTTLAREVYR	LQDKRYLIVIDDIWS	KILKKCGGLPLAIISIASLLA	AFVSVSQKPN	SNRIIATTRVVDVAK	LNDLHSKRLFFKRIF	964
MLOC_79654	LVSIVGESSIGKTTLARKVYQ	MTGRRYLVVVDGSIA	DVFRITRGLPLSIVILAGVLR	TWTVLPPNSR	GSRVVLVTDAAGLEV	LSAENTYELFRRRVF	784
MLOC_80142	VLPVIGPGGMGKTTLIQHIYN	LKHKRFLLVLDDIWQ	EIMGKLKCSPLAAKTVGRLLS	IWICVSFSFN	GSMILVTTRQKEIAD	LEPGEFKKLFLVYIF	679

ACKNOWLEDGEMENTS

Support for this dissertation work came from the USDA-NIFA hatch Projects to M. Nepal (SD00H469-13 and SD00H659-18), South Dakota Agriculture Experiment Station, and Department of Biology & Microbiology at South Dakota State University.

REFERENCES

1. Hammond-Kosack KE, Parker JE. Deciphering plant–pathogen communication: fresh perspectives for molecular resistance breeding. *Current Opinion in Biotechnology.* 2003;14(2):177-193.
2. Jones JD, Dangl JL. The plant immune system. *Nature.* 2006;444(7117):323-329.
3. Hammond-Kosack KE, Jones J. Resistance gene-dependent plant defense responses. *The Plant Cell.* 1996;8(10):1773.
4. Michelmore RW, Meyers BC. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Research.* 1998;8(11):1113-1130.
5. Rivas S. Nuclear dynamics during plant innate immunity. *Plant physiology.* 2012;158(1):87-94.
6. Boller T, Felix G. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annual review of plant biology.* 2009;60:379-406.
7. Michelmore RW, Christopoulou M, Caldwell KS. Impacts of resistance gene genetics, function, and evolution on a durable future. *Annual review of phytopathology.* 2013;51:291-319.

8. Lam E, Kato N, Lawton M. Programmed cell death, mitochondria and the plant hypersensitive response. *Nature.* 2001;411(6839):848-853.

9. Lamb C, Dixon RA. The oxidative burst in plant disease resistance. *Annual review of plant biology.* 1997;48(1):251-275.

10. Hématy K, Cherk C, Somerville S. Host–pathogen warfare at the plant cell wall. *Current opinion in plant biology.* 2009;12(4):406-413.

11. Bolton MD. Primary metabolism and plant defense-fuel for the fire. *Molecular Plant-Microbe Interactions.* 2009;22(5):487-497.

12. Zhang S, Klessig DF. MAPK cascades in plant defense signaling. *Trends in plant science.* 2001;6(11):520-527.

13. Delledonne M, Xia Y, Dixon RA, Lamb C. Nitric oxide functions as a signal in plant disease resistance. *Nature.* 1998;394(6693):585.

14. Lu H. Dissection of salicylic acid-mediated defense signaling networks. *Plant Signal Behav.* 2009;4(8):713-717.

15. Mohr TJ, Mammarella ND, Hoff T, Woffenden BJ, Jelesko JG, McDowell JM. The Arabidopsis downy mildew resistance gene RPP8 is induced by pathogens and salicylic acid and is regulated by W box cis elements. *Molecular plant-microbe interactions.* 2010;23(10):1303-1315.

16. Shivaprasad PV, Chen H-M, Patel K, Bond DM, Santos BA, Baulcombe DC. A microRNA superfamily regulates nucleotide binding site–leucine-rich repeats and other mRNAs. *The Plant Cell.* 2012;24(3):859-874.

17. Flor HH. Current status of the gene-for-gene concept. *Annual review of phytopathology.* 1971;9(1):275-296.

18. Van Der Biezen EA, Jones JD. Plant disease-resistance proteins and the gene-for-gene concept. *Trends in biochemical sciences.* 1998;23(12):454-456.

19. Shao F, Golstein C, Ade J, Stoutemyer M, Dixon JE, Innes RW. Cleavage of Arabidopsis PBS1 by a bacterial type III effector. *Science.* 2003;301(5637):1230-1233.

20. van der Hoorn RA, Kamoun S. From guard to decoy: a new model for perception of plant pathogen effectors. *The Plant Cell.* 2008;20(8):2009-2017.

21. Bergelson J, Kreitman M, Stahl EA, Tian D. Evolutionary dynamics of plant R-genes. *Science.* 2001;292(5525):2281-2285.

22. McGrann GR, Stavrinides A, Russell J, et al. A trade off between mlo resistance to powdery mildew and increased susceptibility of barley to a newly important disease, Ramularia leaf spot. *Journal of experimental botany.* 2014:ert452.

23. Lin X, Zhang Y, Kuang H, Chen J. Frequent loss of lineages and deficient duplications accounted for low copy number of disease resistance genes in Cucurbitaceae. *BMC genomics.* 2013;14(1):335.

24. Gururani MA, Venkatesh J, Upadhyaya CP, Nookaraju A, Pandey SK, Park SW. Plant disease resistance genes: current status and future directions. *Physiological and Molecular Plant Pathology.* 2012;78:51-65.

25. Meyers BC, Kaushik S, Nandety RS. Evolving disease resistance genes. *Current opinion in plant biology.* 2005;8(2):129-134.

26. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of NBS-LRR–encoding genes in Arabidopsis. *The Plant Cell Online.* 2003;15(4):809-834.

27. Yue JX, Meyers BC, Chen JQ, Tian D, Yang S. Tracing the origin and evolutionary history of plant nucleotide‐binding site‐leucine‐rich repeat (NBS‐LRR) genes. *New Phytologist.* 2012;193(4):1049-1063.

28. Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide‐binding superfamily. *The Plant Journal.* 1999;20(3):317-332.

29. Takken FL, Goverse A. How to build a pathogen detector: structural basis of NB-LRR function. *Current opinion in plant biology.* 2012;15(4):375-384.

30. Zhou T, Wang Y, Chen J-Q, et al. Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Molecular Genetics and Genomics.* 2004;271(4):402-415.

31. Benson BV. *Disease Resistance Genes and their Evolutionary History in Six Plant Species*: Biology and Microbiology, South Dakota State University; 2014.

32. Nepal MP, Benson BV. CNL Disease Resistance Genes in Soybean and Their Evolutionary Divergence. *Evol Bioinform.* 2015;11:49-63.

33. Jaillon O, Aury J-M, Noel B, et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *nature.* 2007;449(7161):463-467.

34. Lozano R, Ponce O, Ramirez M, Mostajo N, Orjeda G. Genome-wide identification and mapping of NBS-encoding resistance genes in Solanum tuberosum group phureja. *PLoS One.* 2012;7(4):e34775.

35. Schmutz J, McClean PE, Mamidi S, et al. A reference genome for common bean and genome-wide analysis of dual domestications. *Nature genetics.* 2014;46(7):707-713.

36. Ameline-Torregrosa C, Wang B-B, O'Bleness MS, et al. Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant Medicago truncatula. *Plant physiology.* 2008;146(1):5-21.

37. Porter BW, Paidi M, Ming R, Alam M, Nishijima WT, Zhu YJ. Genome-wide analysis of Carica papaya reveals a small NBS resistance gene family. *Molecular Genetics and Genomics.* 2009;281(6):609-626.

38. Wan H, Yuan W, Bo K, Shen J, Pang X, Chen J. Genome-wide analysis of NBS-encoding disease resistance genes in Cucumis sativus and phylogenetic study of NBS-encoding genes in Cucurbitaceae crops. *BMC genomics.* 2013;14(1):109.

39. Periyannan S, Moore J, Ayliffe M, et al. The gene Sr33, an ortholog of barley Mla genes, encodes resistance to wheat stem rust race Ug99. *Science.* 2013;341(6147):786-788.

40. Saintenac C, Zhang W, Salcedo A, et al. Identification of wheat gene Sr35 that confers resistance to Ug99 stem rust race group. *Science.* 2013;341(6147):783-786.

41. Mayer KFX, Waugh R, Langridge P, et al. A physical, genetic and functional sequence assembly of the barley genome. *Nature.* 2012;491(7426):711-716.

42. Richter TE, Ronald PC. The evolution of disease resistance genes. *Plant Molecular Evolution*: Springer; 2000:195-204.

43. Dineley M. The use of spent grain as animal feed in the Neolithic. *Animals in the Neolithic of Britain and Europe Oxford: Oxbow Books.* 2006.

44. Brueggeman R, Rostoks N, Kudrna D, et al. The barley stem rust-resistance gene Rpg1 is a novel disease-resistance gene with homology to receptor kinases. *Proceedings of the National Academy of Sciences.* 2002;99(14):9328-9333.

45. Chełkowski J, Tyrka M, Sobkiewicz A. Resistance genes in barley (Hordeum vulgare L.) and their identification with molecular markers. *Journal of Applied Genetics.* 2003;44(3):291-309.

46. Hickey L, Lawson W, Platz G, et al. Mapping Rph20: a gene conferring adult plant resistance to Puccinia hordei in barley. *Theoretical and applied genetics.* 2011;123(1):55-68.

47. Hickey LT, Lawson W, Platz GJ, Dieters M, Franckowiak J, Van Deynze A. Origin of leaf rust adult plant resistance gene Rph20 in barley. *Genome.* 2012;55(5):396-399.

48. Büschges R, Hollricher K, Panstruga R, et al. The Barley *Mlo* Gene: A Novel Control Element of Plant Pathogen Resistance. *Cell.* 1997;88(5):695-705.

49. Piffanelli P, Zhou F, Casais C, et al. The barley MLO modulator of defense and cell death is responsive to biotic and abiotic stress stimuli. *Plant Physiology.* 2002;129(3):1076-1085.

50. Wei F, Wing RA, Wise RP. Genome dynamics and evolution of the Mla (powdery mildew) resistance locus in barley. *The Plant Cell.* 2002;14(8):1903-1917.

51. Shen Q-H, Zhou F, Bieri S, Haizel T, Shirasu K, Schulze-Lefert P. Recognition specificity and RAR1/SGT1 dependence in barley Mla disease resistance genes to the powdery mildew fungus. *The Plant Cell Online.* 2003;15(3):732-744.

52. Fu Y-B, Somers D. Population-based resequencing analysis of wild and cultivated barley revealed weak domestication signal of selection and bottleneck in the Rrs2 scald resistance gene region. *Genome.* 2012;55(2):93-104.

53. Nepal MP, Benson BV. CNL Disease Resistance Genes in Soybean and Their Evolutionary Divergence. *Evolutionary bioinformatics online.* 2015;11:49.

54. Kersey PJ, Allen JE, Christensen M, et al. Ensembl Genomes 2013: scaling up access to genome-wide data. *Nucleic acids research.* 2014;42(D1):D546-D552.

55. Goodstein DM, Shu S, Howson R, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic acids research.* 2012;40(D1):D1178-D1186.

56. Larkin MA, Blackshields G, Brown N, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23(21):2947-2948.

57. Finn RD, Clements J, Arndt W, et al. HMMER web server: 2015 update. *Nucleic acids research.* 2015:gkv397.

58. Finn RD, Bateman A, Clements J, et al. Pfam: the protein families database. *Nucleic acids research.* 2013:gkt1223.

59. Jones P, Binns D, Chang H-Y, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014:btu031.

60. Kearse M, Moir R, Wilson A, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012;28(12):1647-1649.

61. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in bipolymers. Department of Computer Science and Engineering, University of California, San Diego; 1994.

62. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution.* 2011;28(10):2731-2739.

63. Rambaldi D, Ciccarelli FD. FancyGene: dynamic visualization of gene structures and protein domain architectures on genomic loci. *Bioinformatics.* 2009;25(17):2281-2282.

64. Jupe F, Pritchard L, Etherington GJ, et al. Identification and localisation of the NB-LRR gene family within the potato genome. *BMC genomics.* 2012;13(1):75.

65. Rozas J. DNA sequence polymorphism analysis using DnaSP. *Bioinformatics for DNA sequence analysis*: Springer; 2009:337-350.

66. Mackey D, Holt BF, Wiig A, Dangl JL. RIN4 interacts with Pseudomonas syringae type III effector molecules and is required for RPM1-mediated resistance in Arabidopsis. *Cell.* 2002;108(6):743-754.

67. Lewis JD, Wu R, Guttman DS, Desveaux D. Allele-specific virulence attenuation of the Pseudomonas syringae HopZ1a type III effector via the Arabidopsis ZAR1 resistance protein. 2010.

68. Henk AD, Warren RF, Innes RW. A new Ac-like transposon of Arabidopsis is associated with a deletion of the RPS5 disease resistance gene. *Genetics.* 1999;151(4):1581-1589.

69. Caicedo AL, Schaal BA, Kunkel BN. Diversity and molecular evolution of the RPS2 resistance gene in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences.* 1999;96(1):302-306.

70. Meng X, Zhang S. MAPK cascades in plant disease resistance signaling. *Annual Review of Phytopathology.* 2013;51:245-266.

71. Grant JJ, Chini A, Basu D, Loake GJ. Targeted activation tagging of the Arabidopsis NBS-LRR gene, ADR1, conveys resistance to virulent pathogens. *Molecular Plant-Microbe Interactions.* 2003;16(8):669-680.

72. McDowell JM, Dhandaydham M, Long TA, et al. Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of Arabidopsis. *The Plant Cell.* 1998;10(11):1861-1874.

73. Rose LE, Bittner-Eddy PD, Langley CH, Holub EB, Michelmore RW, Beynon JL. The maintenance of extreme amino acid diversity at the disease resistance gene, RPP13, in Arabidopsis thaliana. *Genetics.* 2004;166(3):1517-1527.

74. Brooks SA, Huang L, Gill BS, Fellers JP. Analysis of 106 kb of contiguous DNA sequence from the D genome of wheat reveals high gene density and a complex arrangement of genes related to disease resistance. *Genome.* 2002;45(5):963-972.

75. Xu X, Hayashi N, Wang C-T, et al. Rice blast resistance gene Pikahei-1 (t), a member of a resistance gene cluster on chromosome 4, encodes a nucleotide-binding site and leucine-rich repeat protein. *Molecular breeding.* 2014;34(2):691-700.

76. Perumalsamy S, Bharani M, Sudha M, et al. Functional marker‐assisted selection for bacterial leaf blight resistance genes in rice (Oryza sativa L.). *Plant breeding.* 2010;129(4):400-406.

77. Yoshimura S, Yamanouchi U, Katayose Y, et al. Expression of Xa1, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. *Proceedings of the National Academy of Sciences.* 1998;95(4):1663-1668.

78. Collins N, Drake J, Ayliffe M, et al. Molecular characterization of the maize Rp1-D rust resistance haplotype and its mutants. *The Plant Cell.* 1999;11(7):1365-1376.

79. Andersen EJ, Shaw SR, Nepal MP. Identification of Disease Resistance Genes in Aegilops Tauschii Coss. (Poaceae). *Proceedings of the South Dakota Academy of Science.* 2015;94:273-287.

80. Stevens P. onwards. Angiosperm Phylogeny Website. Version 12, July 2012 [and more or less continuously updated since]. 2001.

81. Varshney RK, Koebner RM. *Model plants and crop improvement.* CRC Press; 2006.

82. Wolfe KH, Gouy M, Yang Y-W, Sharp PM, Li W-H. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proceedings of the National Academy of Sciences.* 1989;86(16):6201-6205.

83. Bafna V, Pevzner PA. Sorting by reversals: Genome rearrangements in plant organelles and evolutionary history of X chromosome. Paper presented at: Mol. Biol. and Evol1995.

84. Leister D. Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. *Trends in Genetics.* 2004;20(3):116-122.

85. Ashfield T, Egan AN, Pfeil BE, et al. Evolution of a complex disease resistance gene cluster in diploid Phaseolus and tetraploid Glycine. *Plant physiology.* 2012;159(1):336-354.

86. Tian D, Araki H, Stahl E, Bergelson J, Kreitman M. Signature of balancing selection in Arabidopsis. *Proceedings of the National Academy of Sciences.* 2002;99(17):11525-11530.

87. Lozano R, Hamblin MT, Prochnik S, Jannink J-L. Identification and distribution of the NBS-LRR gene family in the Cassava genome. *BMC genomics.* 2015;16(1):360.

88. Pfeil B, Schlueter J, Shoemaker R, Doyle J. Placing paleopolyploidy in relation to taxon divergence: a phylogenetic analysis in legumes using 39 gene families. *Systematic Biology.* 2005;54(3):441-454.

89. Schmutz J, Cannon SB, Schlueter J, et al. Genome sequence of the palaeopolyploid soybean. *nature.* 2010;463(7278):178-183.

90. Lorang JM, Sweat TA, Wolpert TJ. Plant disease susceptibility conferred by a “resistance” gene. *Proceedings of the National Academy of Sciences.* 2007;104(37):14861-14866.

91. Seeholzer S, Tsuchimatsu T, Jordan T, et al. Diversity at the Mla powdery mildew resistance locus from cultivated barley reveals sites of positive selection. *Molecular plant-microbe interactions.* 2010;23(4):497-509.

92. Chaw S-M, Chang C-C, Chen H-L, Li W-H. Dating the monocot–dicot divergence and the origin of core eudicots using whole chloroplast genomes. *Journal of Molecular Evolution.* 2004;58(4):424-441.

93. Ling H-Q, Zhao S, Liu D, et al. Draft genome of the wheat A-genome progenitor Triticum urartu. *Nature.* 2013;496(7443):87-90.

94. Goff SA, Ricke D, Lan T-H, et al. A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). *Science.* 2002;296(5565):92-100.

95. Tan X, Meyers BC, Kozik A, et al. Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in Arabidopsis. *BMC plant biology.* 2007;7(1):56.

96. Dinesh-Kumar S, Baker BJ. Alternatively spliced N resistance gene transcripts: their possible role in tobacco mosaic virus resistance. *Proceedings of the National Academy of Sciences.* 2000;97(4):1908-1913.

97. Wei F, Gobelman-Werner K, Morroll SM, et al. The Mla (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. *Genetics.* 1999;153(4):1929-1948.

98. Kuang H, Woo S-S, Meyers BC, Nevo E, Michelmore RW. Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *The plant cell.* 2004;16(11):2870-2894.

99. Schlueter JA, Dixon P, Granger C, et al. Mining EST databases to resolve evolutionary events in major crop species. *Genome.* 2004;47(5):868-876.

CHAPTER 3: GENETIC DIVERSITY OF DISEASE RESISTANCE GENES IN FOXTAIL MILLET (*SETARIA ITALICA* L. BEAUV.)

This chapter is published in the journal *Plant Gene* and supplementary data is published in the journal *Data in Brief*:

Andersen, E.J. and M.P. Nepal. 2017. Genetic diversity of disease resistance genes in foxtail millet (*Setaria italica* L.). *Plant Gene*. 10:8-16.

Andersen, E.J. and M.P. Nepal. 2017. Data on the genome-wide identification of CNL R-genes in *Setaria italica* (L.) P. Beauv. *Data in Brief*. 13:259-273.

ABSTRACT

Foxtail millet (*Setaria italica* L. Beauv.) is a potential biofuel plant species, which is also one of the most commonly cultivated millet species for food and fodder. The aims of this study are to conduct a genome-wide identification of Coiled-coil, Nucleotide-binding site, Leucine-rich repeat (CNL) disease resistance genes (R-genes) in foxtail millet and study their evolutionary relationships. A total of 242 CNL genes were identified with domains for NBS-LRR receptor function, with the addition of a few genes that contained transmembrane or zinc finger domains. Of the identified CNL genes, more than half formed gene clusters within the foxtail millet genome, with the majority showing evidence of tandem duplications. Syntenic analysis displayed chromosomal similarities among foxtail millet, rice and barley, identifying strong syntenic relationships between foxtail millet and rice. Approximately 30% of the foxtail millet CNL gene clusters were found on chromosome Si08, exhibiting strong synteny with chromosome Os11 of rice. Selection pressure analysis showed a prevalence of purifying selection among all phylogenetic clades. Foxtail millet homologs of several well-studied R-genes in other

species were identified. The results from this study will have implications in research on CNL gene signaling pathways, and development of durable resistance in foxtail millet and other crop species in general.

INTRODUCTION

Plants face a broad spectrum of pathogens including viruses, bacteria, and fungi [1-3], as well as pests including nematodes, parasitic plants, and insects [4-6]. Several models have been proposed to describe the diverse mechanisms utilized by plants to resist pathogens and pests, such as the Gene-for-Gene model [7], Zig-Zag model [8], Guard model [9,10], and Decoy model [11]. These plant resistance mechanisms can evolve quickly [12-14] and are initiated by pathogen-associated molecular patterns (PAMPs) or pathogen effectors [8,15]. Nucleotide-Binding Site, Leucine-Rich Repeat (NBS-LRR) receptors in plants are activated upon contact with pathogenic effectors, and a pathogen-specific defense response is triggered [15]. Response variation corresponds to pathogen diversity, such as pathogens deriving nutrients from living tissue (biotrophic), dying tissue (necrotrophic), or switching from biotrophy to necrotrophy as the infection progresses (hemibiotrophic) [16,17]. Due to the biotrophic requirement of living host tissue, hypersensitive response, or triggered cell death of infected tissue, is effective against biotrophs [18], but fails to inhibit necrotrophs [16]. Plants can also produce chemicals to inhibit pathogen growth [19], thicken their cell walls [20], or utilize systemic signaling via salicylic acid, ethylene, or jasmonates [16].

Foxtail millet (*Setaria italica* L. Beauv.; family Poaceae; tribe Paniceae) is an important food crop and a potential biofuel plant species. *S. italica* was domesticated from *Setaria viridis* between 8,700 and 5,900 years ago in northern China [21,22], where a dry climate gave the C_4 photosynthetic pathway of this plant an advantage over other species [4,23]. For this reason, foxtail millet is a model for studying phosphate transport mechanisms (Ceasar et al., 2014) and abiotic stress tolerance, particularly in response to draught [24]. Recent availability of the complete *S. italica* genomic sequence [23] facilitates investigation into functional genomics and evolutionary divergence of foxtail millet and closely related species [25]. In a previous investigation using a Genome-Wide Association Study (GWAS), 916 varieties of foxtail millet were sequenced to construct a haplotype map (Jia et al., 2013) [26]. Switchgrass (*Panicum virgatum* L.), another potential biofuel species of the same family and tribe, diverged from foxtail millet about 13 million years ago [25,27]. Therefore, foxtail millet has been used as a reference in the assembly of the switchgrass genome [28]. *S. viridis* is also studied in context with foxtail millet, with both species sharing morphological similarities to other Panicoideae members [29]. Previous analyses have also focused on identifying genomic relationships between foxtail millet, rice, and switchgrass (Devos et al., 1998; Daverdin et al., 2015), with rice being another model grass species and major contributor to the global food supply [30,31]. Among the diseases common to foxtail millet, blast caused by the hemibiotroph *Magnaporthe grisea* (*Magnaporthe oryzae*) is a major issue that can lead to 60% crop yield loss [32]. This pathogen also greatly affects rice [33], classifying *M. grisea* as a significant detriment to the human food supply [34] and a significant focus of research. In rice, the *Pib* gene has been identified as an NBS-LRR R-gene that confers rice blast resistance [35]. Additionally, the

Pi-ta and *Pi9* NBS-LRR genes have been studied extensively and confer resistance to rice blast [36-41]. Other diseases that impact foxtail millet production include downy mildew, leaf rust, and smut, caused by *Sclerospora graminicola*, *Uromyces setariae-italiae*, and *Ustilago crameri*, respectively [42].

The main objectives of this study were to conduct a genome-wide identification of foxtail millet CNL R-genes and assess their evolutionary relationships. Assessment of R-gene evolutionary patterns was accomplished by: 1) construction of a phylogenetic tree including foxtail millet, *Arabidopsis*, rice, and barley; 2) inference of functional homologs using BLAST results; and 3) generation of syntenic maps comparing chromosomes of foxtail millet, rice, and barley. With foxtail millet's status as a model plant for crop and biofuel species, genetic evidence of pathogen resistance can be utilized in other grass species with sequenced genomes. Biofuel alternatives to maize that can grow in areas otherwise unsuitable for agriculture will likely expand the available resources for production.

MATERIALS AND METHODS

CNL R-gene identification

Identification of CNL genes in foxtail millet follows methods for the *Arabidopsis thaliana* [43], soybean [44], and barley [45] genomes. Foxtail millet protein sequences were downloaded from the Ensembl Genomes database [46]. *Arabidopsis* reference NB-ARC sequences [43] were accessed from Phytozome [47] and aligned within ClustalW [48]. A Hidden Markov Model (HMM) was constructed using HMMER version 3.1b2 [49] at a 0.05

stringency. Identified protein sequences were then annotated using Pfam [50] within InterProScan [51] and NB-ARCs were extracted, aligned within ClustalW, and used as reference sequences in a reiterative HMM with a 0.001 stringency. Identified protein sequences were again annotated using Pfam and sequences with NB-ARC and "DiseaseResist" domains were then extracted and imported into Multiple Expectation maximization for Motif Elicitation (MEME) software [52] to isolate sequences with P-loop, Kinase-2, and GLPL motifs.

Phylogenetic analysis

NB-ARCs for all four species (*Arabidopsis*, barley, rice, and foxtail millet) were aligned using ClustalW integrated within the program Geneious [53] and manually trimmed. The alignment was then imported into MEGA 7 [54,55] and a maximum likelihood model test resulted in the JTT+G+F model, which was used in constructing a maximum-likelihood tree with 100 bootstrap replicates. MEME results and clade information was added to the phylogenetic tree using the Interactive Tree Of Life (ITOL) website (Letunic and Bork, 2016). In order to assess homology to other species, BLAST results for each of the 242 foxtail millet NB-ARCs were compiled by: 1) uploading foxtail millet sequences to the PLAN BLAST server [56] set to return the top ten hits, and 2) a script file was composed initiating BLASTP with an e-value threshold of $1x10^{-50}$ using the program BLAST+ [57] in the South Dakota State University High Performance Computing Cluster (HPC Cluster). The Biomart function within the Ensembl Genomes database was used to access additional InterProScan and Gene Ontology (GO) annotation data [58] for the identified CNL genes in order to make inferences regarding gene function. InterProScan

and MEME analyses were used to identify likely CNL genes among 88 *Sorghum bicolor* NBS-LRR sequences [59] and 405 *Panicum virgatum* CC and NB-ARC containing sequences [60], resulting in 80 and 234 genes, respectively. These genes were aligned with foxtail millet and *Arabidopsis* sequences and a maximum-likelihood model test was performed, leading to the construction a maximum-likelihood tree in MEGA 7 (model: JTT+G+F) with 100 bootstrap replicates.

Gene structure, clustering, and synteny

Exon locations accessed from Ensembl Genomes site were uploaded to the Gene Structure Display Server [61] to generate exon maps. Coding sequence locations, also accessed from Ensembl Genomes, were used to annotate chromosome maps and detect R-gene clustering as in previous studies [62]. The program R [63] was used to generate a plot describing the relative densities of R-genes across foxtail millet, rice, and barley chromosomes. Coding sequence nonsynonymous substitutions per nonsynonymous site (K_a) and synonymous substitutions per synonymous site (K_s) were calculated using DnaSP version 5.10.1 [64] to compare selection pressure and relative age between CNL clades and clusters. Using the HPC cluster, syntenic maps of barley, rice, and foxtail millet were generated from whole chromosome sequences and CNL R-gene location annotation files in the program SyMAP version 4.2 [65].

RESULTS

CNL R-genes Identification

I identified 242 CNL genes in the foxtail millet genome. The nesting patterns of the identified genes showed a clear expansion of CNL-C (237 members) and a reduction in CNL-A (one member: Seita.3G369800), CNL-B (four members: Seita.5G344600, Seita.7G165200, Seita.7G164900, and Seita.9G466500), and a complete absence of CNL-D (**Figure 3.1**). The NB-ARC signature motifs can be seen in Fig. 1 with RNBS-A, B, and C motifs identified in 225, 236, and 210 genes, respectively. InterProScan and GO annotations (see **Supplementary Table 3.1**), include the following sequence descriptions: ATPase (10 genes), ATP binding site (20 genes), ADP binding (242 genes), DNA-binding (Seita.8G039400, Seita.7G242200, and Seita.J025700), DNA-binding WRKY (Seita.2G175200), mannose-binding lectin (Seita.8G124300), no apical meristem protein (Seita.8G039400), phospholipase C (Seita.3G100200), P-loop containing nucleoside triphosphate hydrolase (242 genes), tetratricopeptide repeat (Seita.8G133400), phosphorus-containing group transferase activity (Seita.8G100000 and Seita.4G035100), zinc finger (Seita.7G242200, Seita.J025700, and Seita.1G053100), and multiple kinase domains (Seita.4G035100 and Seita.8G100000). It should be noted that the ADP binding and nucleoside triphosphate hydrolase annotation were found in all 242 of foxtail millet's R-genes. Foxtail millet R-gene exon composition ranged from one (37 genes) to thirteen (i.e. Seita.8G100000) exons, with an average of 3.1 exons per gene (**Supplementary Figure 3.1**).

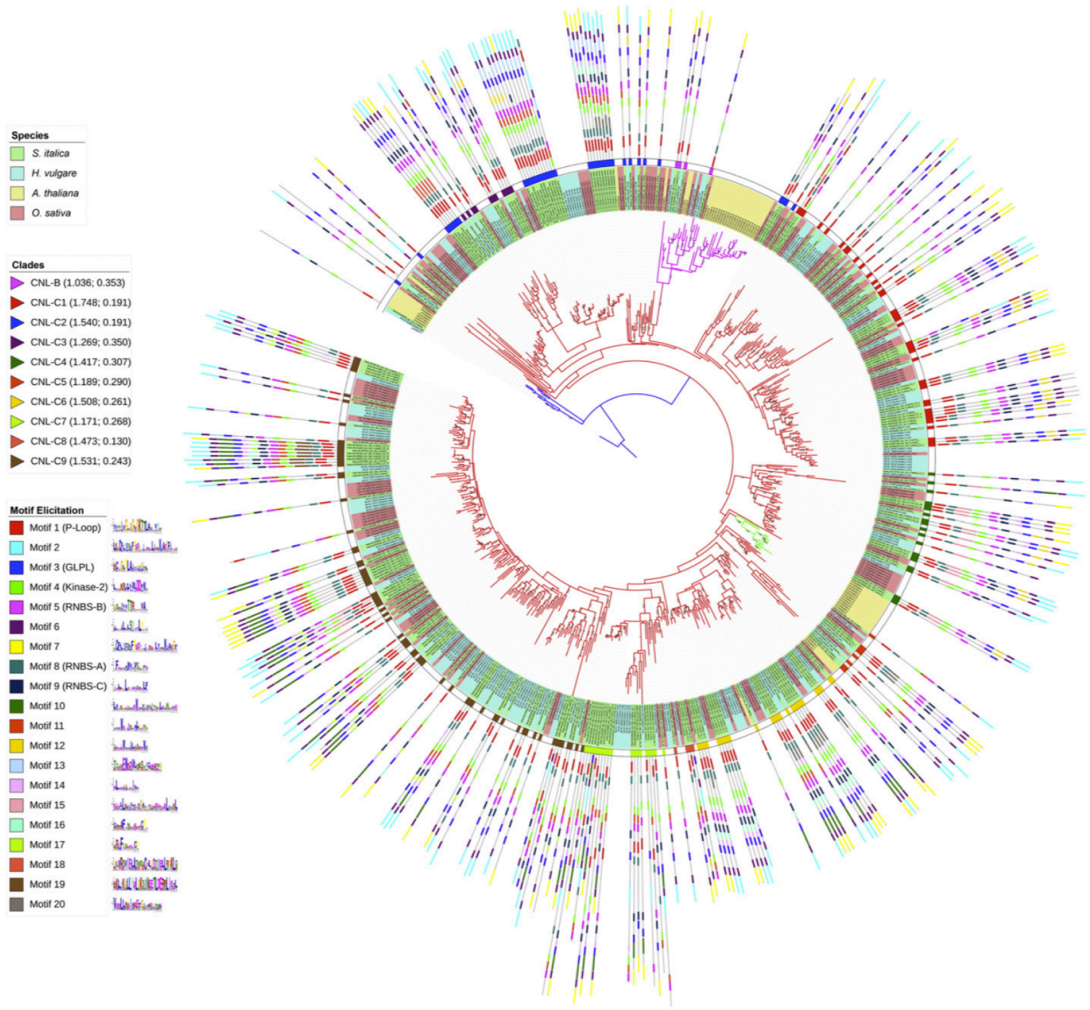


Figure 3.1 Maximum-likelihood phylogenetic analysis of the NB-ARC amino acid sequences of CNL genes in *Setaria italica* (Seita), *Hordeum vulgare* (MLOC), *Arabidopsis thaliana* (AT), and *Oryza sativa* (LOC) with labels shaded in green, blue, yellow, and pink, respectively. The JTT+G+F model with 100 bootstrap replicates was used to construct this tree rooted on the outgroup p25941 from *Streptomyces coelicolor*. Foxtail millet accessions are followed by their respective number of exons, chromosome, GC-content percentage, and genomic cluster as listed in Table 3.1**, respectively. Barley and rice accession names are also followed by clade and chromosome location, respectively. Clades are labeled by color in the strip**

surrounding the accessions and protein motifs from the MEME analysis are included radiating outward from the accession and clade information.

Genomic structure and synteny

Approximately 55% of foxtail millet CNL genes formed gene clusters (see **Figure 3.2** and **Table 3.1**). Of the 53 clusters, 36 were composed of two genes each. Many of the clusters presented evidence of tandem duplication, especially the approximate 66% of clusters that formed intraspecific clades in the phylogenetic tree (see **Figure 3.1**). Chromosome Os11 in rice is densely populated with CNL gene clusters and contains 40 of the 149 CNL orthologs (approximately 27%). A similar case exists in the foxtail millet genome, with chromosome Si08 containing 79 of 242 genes (approximately 33%). Syntenic analysis showed a very strong similarity between chromosomes Os11 and Si08 (see **Figure 3.3**). Several instances of possible chromosomal inversion were found between rice and foxtail millet. For instance, the region between 5.25-8.75 Mb on Os11 and 8-13 Mb on Si08 show a likely chromosomal inversion in the evolutionary history of one of the two species (see **Supplementary Figure 3.2**). Evidence of a similar event was witnessed between Os02 and Si01, where the first 10 Mb of the chromosomes appear to be inverted. Synteny between barley and foxtail millet is much less prevalent and blocks were more spread out (**Figure 3.3**). Foxtail millet R-genes were more evenly distributed across chromosomes than those in barley, prominently seen on chromosomes Si02, Si04, Si07, Si08, and Si09 (**Figure 3.2**; also see **Supplementary Figure 3.3**).

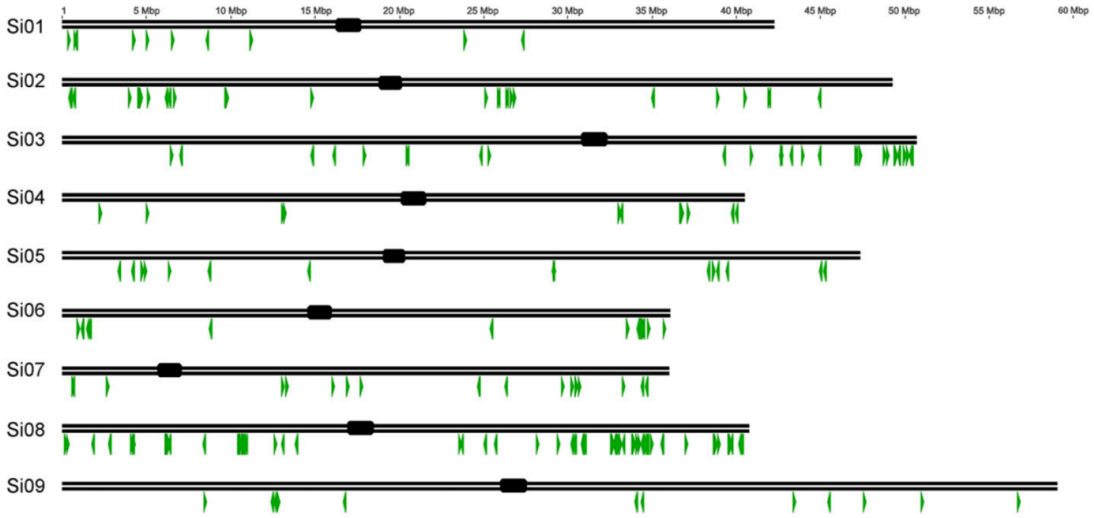


Figure 3.2 Chromosomal map of the CNL R-genes found in foxtail millet. Black bars represent each of foxtail millet's nine chromosomes with black rectangles denoting centromere positions. Arrows represent gene locations.

Gene homology and evolution

Homology inferred from BLAST results for each of the 242 foxtail millet sequences showed a prevalence of homologs in *Brachypodium distachyon* and rice, with some to wheat and barley (**Supplementary Table 3.2**). Rice homologs *Lr21* (Seita.6G021300), *Pi* (34 genes), *MLA* (seven genes), *Yr10* (Seita.1G006400 and Seita.2G050400), *Xa1* (five genes), *YNR1-5* (Seita.8G184900 and Seita.8G184000), and *Pm3b* (five genes) along with *Brachypodium* homologs *RPM1* (57 genes), *RPP13* (29 genes), and *RPS2* (four genes) were found (see **Supplementary Table 3.2**). Homologs of wheat *Lr21* (Seita.6G021300), *Pm3b* (Seita.3G406200 and Seita.3G402500), *Tsn1* (Seita.8G097100), *PM* (Seita.3G107100 and Seita.2G055800), and *Yr10* (Seita.8G133400) were identified, along with a few select homologs in *Zea mays* (*PIC21*,

Rp1, and *RXO*), *Sorghum bicolor* (*Pc* and *Yr10/Mla1*), *Hordeum vulgare* (*NBS3-RDG2A*), and *Vitis vinifera* (*RPM1*, *RPP8*, and *RDL6/RF9*) (see **Supplementary Table 3.2**). *S. bicolor* and *P. virgatum*, both nested closely with foxtail millet sequences, with all the grasses showing reductions in CNL-A and CNL-B, expansion in CNL-C, and absence of CNL-D (**Supplementary Figure 3.4**). K_a/K_s pairwise comparisons of all genes within the same clade were substantially less than one (Fig. 1), with the only exceptions being two pairs from CNL-C4 (Seita.8G208400-Seita.9G134600 and Seita.2G172800-Seita.9G549300) and one pair from CNL-C9 (Seita.2G009000-Seita.2G103400). K_s values for each of these clades showed that CNL-C1 had the highest K_s value of 1.748 and CNL-B had the lowest K_s value of 1.036 (**Figure 3.1**).

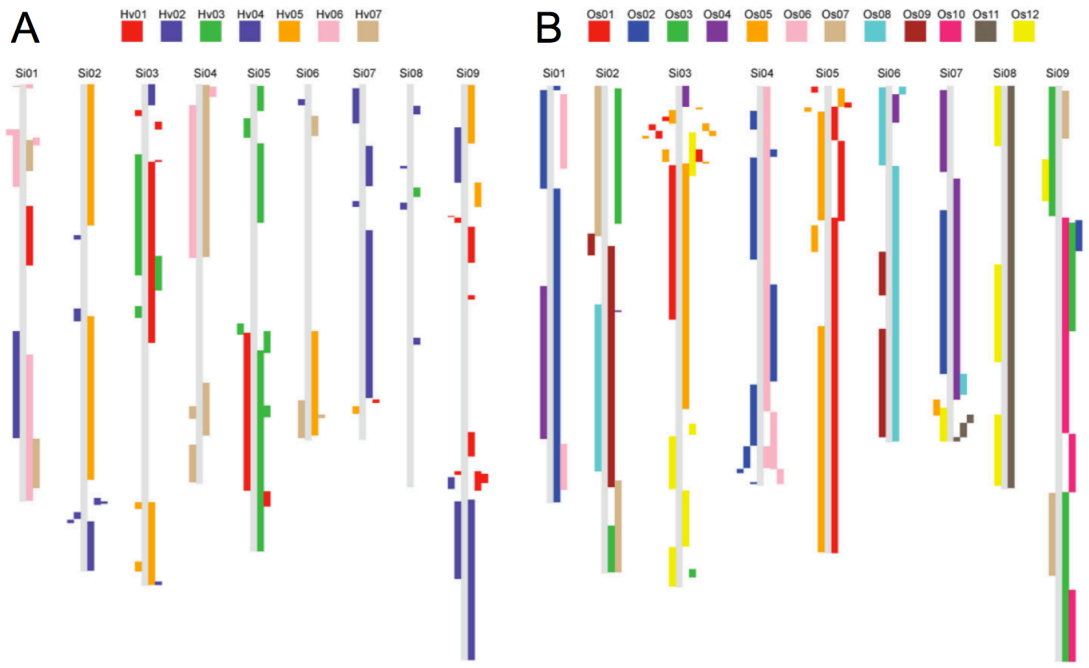


Figure 3.3 Syntenic map of foxtail millet CNL genes with barley (A) and rice (B).

Table 3.1 Gene clusters within the foxtail millet genome, named with the first number representing the chromosome in which the cluster is found and the second number representing the order of the clusters on that chromosome. Clade and K_s values for each cluster are also included; "NA" is used when a particular K_s value could not be calculated.

Cluster	Clustered Genes	K_s value
1_1	Seita.1G010400 (C1), Seita.1G011100 (C1)	0.000
1_2	Seita.1G126600 (C1), Seita.1G126700 (C1), Seita.1G126800 (C6)	0.745
2_1	Seita.2G008500 (C9), Seita.2G009000 (C9)	0.040
2_2	Seita.2G055500 (C2), Seita.2G055800 (C2), Seita.2G056400 (C2)	0.055
2_3	Seita.2G056600 (C2), Seita.2G057000 (C2)	0.082
2_4	Seita.2G103300 (C9), Seita.2G103400 (C9)	0.129
2_5	Seita.2G169100 (C1), Seita.2G169300 (C1)	0.221
2_6	Seita.2G172700 (C4), Seita.2G172800 (C4)	NA
2_7	Seita.2G175100 (C9), Seita.2G175200 (C9)	0.894
2_8	Seita.2G335000 (C2), Seita.2G335100 (C2)	NA
3_1	Seita.3G099900 (C2), Seita.3G100200 (C2)	0.042
3_2	Seita.3G107100 (C2), Seita.3G107300 (C2)	0.090
3_3	Seita.3G241300 (C1), Seita.3G241400 (C1)	3.124
3_4	Seita.3G333300 (C9), Seita.3G333400 (C9)	2.204
3_5	Seita.3G367600 (C1), Seita.3G368000 (C1)	0.000
3_6	Seita.3G393500 (C1), Seita.3G393700 (C1)	0.000
3_7	Seita.3G395700 (C9), Seita.3G396100 (C1), Seita.3G396300 (C7)	2.003
3_8	Seita.3G400200 (C2), Seita.3G400300 (C2)	0.085
3_9	Seita.3G406200 (C2), Seita.3G406300 (C2)	0.040

4_1	Seita.4G126400 (C8), Seita.4G126900 (C8), Seita.4G127000 (C8)	1.599
4_2	Seita.4G243900 (C6), Seita.4G244000 (C6), Seita.4G244400 (C6)	0.928
5_1	Seita.5G230900 (C9), Seita.5G231000 (C9)	NA
6_1	Seita.6G014500 (C9), Seita.6G014600 (C9), Seita.6G014700 (C9), Seita.6G014800 (C9)	0.000
6_2	Seita.6G017100 (C9), Seita.6G017500 (C3)	NA
6_3	Seita.6G023500 (C9), Seita.6G023600 (C9)	0.000
6_4	Seita.6G229200 (C7), Seita.6G229300 (C7), Seita.6G229600 (C7), Seita.6G229700 (C7)	0.430
6_5	Seita.6G232800 (C2), Seita.6G233200 (C2), Seita.6G233300 (C2), Seita.6G233400 (C2)	0.318
7_1	Seita.7G003900 (C9), Seita.7G004000 (C3)	2.283
7_2	Seita.7G164900 (B), Seita.7G165200 (B)	1.318
7_3	Seita.7G241500 (C1), Seita.7G241600 (C1), Seita.7G242200 (C1)	1.085
7_4	Seita.7G245900 (C1), Seita.7G246200 (C1), Seita.7G246400 (C1)	0.303
8_1	Seita.8G039400 (C9), Seita.8G039500 (C9)	0.920
8_2	Seita.8G049900 (C9), Seita.8G050000 (C9)	0.702
8_3	Seita.8G064200 (C6), Seita.8G064300 (C6), Seita.8G064400 (C6), Seita.8G064700 (C6), Seita.8G064900 (C6)	0.395
8_4	Seita.8G087200 (C9), Seita.8G087300 (C6)	1.805
8_5	Seita.8G088100 (C6), Seita.8G088200 (C6), Seita.8G088300 (C6)	1.904
8_6	Seita.8G088900 (C6), Seita.8G089000 (C9), Seita.8G089100 (C9), Seita.8G089200 (C6), Seita.8G089500 (C7)	1.929
8_7	Seita.8G090100 (C7), Seita.8G090200 (C6)	NA
8_8	Seita.8G123800 (C), Seita.8G124300 (C9)	3.419
8_9	Seita.8G162600 (C4), Seita.8G162700 (C4)	0.044
8_10	Seita.8G166700 (C7), Seita.8G166900 (C7), Seita.8G167100 (C7), Seita.8G167300 (C7), Seita.8G167500 (C7)	0.026
8_11	Seita.8G181000 (C6), Seita.8G181900 (C7)	2.866
8_12	Seita.8G182600 (C9), Seita.8G183100 (C6)	NA
8_13	Seita.8G183400 (C6), Seita.8G184000 (C9), Seita.8G184200 (C9)	NA
8_14	Seita.8G184600 (C9), Seita.8G184900 (C9)	1.093
8_15	Seita.8G191900 (C9), Seita.8G192200 (C9)	0.129
8_16	Seita.8G199500 (C9), Seita.8G199600 (C9)	0.539
8_17	Seita.8G200100 (C9), Seita.8G200400 (C9), Seita.8G201100 (C4)	1.074

8_18	Seita.8G208200 (C4), Seita.8G208400 (C4)	0.042
8_19	Seita.8G234200 (C2), Seita.8G235200 (C2)	0.000
8_20	Seita.8G242600 (C2), Seita.8G242800 (C2), Seita.8G243100 (C7)	1.556
8_21	Seita.8G249100 (C4), Seita.8G249300 (C4)	0.092
9_1	Seita.9G374900 (C5), Seita.9G375200 (C5)	0.000

DISCUSSION

Genetic structure

In a previous study, 96 NBS-LRR genes were identified in the foxtail millet genome, 45 of which were CNL [66]. The 242 genes identified in the present study indicate that the number of CNL genes in foxtail millet is over five times higher than formerly reported and that the majority of them are CNL-C members. CNL-D members were completely absent in foxtail millet (**Figure 3.1**), consistent to the findings in several previous studies [44,45,67]. Two key differences stood out between the gene nesting of foxtail millet with that of barley: 1) CNL-C appears to be ancestral to CNL-B in foxtail millet and 2) CNL-C2 is basal to clades B, C1, and C3 in foxtail millet (see Fig. 1). These results indicate uncertainty regarding the clade positions in the phylogenetic tree of the CNL genes in foxtail millet and barley. In fact, there was weak or no bootstrap support for the four major clades of the barley CNL genes. Activation of NBS-LRR proteins through the removal of ADP and binding to ATP [68] is supported by annotation of ADP and ATP-binding sites in the identified genes, with specific annotations for each gene being found in **Supplementary Table 3.1**. An earlier study [69] showed a prevalence of the amino acid residues DDVD in TNL R-genes and DDVW in non-TNL genes, which this study confirms within foxtail millet, pointing out the highly conserved nature of the NB-

ARC within plant species. CNL proteins with unique domains were identified, such as zinc fingers (amino acid residues: CxxCxxCxx…HxxxH) in accessions Seita.7G242200, Seita.J025700, and Seita.1G053100 which potentially interact with DNA to directly initiate a defense response after activation by pathogen effectors [70]. Seita.7G242200 and Seita.J025700 contain "DNA-binding" annotations, strengthening this hypothesis. Nineteen of the 242 genes contained "nucleotide binding" annotations, with Seita.2G175200 containing annotations for "transcription factor activity" and "DNA-binding WRKY" (**Supplementary Table 3.1**). Seita.3G100200 contained a "lipid metabolic process" annotation, which may play a role in the change in metabolism that occurs during plant defense response [71]. Exon-intron variation (see **Supplementary Figure 3.1**) is very similar between foxtail millet and barley [45], with such a prevalence of introns in particular genes indicating that alternative splicing may be an important regulatory mechanism of R-gene expression [72].

Six of the foxtail millet protein sequences (Seita.2G315000, Seita.6G021300, Seita.3G396300, Seita.8G192200, Seita.2G172700, and Seita.9G466500) have evidence of transmembrane domains, indicating that some *S. italica* CNL proteins are localized within cellular membranes, in contrast to the prevalence of cytoplasmic NBS-LRR proteins [73]. Of these transmembrane domain-containing genes, Seita.9G466500 and Seita.3G396300 showed strong homology to RPS2 and RPM1, respectively, each of which are membrane-localized R-proteins that initiate defensive cellular signaling [74,75]. Possessing such characteristics, it can be inferred that Seita.9G466500 and Seita.3G396300 are localized within the plasma membrane and confer resistance to *Pseudomonas syringae*. The other genes listed (Seita.2G315000, Seita.6G021300,

Seita.8G192200, and Seita.2G172700) do not explicitly show homology to membrane localized R-genes so further investigation will be necessary to confirm their localization. These four genes show homology to R-protein RPP13, which has been shown to be localized in the cytoplasm [76].

Genomic structure and synteny

The number of gene clusters in foxtail millet (**Figure 3.2** and **Table 3.1**) is much higher than that in barley but similar to rice [45,77]. The genomes of foxtail millet and rice possess many conserved attributes that two earlier studies show in their syntenic maps [23,78], corresponding heavily to **Figure 3.3** (also see **Supplementary Figure 3.2**). This confirms the chromosomal relationships in the following ways: (Si01=Os02); (Si02=Os07+Os09); (Si03=Os05+Os12); (Si04=Os06); (Si05=Os01); (Si06=Os08); (Si07=Os04); (Si08=Os11); and (Si09=Os03+Os10) [23,78]. Syntenic relationships between foxtail millet and barley R-gene-containing regions exist within the extra-pericentromeric regions of chromosomes, the locations of barley's CNL genes [45], showing the large divergence in their respective genome sizes (**Figure 3.3**). This indicates that the ancestor of barley and wheat experienced substantial evolutionary divergence from their common ancestor with rice, and that each lineage underwent a diverse, independent evolutionary history. A remnant of this can be seen by the fact that no single barley chromosome contains such a substantial amount of its CNL genes [45] and foxtail millet R-genes are more evenly distributed across chromosomes (**Supplementary Figure 3.3**).

Gene homology and evolution

Rice, barley, and foxtail millet all share a common ancestor about 57 million years ago [79]. From the prevalence of homologs found in foxtail millet (**Supplementary Table 3.2**), it can be inferred that Seita.6G021300 is the likely *S. italica* ortholog of *Lr21* and confers resistance to leaf rust [80]; Seita.3G406200 and Seita.3G402500 confer resistance to powdery mildew as *Pm3b* [81]; and Seita.J025700 and Seita.9G392900 confer resistance to bacterial blight as *Xa1* [82]. Fungal diseases like rusts and powdery mildew are particularly damaging to cereal crops and resistance to these diseases have been studied extensively, especially in wheat [83-86]. Several genes, including all three genes of cluster 7_4 (Seita.7G245900, Seita.7G246200, and Seita.7G246400), show homology to fungal R-protein RGA4, which functionally interacts with RGA5 in the recognition of *M. oryzae* effectors AVR-Pia and AVR1-CO39 [87,88]. Cluster 6_5 (Seita.6G232800, Seita.6G233200, Seita.6G233300, and Seita.6G233400) contains homologs of leaf stripe (*Pyrenophora graminea*) R-gene *Rdg2a* [89]. Clusters 8_15 (Seita.8G191900 and Seita.8G192200), 8_16 (Seita.8G199500 and Seita.8G199600), and 8_17 (Seita.8G200100, Seita.8G200400, and Seita.8G201100) all show homology to stripe rust R-gene *Yr10* and nest within C9, with *S. bicolor Yr10* having eight homologs in foxtail millet (see **Table 3.2** and **Supplementary Figure 3.4**). *S. bicolor* NBS-LRR genes, along with *P. virgatum*, share a similar nesting pattern to foxtail millet (**Supplementary Figure 3.4**), with *Zea mays* and *Sorghum bicolor* sharing a more recent common ancestor with foxtail millet than rice and barley [90]. Numerous foxtail millet genes share homology with maize *RGA4*, *rp3*, *Rp1*, and *RXO1*, along with sequences from the genus *Solanum* and the species *Vitis vinifera*

(for accessions see **Supplementary Table 3.2**), displaying the highly conserved nature of the NB-ARC.

The K_a/K_s ratios and K_s values for pairwise comparisons give information regarding the selection pressures and relative ages of gene groups, respectively. K_a/K_s ratios show a strong prevalence of purifying selection among all clades in **Figure 3.1**, as seen in barley [45], which can be expected for crop species [91]. The K_s-derived relative ages of the clades differ slightly from results found in barley [45], but concur with the nesting patterns of the phylogenetic tree where CNL-C is ancestral to CNL-B (**Figure 3.1**). It can also be inferred that clusters 8_8, 3_3, 8_11, 7_1, 3_4, and 3_7 are the oldest clusters to have formed (see **Table 3.1** for accessions within each cluster). With the exception of 3_7, these clusters are each composed of two genes. In the case of larger gene clusters, pairwise comparisons between more recent duplication events can lower the overall age of the cluster. For instance, cluster 8_6 is composed of five genes (Seita.8G088900, Seita.8G089000, Seita.8G089100, Seita.8G089200, and Seita.8G089500) and has a K_s value of 1.929. When specific pairwise comparisons are investigated, Seita.8G088900 and Seita.8G089500 have a K_s value of 3.542, making the duplication event that formed the pair much older than Seita.8G089200 and Seita.8G089100, with a K_s value of 1.018. Therefore, it can be inferred that within cluster 8_6, the Seita.8G088900 and Seita.8G089500 pair formed through duplication long before later duplicating to form Seita.8G089200 and Seita.8G089100. Homology to these genes also shows a difference, with Seita.8G088900 and Seita.8G089500 showing most homology to *RPM1* and Seita.8G089100 appearing to have diverged. Almost all foxtail millet *RPM1* homologs nested within either clade C6 or C9. *RPM1* homologs in C6 mostly showed homology

only to *RPM1*, with homologs in C9 also showing homology to *RPP13* or other R-genes. *RPM1* and *RPP13* each show a high degree of diversity that is maintained by balancing selection [92-94]. These findings lead to the prediction that these two genes either contain a highly conserved NB-ARC or have experienced convergent evolution. Convergent evolution has been observed in resistance to *Pseudomonas syringae* through detection of effectors AvrB and AvrRpm1, which evolved separately in soybean and *Arabidopsis* [95]; and variation in *RPM1* has resulted in loss-of-function alleles possibly due to the cost of R-gene expression [96].

Due to the destructive nature of blast, dozens of blast R-genes have been identified and cloned in rice, most of those being CNL genes [40]. Clade C7 contains many homologs of the *Pib* blast R-gene, which has been previously investigated in rice [35]. These homologs form an expansion in foxtail millet, which makes up cluster 8_10 (Seita.8G167100, Seita.8G166900, Seita.8G167300, Seita.8G166700, and Seita.8G167500) and five genes from chromosome 6 (Seita.6G229300, Seita.6G229600, Si13130, Seita.6G229700, Seita.6G229200), four of which form cluster 6_4. It can be inferred that blast exposure has heavily selected for an expansion of these genes and duplication provided a mechanism for this expansion. Gene pairs within cluster 8_10 all have very low K_s values (0-0.044), which correspond to cluster 6_4 (0.045-0.650). Since these two clusters also nested together, it can be hypothesized that cluster 6_4 formed first and the entire region may have duplicated onto chromosome 8. Another closely related cluster, 4_1 (Seita.4G126400, Seita.4G126900, and Seita.4G127000) may be the progenitor to cluster 6_4, having three genes with K_s values ranging from 1.477 to 1.727, one of which showing homology to *Pi9*, which also confers resistance to blast [39].

Accession Seita.8G181900 from the neighboring cluster 8_11 (clustered with accession Seita.8G181000) also shows homology to *Pib*. Other fungal R-genes are homologs of *Pm3b* (Seita.3G400200, Seita.3G400300, Seita.3G406200, Seita.3G406300, Seita.8G242800, and Seita.8G242600) and *Pc* (Seita.2G172800) genes, conferring resistance to *Blumeria graminis* [81] and *Periconia circinata* [97], respectively. Evidence of a more recent duplication event can be seen in cluster 6_1 (Seita.6G014500, Seita.6G014600, Seita.6G014700, and Seita.6G014800) where four homologs of *RGH1A* each have K_s values of 0.000, indicating recent origin not allowing enough time for diversification. Genes conferring resistance to pests were also found, especially within clusters 3_1 (Seita.3G099900 and Seita.3G100200), 3_2 (Seita.3G107100 and Seita.3G107300), and 3_8 (Seita.3G400200 and Seita.3G400300), showing homology to brown planthopper R-gene *Bph14* [4,98], conferring resistance to *Nilaparvata lugens*, a devastating pest throughout Asia [99]. These genes also showed homology to powdery mildew R-genes, another evidence of the highly conserved nature of the NB-ARC. However, since entomopathogenic fungal species such as *Beauveria bassiana* and *Metarhizium anisopliae* have been shown to control brown planthopper populations [100,101], future research should investigate the possibility that plants use fungal R-genes to select for the growth of species that reduce damage from insects.

Future directions

Foxtail millet is both a substantial crop and a model for the study of closely related grass species. Future research should aim at increasing the available knowledge about molecular pathways within foxtail millet, specifically regarding host-pathogen

interactions. Genomic data from other grass species (e.g. *Panicum hallii*, *P. virgatum*, and *S. viridis*) will further develop the understanding of R-gene evolution in *S. italica*. A demand for efficient biofuel production will undoubtedly require future breeders to develop durable resistance in species that previously did not have such global impact, especially as increased cultivation causes a greater pathogen exposure. These results have potential implications in elucidating specific signaling pathways, measuring expression before and after pathogen exposure, as well as further development of abiotic stress resistance in foxtail millet and other closely related grass species.

CONCLUSION

In this study, 242 CNL foxtail millet genes were identified. Conserved domains within the protein sequences concurred with expectations based upon existing knowledge of NBS-LRR receptor function, with the presence of additional transmembrane (six genes) and zinc finger domains (three genes). Syntenic analyses among foxtail millet, rice, and barley showed a high degree of genomic similarity between foxtail millet and rice, as previously reported. Approximately 55% of identified genes formed clusters and many presented evidence of tandem duplications as a means of diversification in response to rapidly diversifying pathogens (e.g. evolution of blast resistance). Homology of foxtail millet R genes was confirmed (i.e. *Lr21*, *P9*, *Pib*, *MLA*, *Yr10*, *Xa1*, *YNR1-5*, *RGA4*, *RPM1*, *RPP13* and *Pm3b*) to that of other crop species, and a prevalence of purifying selection as means of R-gene evolution in response to corresponding pathogens. Future efforts should include thorough analysis of disease resistance signaling mechanisms and

investigation of cross-talk with abiotic stress resistance pathways for cultivars of foxtail millet and other crop species.

SUPPLEMENTARY DATA

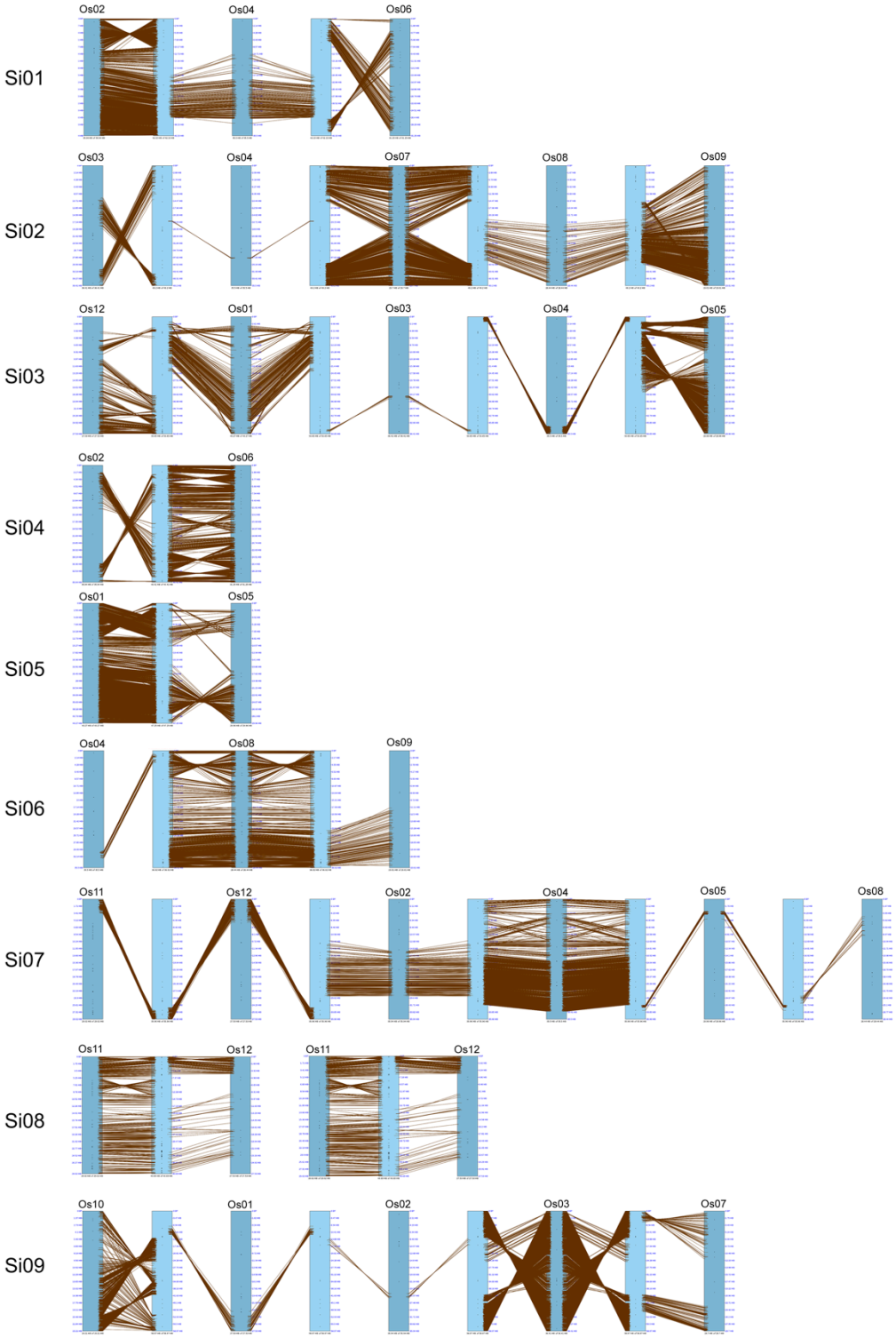
I report data on the identification of 242 disease resistance genes (R-genes) in the *Setaria italica* genome as presented in "Genetic diversity of disease resistance genes in foxtail millet (*Setaria italica* L.)"[102]. This data describe the structure and evolution of plant Coiled-coil, Nucleotide-binding site, Leucine-rich repeat (CNL) R-genes, and were obtained through rigorous extraction and analysis of recently available plant genome sequences using cutting-edge analytical software. In order to display the data, figures include gene structure diagrams, chromosomal syntenic maps, a chromosomal density plot, and a maximum-likelihood phylogenetic tree comparing: *Sorghum bicolor*, *Panicum virgatum*, *Setaria italica*, and *Arabidopsis thaliana*. Compilation of InterProScan annotations, Gene Ontology (GO) annotations, and Basic Local Alignment Search Tool (BLAST) results for the 242 R-genes identified in the foxtail millet genome are also included in tabular format.

These data provide details regarding foxtail millet R-genes[102]. Generated from sequence annotations, gene structure diagrams for each of the 242 genes were compiled (**Supplementary Figure 3.1**), along with InterProScan and Gene Ontology (GO) annotations (**Supplementary Table 3.1**). Utilizing R-gene genomic locations, chromosomal syntenic maps (**Supplementary Figure 3.2**) and R-gene density (**Supplementary Figure 3.3**) are displayed. Based on R-protein sequences, homologs were compiled (**Supplementary Table 3.2**) and a maximum-likelihood phylogenetic tree

(**Supplementary Figure 3.4**) displays the evolutionary relationships between *Setaria italica*, *Arabidopsis thaliana*, *Sorghum bicolor*, and *Panicum virgatum* accessions.

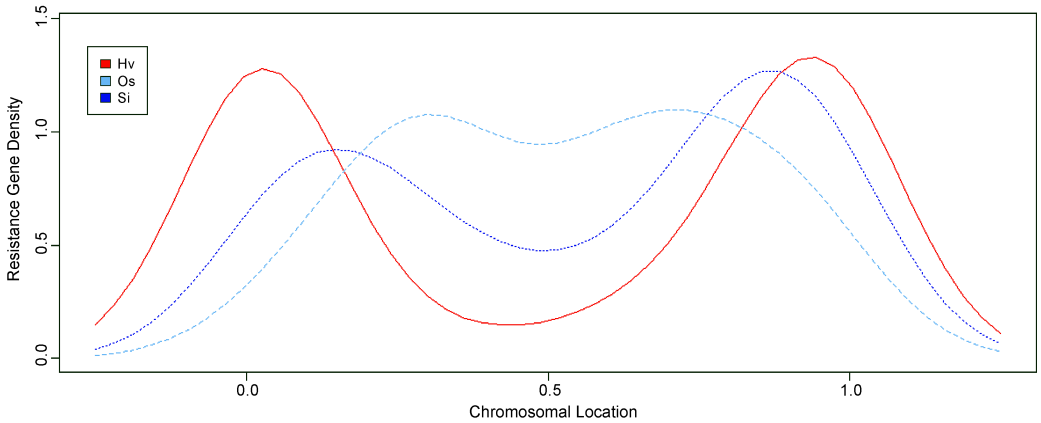
Exon and intron locations for the 242 identified R-genes were accessed from the Ensembl Genomes database[46] and uploaded to the Gene Structure Display Server[61] in BED format to generate the visual display of structure. InterProScan and GO annotations[51], also from the Ensemble Genomes Biomart, were downloaded and summarized in tabular format. Chromosome sequences for foxtail millet, rice, and barley genomes were uploaded in FASTA format to SyMAP version 4.2[65] along with GFF3 annotation files. Using the 2D Chromosome Explorer function within SyMAP, figures were generated comparing chromosomes. Chromosomal locations for each of the R-genes of foxtail millet, rice, and barley were uploaded to the program R[63] to generate a density plot visualization. *S. bicolor* and *P. virgatum* R-genes[60,103] were compiled with *S. italica* and *A. thaliana* sequences for the construction of a phylogenetic tree. Nucleotide-Binding, Apoptosis protease-activating factor-1, R-protein, *Caenorhabditis elegans* death-4 protein (NB-ARC) sequences were aligned within ClustalW[48] and a maximum-likelihood phylogenetic tree (100 bootstrap replicates) was generated in MEGA 7[54,55] using the JTT+G+F model, selected based upon a maximum-likelihood model test. This tree was edited in the Interactive Tree of Life server[104]. Using the South Dakota State University High Performance Computing Cluster and the PLAN server[56], BLAST results of the foxtail millet NB-ARC protein sequences were acquired and summarized in tabular format. *S. italica* accessions were updated to version 2.2 of the genome, available in Phytozome[28,47].

Supplementary Figure 3.1 Exon variation across foxtail millet CNL gene sequences, with exons and introns represented by yellow bars and grey lines, respectively.

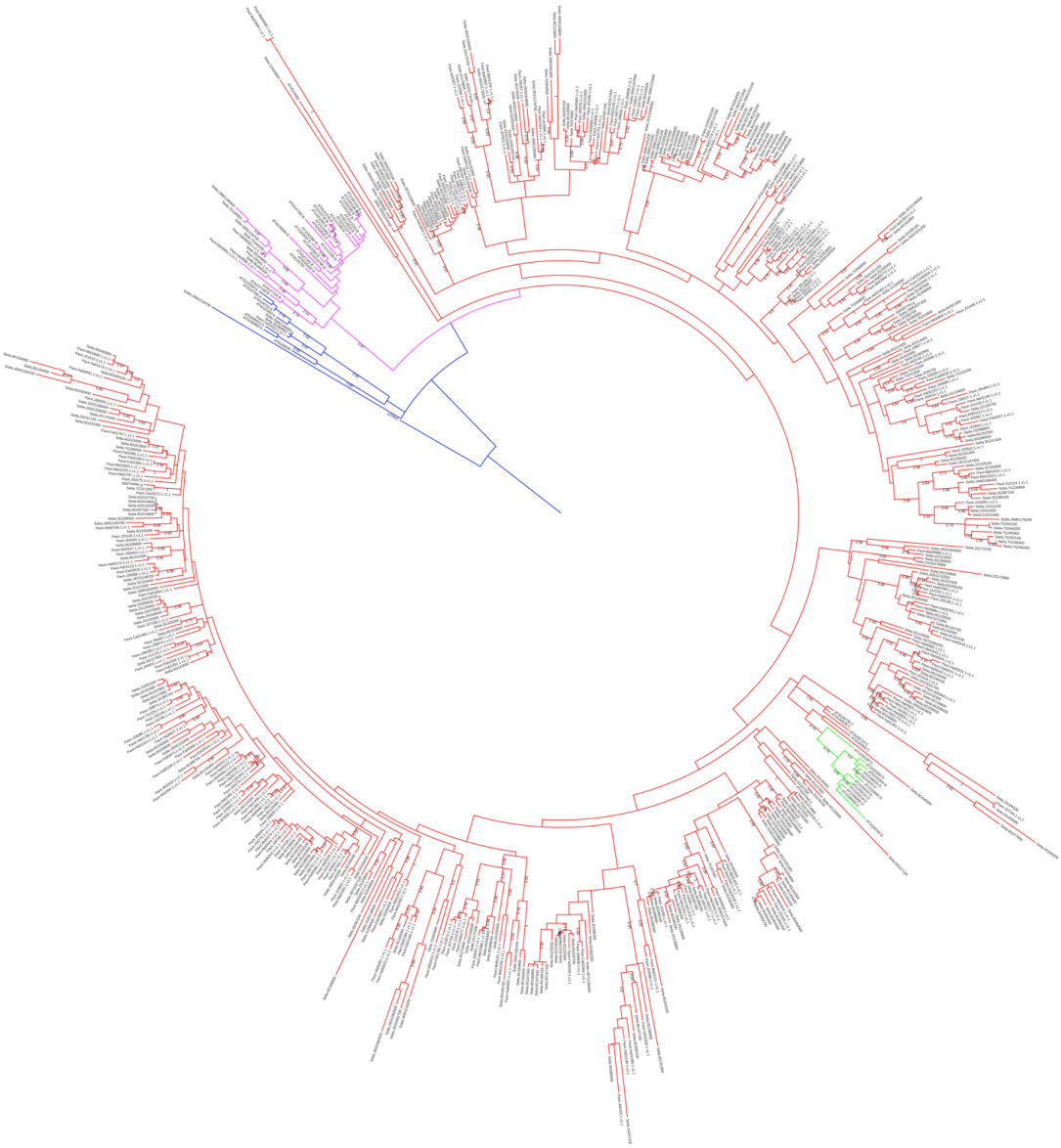


Supplementary Figure 3.2 Synteny between selected foxtail millet and rice chromosomes, illustrating orthologous relationships of CNL genes. Comparisons are

shown for substantial chromosomal inversions and duplications. Maps are arranged by foxtail millet chromosomes (Si) in rows, with rice chromosomes (Os) being individually labeled.



Supplementary Figure 3.3 Density plot of chromosomal R-gene locations for *H. vulgare*, *O. sativa*, and *S. italica*, labeled as Hv, Os, and Si, respectively.



Supplementary Figure 3.4 Maximum-likelihood phylogenetic analysis of the NB-ARC amino acid sequences of R-genes in *Setaria italica* (Seita), *Sorghum bicolor* (Sobic), *Panicum virgatum* (Pavir), and *Arabidopsis thaliana* (AT) using the JTT+G+F model and 100 bootstrap replicates, rooted on the outgroup p25941 from *Streptomyces coelicolor*. CNL clades A, B, C, and D are shown in blue, pink, red, and green, respectively.

Supplementary Table 3.1 InterProScan and GO annotation identities and descriptions for the foxtail millet CNL sequences, accessed from the Biomart function of Ensembl Genomes.

Annotation Name:	Annotation ID:	Accessions:									
AAA+ ATPase domain	IPR003593	Seita.7G250100	Seita.6G014500	Seita.6G014700	Seita.2G315000	Seita.5G158300	Seita.3G387100	Seita.8G039400	Seita.7G164900	Seita.8G087200	Seita.6G014600
Armadillo-like helical	IPR011989	Seita.1G043600									
Armadillo-type fold	IPR016024	Seita.1G043600									
DNA-binding WRKY	IPR003657	Seita.2G175200									
FNIP	IPR008615	Seita.8G242600									
Leucine-rich repeat	IPR001611, IPR032675, IPR006553, and IPR003591	Seita.8G029500	Si027449m.g	Seita.8G208200	Seita.4G244000	Seita.6G233400	Seita.9G134600	Seita.3G241300	Seita.9G182800	Seita.3G278200	Seita.8G103500
		Seita.9G224800	Seita.8G048800	Seita.2G178600	Seita.6G233300	Seita.7G311300	Seita.6G232200	Seita.5G344600	Seita.2G062100	Seita.6G232800	Seita.7G164900
		Seita.2G172700	Seita.7G165200	Seita.8G133400	Seita.8G191900	Seita.7G074000	Seita.9G466500	Seita.8G100000	Seita.6G235800	Seita.2G075000	Seita.3G393500
		Seita.8G185700	Seita.3G333400	Seita.6G229700	Seita.1G126700	Seita.8G217800	Seita.7G288400	Seita.5G046800	Seita.8G161100	Seita.9G375200	Seita.2G103300
		Seita.8G089100	Seita.6G023500	Seita.3G342000	Seita.2G078700	Seita.5G036400	Seita.7G250100	Seita.6G023600	Seita.7G066800	Seita.8G088100	Seita.4G284200
		Seita.8G181000	Seita.1G072100	Seita.7G060600	Seita.8G124300	Seita.8G184200	Si005074m.g	Seita.8G064700	Seita.3G107100	Seita.3G369800	Seita.3G395700
		Seita.4G126400	Seita.8G039500	Seita.7G245900	Seita.2G169100	Seita.8G183100	Seita.8G249100	Seita.6G229300	Seita.3G195000	Seita.8G200100	Seita.7G306800
		Seita.8G167500	Seita.8G236700	Seita.2G172800	Seita.8G087300	Seita.4G250500	Seita.3G274400	Seita.8G187100	Seita.3G207600	Seita.8G202300	Seita.8G184000
		Seita.8G166700	Seita.5G074500	Seita.6G252300	Seita.9G392900	Seita.8G199600	Seita.3G400300	Seita.8G201100	Seita.5G337400	Seita.3G367600	Seita.7G043100
		Seita.8G064300	Seita.4G213400	Seita.2G378800	Seita.1G166800	Seita.8G162600	Seita.3G325300	Seita.6G014500	Seita.3G350300	Seita.3G400200	Seita.6G220900
		Seita.7G044200	Seita.7G014800	Seita.5G053900	Seita.1G053100	Seita.2G128800	Seita.8G088200	Seita.2G103400	Seita.7G241600	Seita.8G146600	Seita.7G241500
		Seita.7G242200	Seita.5G055400	Seita.2G055800	Seita.6G014700	Seita.9G296900	Seita.3G338500	Seita.2G008500	Seita.1G098800	Seita.8G181900	Seita.2G246400
		Seita.1G191600	Seita.8G247800	Seita.2G175100	Seita.2G076700	Seita.8G182600	Seita.8G050000	Seita.3G406300	Seita.3G393700	Seita.8G184900	Seita.8G090200
		Seita.7G246200	Seita.3G100200	Seita.2G335000	Seita.8G199500	Seita.8G130400	Seita.4G214400	Seita.9G374900	Seita.3G241400	Seita.2G315000	Seita.3G107300
		Seita.5G353600	Seita.8G244000	Seita.9G298900	Seita.6G021300	Seita.3G396300	Seita.4G288900	Seita.2G009000	Seita.8G234200	Seita.5G158300	Seita.7G246400
		Seita.4G067000	Seita.2G011700	Seita.2G175200	Seita.2G294100	Seita.6G017100	Seita.9G181300	Seita.2G056400	Seita.8G089500	Seita.7G003900	Seita.1G006400

		Seita.8G208400	Seita.3G368000	Seita.8G242800	Seita.2G055500	Seita.4G126900	Seita.8G194100	Seita.1G126800	Seita.5G231000	Seita.6G014800	Seita.8G235200
		Seita.2G050400	Seita.8G249300	Seita.8G242600	Seita.6G229600	Seita.3G387100	Seita.6G229200	Seita.3G221500	Seita.8G198300	Seita.8G192200	Seita.8G039400
		Seita.8G184600	Seita.8G090100	Seita.8G243100	Seita.8G195900	Seita.J025700	Seita.8G065300	Seita.8G167100	Seita.8G006800	Seita.1G126600	Seita.8G200400
		Seita.8G064200	Seita.8G049900	Seita.3G406200	Seita.6G233200	Seita.5G103000	Seita.8G162700	Seita.3G388700	Seita.8G064900	Seita.8G087200	Seita.8G183400
		Seita.3G317600	Seita.8G089000	Seita.5G230900	Seita.1G011600	Seita.2G056600	Seita.5G435000	Seita.3G333300	Seita.6G092200	Seita.2G057000	Seita.1G011100
		Seita.8G089200	Seita.3G402500	Seita.4G243900	Seita.8G088300	Seita.8G064400	Seita.8G077900	Seita.1G010400	Seita.5G432300	Seita.4G244400	Seita.9G185000
		Seita.6G014600	Seita.3G099900	Seita.2G169300	Seita.8G166900	Seita.7G187000	Seita.8G155500	Seita.8G167300	Seita.6G143000	Seita.8G088900	Seita.4G127000
		Seita.2G335100	Seita.9G419800								
Mannose-binding lectin	IPR001229	Seita.8G124300									
		Seita.2G075000	Seita.3G393500	Seita.8G185700	Seita.3G333400	Seita.6G229700	Seita.8G029500	Seita.1G126700	Seita.8G217800	Seita.7G288400	Si027449m.g
		Seita.5G046800	Seita.8G208200	Seita.8G161100	Seita.9G375200	Seita.2G103300	Seita.8G089100	Seita.6G023500	Seita.3G342000	Seita.2G078700	Seita.5G036400
		Seita.9G549300	Seita.7G250100	Seita.2G176700	Seita.6G023600	Seita.7G066800	Seita.8G088100	Seita.4G284200	Seita.8G181000	Seita.1G072100	Seita.7G060600
		Seita.8G124300	Seita.4G244000	Seita.8G184200	Si005074m.g	Seita.8G064700	Seita.3G107100	Seita.3G369800	Seita.3G395700	Seita.4G126400	Seita.8G039500
		Seita.7G245900	Seita.2G169100	Seita.8G183100	Seita.8G249100	Seita.6G229300	Seita.3G195000	Seita.8G200100	Seita.7G306800	Seita.8G167500	Seita.8G236700
		Seita.7G234900	Seita.2G172800	Seita.8G087300	Seita.4G250500	Seita.3G396100	Seita.3G274400	Seita.8G187100	Seita.3G207600	Seita.8G202300	Seita.6G233400
		Seita.8G184000	Seita.8G166700	Seita.5G074500	Seita.6G252300	Seita.9G392900	Seita.8G199600	Seita.3G400300	Seita.8G201100	Seita.5G337400	Seita.3G367600
		Seita.7G043100	Seita.8G064300	Seita.4G213400	Seita.2G378800	Seita.1G166800	Seita.8G162600	Seita.3G325300	Seita.6G014500	Seita.3G350300	Seita.3G400200
		Seita.6G220900	Seita.7G044200	Seita.7G014800	Seita.5G053900	Seita.1G053100	Seita.8G097100	Seita.2G128800	Seita.8G088200	Seita.2G103400	Seita.7G241600
		Seita.9G134600	Seita.8G146600	Seita.7G241500	Seita.7G242200	Seita.5G055400	Seita.2G055800	Seita.6G014700	Seita.9G296900	Seita.3G338500	Seita.3G241300
NB-ARC	IPR002182	Seita.2G008500	Seita.4G035100	Seita.1G098800	Seita.9G182800	Seita.8G181900	Seita.2G246400	Seita.1G191600	Seita.8G247800	Seita.2G175100	Seita.2G076700
		Seita.8G182600	Seita.3G278200	Seita.8G103500	Seita.8G050000	Seita.3G406300	Seita.3G393700	Seita.8G184900	Seita.8G090200	Seita.7G246200	Seita.3G100200
		Seita.2G335000	Seita.8G199500	Seita.8G130400	Seita.4G214400	Seita.9G374900	Seita.3G241400	Seita.2G315000	Seita.3G107300	Seita.5G353600	Seita.8G244000
		Seita.9G298900	Seita.9G224800	Seita.6G021300	Seita.3G396300	Seita.4G288900	Seita.2G009000	Seita.8G234200	Seita.5G158300	Seita.7G246400	Seita.4G067000
		Seita.8G123800	Seita.2G011700	Seita.8G048800	Seita.2G175200	Seita.2G294100	Seita.6G017100	Seita.9G181300	Seita.2G056400	Seita.8G089500	Seita.7G003900
		Seita.1G006400	Seita.8G208400	Seita.3G368000	Seita.8G242800	Seita.2G055500	Seita.4G126900	Seita.2G178600	Seita.8G194100	Seita.6G233300	Seita.1G126800
		Seita.5G231000	Seita.6G014800	Seita.8G235200	Seita.2G050400	Seita.8G249300	Seita.8G242600	Seita.6G229600	Seita.7G311300	Seita.3G387100	Seita.6G229200
		Seita.3G221500	Seita.6G232200	Seita.8G198300	Seita.8G192200	Seita.5G344600	Seita.8G039400	Seita.8G184600	Seita.8G090100	Seita.8G243100	Seita.2G062100
		Seita.8G195900	Seita.6G232800	Seita.7G164900	Seita.J025700	Seita.8G065300	Seita.8G167100	Seita.8G006800	Seita.1G126600	Seita.8G200400	Seita.8G064200
		Seita.8G049900	Seita.3G406200	Seita.6G233200	Seita.5G103000	Seita.8G162700	Seita.3G388700	Seita.8G064900	Seita.8G087200	Seita.2G172700	Seita.8G183400
		Seita.3G317600	Seita.7G165200	Seita.8G089000	Seita.5G230900	Seita.8G133400	Seita.1G011600	Seita.2G056600	Seita.1G043600	Seita.5G435000	Seita.3G333300
		Seita.6G092	Seita.2G057	Seita.1G011	Seita.8G089	Seita.8G191	Seita.3G402	Seita.7G074	Seita.4G243	Seita.8G088	Seita.9G466

		200	000	100	200	900	500	000	900	300	500
		Seita.8G064400	Seita.8G077900	Seita.1G010400	Seita.5G432300	Seita.4G244400	Seita.9G185000	Seita.6G014600	Seita.3G099900	Seita.2G169300	Seita.8G100000
		Seita.8G166900	Seita.7G187000	Seita.8G155500	Seita.7G004000	Seita.8G167300	Seita.6G235800	Seita.6G143000	Seita.8G088900	Seita.4G127000	Seita.2G335100
		Seita.6G017500	Seita.9G419800								
No apical meristem (NAM) protein	IPR003441	Seita.8G039400									
Phospholipase C, phosphatidylinositol-specific, Y domain	IPR001711	Seita.3G100200									
		Seita.2G075000	Seita.3G393500	Seita.8G185700	Seita.3G333400	Seita.6G229700	Seita.8G029500	Seita.1G126700	Seita.8G217800	Seita.7G288400	Si027449m.g
		Seita.5G046800	Seita.8G208200	Seita.8G161100	Seita.9G375200	Seita.2G103300	Seita.8G089100	Seita.6G023500	Seita.3G342000	Seita.2G078700	Seita.5G036400
		Seita.9G549300	Seita.7G250100	Seita.2G176700	Seita.6G023600	Seita.7G066800	Seita.8G088100	Seita.4G284200	Seita.8G181000	Seita.1G072100	Seita.7G060600
		Seita.8G124300	Seita.4G244000	Seita.8G184200	Si005074m.g	Seita.8G064700	Seita.3G107100	Seita.3G369800	Seita.3G395700	Seita.4G126400	Seita.8G039500
		Seita.7G245900	Seita.2G169100	Seita.8G183100	Seita.8G249100	Seita.6G229300	Seita.3G195000	Seita.8G200100	Seita.7G306800	Seita.8G167500	Seita.8G236700
		Seita.7G234900	Seita.2G172800	Seita.8G087300	Seita.4G250500	Seita.3G396100	Seita.3G274400	Seita.8G187100	Seita.3G207600	Seita.8G202300	Seita.6G233400
		Seita.8G184000	Seita.8G166700	Seita.5G074500	Seita.6G252300	Seita.9G392900	Seita.8G199600	Seita.3G400300	Seita.8G201100	Seita.5G337400	Seita.3G367600
		Seita.7G043100	Seita.8G064300	Seita.4G213400	Seita.2G378800	Seita.1G166800	Seita.8G162600	Seita.3G325300	Seita.6G014500	Seita.3G350300	Seita.3G400200
		Seita.6G220900	Seita.7G044200	Seita.7G014800	Seita.5G053900	Seita.1G053100	Seita.8G097100	Seita.2G128800	Seita.8G088200	Seita.2G103400	Seita.7G241600
		Seita.9G134600	Seita.8G146600	Seita.7G241500	Seita.7G242200	Seita.5G055400	Seita.2G055800	Seita.6G014700	Seita.9G296900	Seita.3G338500	Seita.3G241300
		Seita.2G008500	Seita.4G035100	Seita.1G098800	Seita.9G182800	Seita.8G181900	Seita.2G246400	Seita.1G191600	Seita.8G247800	Seita.2G175100	Seita.2G076700
		Seita.8G182600	Seita.3G278200	Seita.8G103500	Seita.8G050000	Seita.3G406300	Seita.3G393700	Seita.8G184900	Seita.8G090200	Seita.7G246200	Seita.3G100200
P-loop containing nucleoside triphosphate hydrolase	IPR027417	Seita.2G335000	Seita.8G199500	Seita.8G130400	Seita.4G214400	Seita.9G374900	Seita.3G241400	Seita.2G315000	Seita.3G107300	Seita.5G353600	Seita.8G244000
		Seita.9G298900	Seita.9G224800	Seita.6G021300	Seita.3G396300	Seita.4G288900	Seita.2G009000	Seita.8G234200	Seita.5G158300	Seita.7G246400	Seita.4G067000
		Seita.8G123800	Seita.2G011700	Seita.8G048800	Seita.2G175200	Seita.2G294100	Seita.6G017100	Seita.9G181300	Seita.2G056400	Seita.8G089500	Seita.7G003900
		Seita.1G006400	Seita.8G208400	Seita.3G368000	Seita.8G242800	Seita.2G055500	Seita.4G126900	Seita.2G178600	Seita.8G194100	Seita.6G233300	Seita.1G126800
		Seita.5G231000	Seita.6G014800	Seita.8G235200	Seita.2G050400	Seita.8G249300	Seita.8G242600	Seita.6G229600	Seita.7G311300	Seita.3G387100	Seita.6G229200
		Seita.3G221500	Seita.6G232200	Seita.8G198300	Seita.8G192200	Seita.5G344600	Seita.8G039400	Seita.8G184600	Seita.8G090100	Seita.8G243100	Seita.2G062100
		Seita.8G195900	Seita.6G232800	Seita.7G164900	Seita.J025700	Seita.8G065300	Seita.8G167100	Seita.8G006800	Seita.1G126600	Seita.8G200400	Seita.8G064200
		Seita.8G049900	Seita.3G406200	Seita.6G233200	Seita.5G103000	Seita.8G162700	Seita.3G388700	Seita.8G064900	Seita.8G087200	Seita.2G172700	Seita.8G183400
		Seita.3G317600	Seita.7G165200	Seita.8G089000	Seita.5G230900	Seita.8G133400	Seita.1G011600	Seita.2G056600	Seita.1G043600	Seita.5G435000	Seita.3G333300
		Seita.6G092200	Seita.2G057000	Seita.1G011100	Seita.8G089200	Seita.8G191900	Seita.3G402500	Seita.7G074000	Seita.4G243900	Seita.8G088300	Seita.9G466500
		Seita.8G064400	Seita.8G077900	Seita.1G010400	Seita.5G432300	Seita.4G244400	Seita.9G185000	Seita.6G014600	Seita.3G099900	Seita.2G169300	Seita.8G100000
		Seita.8G166900	Seita.7G187000	Seita.8G155500	Seita.7G004000	Seita.8G167300	Seita.6G235800	Seita.6G143000	Seita.8G088900	Seita.4G127000	Seita.2G335100
		Seita.6G017	Seita.9G419								

		Seita.4G035500	Seita.8G100800
Powdery mildew resistance protein, RPW8 domain	IPR008808	Seita.3G369800	
Protein kinase, ATP binding site	IPR017441	Seita.4G035100	Seita.8G100000
Protein kinase, catalytic domain	IPR000719	Seita.4G035100	Seita.8G100000
Protein kinase-like domain	IPR011009	Seita.4G035100	Seita.8G100000
Serine/threonine- / dual specificity protein kinase, catalytic domain	IPR002290	Seita.4G035100	Seita.8G100000
Serine/threonine-protein kinase, active site	IPR008271	Seita.4G035100	Seita.8G100000
Tetratricopeptide repeat	IPR019734	Seita.8G133400	
Zinc finger, BED-type predicted	IPR003656	Seita.7G242200	Seita.J025700
Zinc finger, C2H2	IPR007087	Seita.1G053100	
ADP binding	GO:0043531	Seita.7G250100 Seita.1G043600 Seita.9G182800 Seita.2G062100 Seita.2G075000 Seita.9G375200 Seita.4G284200 Seita.3G395700 Seita.7G306800 Seita.8G184000 Seita.7G043100 Seita.7G044200 Seita.7G242200 Seita.8G247800 Seita.3G100200 Seita.9G298900 Seita.6G017100 Seita.4G126900 Seita.3G221500 Seita.8G006800 Seita.8G064	Seita.6G014500 Seita.2G175200 Seita.3G278200 Seita.6G232800 Seita.3G393500 Seita.2G103300 Seita.8G181000 Seita.4G126400 Seita.8G167500 Seita.8G166700 Seita.8G064300 Seita.7G014800 Seita.5G055400 Seita.2G175100 Seita.2G335000 Seita.6G021300 Seita.9G181300 Seita.8G194100 Seita.8G198300 Seita.1G126600 Seita.8G183	Seita.6G014700 Seita.8G242600 Seita.8G103500 Seita.2G172700 Seita.8G185700 Seita.8G089100 Seita.1G072100 Seita.8G039500 Seita.8G236700 Seita.5G074500 Seita.4G213400 Seita.5G053900 Seita.2G055800 Seita.2G076700 Seita.8G199500 Seita.3G396300 Seita.2G056400 Seita.1G126800 Seita.8G192200 Seita.8G200400 Seita.3G317	Seita.2G315000 Seita.8G029500 Seita.9G224800 Seita.7G165200 Seita.3G333400 Seita.6G023500 Seita.7G060600 Seita.7G245900 Seita.8G236800 Seita.6G252300 Seita.2G378800 Seita.2G128800 Seita.9G296900 Seita.2G076600 Seita.8G130400 Seita.4G288900 Seita.8G089500 Seita.5G231000 Seita.8G184600 Seita.8G064200 Seita.8G089	Seita.5G158300 Si027449m.g Seita.8G048800 Seita.8G133400 Seita.6G229700 Seita.8G124300 Seita.2G169100 Seita.8G087300 Seita.9G392900 Seita.1G166800 Seita.8G050000 Seita.2G009000 Seita.7G003900 Seita.6G014800 Seita.8G090100 Seita.8G049900 Seita.5G230	Seita.3G387100 Seita.8G208200 Seita.2G178600 Seita.8G191900 Seita.1G126700 Seita.2G078700 Seita.3G342000 Seita.8G184200 Seita.8G183100 Seita.4G250500 Seita.8G199600 Seita.8G162600 Seita.3G406300 Seita.9G374900 Seita.8G234200 Seita.1G006400 Seita.8G235200 Seita.8G243100 Seita.3G406200 Seita.1G011	Seita.8G039400 Seita.4G244000 Seita.6G233300 Seita.7G074000 Seita.8G217800 Seita.5G036400 Si005074m.g Seita.8G249100 Seita.3G274400 Seita.3G400300 Seita.3G325300 Seita.2G103400 Seita.1G098800 Seita.3G393700 Seita.7G246400 Seita.8G208400 Seita.8G195900 Seita.6G233200 Seita.2G056	Seita.7G164900 Seita.6G233400 Seita.7G311300 Seita.9G466500 Seita.7G288400 Seita.6G023600 Seita.8G064700 Seita.6G229300 Seita.8G187100 Seita.8G201100 Seita.3G350300 Seita.7G241600 Seita.8G181900 Seita.8G184900 Seita.4G067000 Seita.3G368000 Seita.2G050400 Seita.J025700 Seita.5G103000 Seita.5G435	Seita.8G087200 Seita.9G134600 Seita.6G232200 Seita.8G100000 Seita.5G046800 Seita.7G066800 Seita.3G107100 Seita.3G195000 Seita.3G207600 Seita.5G337400 Seita.3G400200 Seita.8G146600 Seita.2G246400 Seita.8G090200 Seita.2G011700 Seita.8G242800 Seita.6G229600 Seita.8G065300 Seita.8G162700 Seita.3G333	Seita.6G014600 Seita.3G241300 Seita.5G344600 Seita.6G235800 Seita.8G161100 Seita.8G088100 Seita.3G369800 Seita.8G200100 Seita.8G202300 Seita.3G367600 Seita.6G220900 Seita.7G241500 Seita.1G191600 Seita.7G246200 Seita.8G244000 Seita.2G294100 Seita.2G055500 Seita.6G229200 Seita.8G167100 Seita.3G388700 Seita.6G092

		900	400	600	000	900	600	600	000	300	200
		Seita.2G057000	Seita.1G011100	Seita.8G089200	Seita.3G402500	Seita.4G243900	Seita.8G088300	Seita.8G064400	Seita.8G077900	Seita.1G010400	Seita.5G432300
		Seita.4G244400	Seita.9G185000	Seita.3G099900	Seita.2G169300	Seita.8G166900	Seita.7G187000	Seita.8G155500	Seita.8G167300	Seita.6G143000	Seita.8G088900
		Seita.4G127000	Seita.2G335100	Seita.9G419800	Seita.9G549300	Seita.2G176700	Seita.7G234900	Seita.3G396100	Seita.8G097100	Seita.4G035100	Seita.8G123800
		Seita.7G004000	Seita.6G017500								
ATP binding	GO:0005524	Seita.7G250100	Seita.6G014500	Seita.6G014700	Seita.2G315000	Seita.5G158300	Seita.3G387100	Seita.8G039400	Seita.7G164900	Seita.8G087200	Seita.6G014600
		Seita.8G100000	Seita.8G039500	Seita.8G184000	Seita.7G246200	Seita.6G021300	Seita.2G009000	Seita.7G246400	Seita.6G014800	Seita.2G176700	Seita.4G035100
binding	GO:0005488	Seita.1G043600									
		Seita.7G250100	Seita.6G014500	Seita.6G014700	Seita.2G315000	Seita.5G158300	Seita.3G387100	Seita.8G039400	Seita.7G164900	Seita.8G087200	Seita.6G014600
		Seita.1G043600	Seita.2G175200	Seita.8G242600	Seita.8G029500	Si027449m.g	Seita.8G208200	Seita.4G244000	Seita.6G233400	Seita.9G134600	Seita.3G241300
		Seita.9G182800	Seita.3G278200	Seita.8G103500	Seita.9G224800	Seita.8G048800	Seita.2G178600	Seita.6G233300	Seita.7G311300	Seita.6G232200	Seita.5G344600
		Seita.2G062100	Seita.6G232800	Seita.2G172700	Seita.7G165200	Seita.8G133400	Seita.8G191900	Seita.7G074000	Seita.9G466500	Seita.8G100000	Seita.6G235800
		Seita.2G075000	Seita.3G393500	Seita.8G185700	Seita.3G333400	Seita.6G229700	Seita.1G126700	Seita.8G217800	Seita.7G288400	Seita.5G046800	Seita.8G161100
		Seita.9G375200	Seita.2G103300	Seita.8G089100	Seita.6G023500	Seita.3G342000	Seita.2G078700	Seita.5G036400	Seita.6G023600	Seita.7G066800	Seita.8G088100
		Seita.4G284200	Seita.8G181000	Seita.1G072100	Seita.7G060600	Seita.8G124300	Seita.8G184200	Si005074m.g	Seita.8G064700	Seita.3G107100	Seita.3G369800
		Seita.3G395700	Seita.4G126400	Seita.8G039500	Seita.7G245900	Seita.2G169100	Seita.8G183100	Seita.8G249100	Seita.6G229300	Seita.3G195000	Seita.8G200100
		Seita.7G306800	Seita.8G167500	Seita.8G236700	Seita.2G172800	Seita.8G087300	Seita.4G250500	Seita.3G274400	Seita.8G187100	Seita.3G207600	Seita.8G202300
		Seita.8G184000	Seita.8G166700	Seita.5G074500	Seita.6G252300	Seita.9G392900	Seita.8G199600	Seita.3G400300	Seita.8G201100	Seita.5G337400	Seita.3G367600
		Seita.7G043100	Seita.8G064300	Seita.4G213400	Seita.2G378800	Seita.1G166800	Seita.8G162600	Seita.3G325300	Seita.3G350300	Seita.3G400200	Seita.6G220900
defense response	GO:0006952	Seita.7G044200	Seita.7G014800	Seita.5G053900	Seita.1G053100	Seita.2G128800	Seita.8G088200	Seita.2G103400	Seita.7G241600	Seita.8G146600	Seita.7G241500
		Seita.7G242200	Seita.5G055400	Seita.2G055800	Seita.9G296900	Seita.3G338500	Seita.2G008500	Seita.1G098800	Seita.8G181900	Seita.2G246400	Seita.1G191600
		Seita.8G247800	Seita.2G175100	Seita.2G076700	Seita.8G182600	Seita.8G050000	Seita.3G406300	Seita.3G393700	Seita.8G184900	Seita.8G090200	Seita.7G246200
		Seita.3G100200	Seita.2G335000	Seita.8G199500	Seita.8G130400	Seita.4G214400	Seita.9G374900	Seita.3G241400	Seita.3G107300	Seita.5G353600	Seita.8G244000
		Seita.9G298900	Seita.6G021300	Seita.3G396300	Seita.4G288900	Seita.2G009000	Seita.8G234200	Seita.7G246400	Seita.4G067000	Seita.2G011700	Seita.2G294100
		Seita.6G017100	Seita.9G181300	Seita.2G056400	Seita.8G089500	Seita.7G003900	Seita.1G006400	Seita.8G208400	Seita.3G368000	Seita.8G242800	Seita.2G055500
		Seita.4G126900	Seita.8G194100	Seita.1G126800	Seita.5G231000	Seita.6G014800	Seita.8G235200	Seita.2G050400	Seita.8G249300	Seita.6G229600	Seita.6G229200
		Seita.3G221500	Seita.8G198300	Seita.8G192200	Seita.8G184600	Seita.8G090100	Seita.8G243100	Seita.8G195900	Seita.J025700	Seita.8G065300	Seita.8G167100
		Seita.8G006800	Seita.1G126600	Seita.8G200400	Seita.8G064200	Seita.8G049900	Seita.3G406200	Seita.6G233200	Seita.5G103000	Seita.8G162700	Seita.3G388700
		Seita.8G064900	Seita.8G183400	Seita.3G317600	Seita.8G089000	Seita.5G230900	Seita.1G011600	Seita.2G056600	Seita.5G435000	Seita.3G333300	Seita.6G092200
		Seita.2G057000	Seita.1G011100	Seita.8G089200	Seita.3G402500	Seita.4G243900	Seita.8G088300	Seita.8G064400	Seita.8G077900	Seita.1G010400	Seita.5G432300
		Seita.4G244	Seita.9G185	Seita.3G099	Seita.2G169	Seita.8G166	Seita.7G187	Seita.8G155	Seita.8G167	Seita.6G143	Seita.8G088

		400	000	900	300	900	000	500	300	000	900
		Seita.4G127000	Seita.2G335100	Seita.9G419800	Seita.9G549300	Seita.2G176700	Seita.7G234900	Seita.3G396100	Seita.8G097100	Seita.4G035100	Seita.8G123800
		Seita.7G004000	Seita.6G017500								
DNA binding	GO:0003677	Seita.8G039400	Seita.7G242200	Seita.J025700							
intracellular signal transduction	GO:0035556	Seita.3G100200									
lipid metabolic process	GO:0006629	Seita.3G100200									
metal ion binding	GO:0046872	Seita.1G053100									
nucleoside-triphosphatase activity	GO:0017111	Seita.7G250100	Seita.6G014500	Seita.6G014700	Seita.2G315000	Seita.5G158300	Seita.3G387100	Seita.8G039400	Seita.7G164900	Seita.8G087200	Seita.6G014600
		Seita.8G039500	Seita.8G184000	Seita.7G246200	Seita.6G021300	Seita.2G009000	Seita.7G246400	Seita.6G014800	Seita.2G176700		
nucleotide binding	GO:0000166	Seita.7G250100	Seita.6G014500	Seita.6G014700	Seita.2G315000	Seita.5G158300	Seita.3G387100	Seita.8G039400	Seita.7G164900	Seita.8G087200	Seita.6G014600
		Seita.8G039500	Seita.8G184000	Seita.7G246200	Seita.6G021300	Seita.2G009000	Seita.7G246400	Seita.6G014800	Seita.2G176700	Seita.4G035100	
nucleus	GO:0005634	Seita.8G039400									
phosphatidylinositol phospholipase C activity	GO:0004435	Seita.3G100200									
protein binding	GO:0005515	Seita.7G164900	Seita.8G029500	Si027449m.g	Seita.8G208200	Seita.4G244000	Seita.6G233400	Seita.9G134600	Seita.3G241300	Seita.9G182800	Seita.3G278200
		Seita.8G103500	Seita.9G224800	Seita.8G048800	Seita.2G178600	Seita.6G233300	Seita.7G311300	Seita.6G232200	Seita.5G344600	Seita.2G062100	Seita.6G232800
		Seita.2G172700	Seita.7G165200	Seita.8G133400	Seita.8G191900	Seita.7G074000	Seita.9G466500	Seita.8G100000	Seita.6G235800		
protein kinase activity	GO:0004672	Seita.8G100000	Seita.4G035100								
protein phosphorylation	GO:0006468	Seita.8G100000	Seita.4G035100								
protein serine/threonine kinase activity	GO:0004674	Seita.8G100000	Seita.4G035100								
regulation of transcription, DNA-templated	GO:0006355	Seita.8G039400	Seita.2G175200								
sequence-specific DNA binding	GO:0043565	Seita.2G175200									
signal transduction	GO:0007165	Seita.3G100200									
transcription factor activity, sequence-specific DNA binding	GO:0003700	Seita.2G175200									
transcription, DNA-templated	GO:0006351	Seita.8G039400									
transferase activity, transferring phosphorus-containing groups	GO:0016772	Seita.8G100000	Seita.4G035100								

Supplementary Table 3.2 Homology generated from BLAST results for the clustered and unclustered foxtail millet sequences.

Clustered Genes						Non-Clustered Genes			
Accession	Homolog	Cluster	Accession	Homolog	Cluster	Accession	Homolog	Accession	Homolog
Seita.2G008500	RPM1	2_1	Seita.7G003900	RPM1	7_1	Seita.5G055400	RGA4	Si027449m.g	RPM1
Seita.2G009000	RPM1	2_1	Seita.7G004000	RGA1	7_1	Seita.5G055400	At3g14460	Seita.8G195900	RGA4
Seita.2G103300	RPM1	2_4	Seita.7G165200	RPS2	7_2	Seita.5G055400	RGA3	Seita.8G097100	Tsn1
Seita.2G103400	RPM1	2_4	Seita.7G165200	NBS-LRR disease resistance protein homologue	7_2	Seita.5G353600	RGA3	Seita.8G130400	RDL5/RF45
Seita.2G103400	RPP13	2_4	Seita.7G165200	O1	7_2	Seita.5G036400	RPP8	Seita.8G130400	RGH1A
Seita.2G103400	RGA1	2_4	Seita.7G165200	O2	7_2	Seita.5G036400	MLA6	Seita.8G194100	Yr10/Mla1
Seita.2G103400	MLA6	2_4	Seita.7G164900	RPS2	7_2	Seita.5G435000	RPP13	Seita.8G161100	RGA1
Seita.2G172700	RGA3	2_6	Seita.7G164900	O1	7_2	Seita.5G053900	RPM1	Seita.8G161100	powdery mildew resistance protein PM3b
Seita.2G172800	RGA4	2_6	Seita.7G164900	O2	7_2	Seita.5G074500	Yr10/Mla1	Seita.8G187100	RPM1
Seita.2G172800	Pc protein A	2_6	Seita.7G242200	RGA4	7_3	Si005074m.g	Rp1	Seita.8G187100	RPP13
Seita.2G172800	Pc protein C	2_6	Seita.7G242200	B0809H07.6	7_3	Si005074m.g	Pi37	Seita.8G077900	RPP13
Seita.2G172800	pollen signalling protein with adenylyl cyclase activity-like	2_6	Seita.7G242200	XA1	7_3	Seita.5G344600	RPS2	Seita.8G198300	RPM1
Seita.2G175200	RPP13	2_7	Seita.7G241500	B0809H07.6	7_3	Seita.5G344600	PIC21	Seita.8G198300	RPP13
Seita.2G175200	WRKY transcription factor 41	2_7	Seita.7G241500	XA1	7_3	Seita.5G337400	At1g50180	Seita.2G056400	BPH14-1
Seita.2G175100	RPP13	2_7	Seita.7G241600	B0809H07.6	7_3	Seita.5G337400	RDL6/RF9	Seita.2G056400	BPH14-2
Seita.3G100200	BPH14-1	3_1	Seita.7G241600	XA1	7_3	Seita.5G337400	RPP8	Seita.2G055800	BPH14-1
Seita.3G100200	BPH14-2	3_1	Seita.7G245900	RGA4	7_4	Seita.4G288900	RPP13	Seita.2G055800	powdery mildew resistance protein PM
Seita.3G099900	BPH14-1	3_1	Seita.7G245900	NBS-LRR disease resistance protein homologue	7_4	Seita.4G288900	pollen signalling protein with adenylyl cyclase activity	Seita.2G315000	RPP13
Seita.3G099900	BPH14-2	3_1	Seita.7G246200	RGA4	7_4	Seita.4G250500	blight resistance protein B149	Seita.2G056600	BPH14-1
Seita.3G107100	powdery mildew resistance protein PM	3_2	Seita.7G246400	RGA4	7_4	Seita.4G250500	blight resistance protein RGA4	Seita.2G294100	disease resistance protein At1g50180
Seita.3G107100	BPH14-2	3_2	Seita.8G039400	RPM1	8_1	Seita.4G250500	blight resistance protein SH20	Seita.2G294100	RPM1
Seita.3G107300	BPH14-1	3_2	Seita.8G039400	RPP13	8_1	Seita.4G250500	Disease resistant protein rga3	Seita.2G050400	Yr10
Seita.3G107300	BPH14-2	3_2	Seita.8G039400	Nitrate-induced NOI protein	8_1	Seita.4G214400	RPP13	Seita.2G103300	MLA6
Seita.3G241400	RPP13	3_3	Seita.8G039500	RGH1A	8_1	Seita.4G284200	RPP13	Seita.2G062100	RPM1
Seita.3G333300	RPM1	3_4	Seita.8G167300	Pi-b protein	8_10	Seita.7G250100	RGA4	Seita.2G176700	RGA2

Seita.3G393700	Rp1	3_6
Seita.3G393700	rust resistance protein	3_6
Seita.3G393700	Rp1-D	3_6
Seita.3G393500	Rp1	3_6
Seita.3G393500	Rp1-D	3_6
Seita.3G396300	RPM1	3_7
Seita.3G395700	RPM1	3_7
Seita.3G395700	RPP13	3_7
Seita.3G396100	RGA2	3_7
Seita.3G396100	NBS-LRR disease resistance protein homologue	3_7
Seita.3G400300	RGA1	3_8
Seita.3G400300	BPH14-1	3_8
Seita.3G400300	powdery mildew resistance protein PM3b	3_8
Seita.3G400200	Pm3b	3_8
Seita.3G400200	powdery mildew resistance protein PM3b	3_8
Seita.3G406200	Pm3b	3_9
Seita.3G406200	powdery mildew resistance protein PM3b	3_9
Seita.3G406300	Pm3b	3_9
Seita.3G406300	powdery mildew resistance protein PM3b	3_9
Seita.4G126400	Nbs1-ON	4_1
Seita.4G126400	Nbs3-OP	4_1
Seita.4G126400	Nbs7-75	4_1
Seita.4G126400	Nbs1-Pi2	4_1
Seita.4G126400	Pi9	4_1
Seita.4G126900	Nbs1-Pi2	4_1
Seita.4G127000	NBA5	4_1
Seita.4G244000	RPM1	4_2
Seita.4G243900	RPM1	4_2
Seita.4G244400	RPM1	4_2
Seita.5G230900	RPM1	5_1
Seita.6G014600	RPM1	6_1
Seita.8G166700	Pi-b protein	8_10
Seita.8G166900	RPM1	8_10
Seita.8G166900	Pib	8_10
Seita.8G166900	Pi-b protein	8_10
Seita.8G167500	Pi-b protein	8_10
Seita.8G167100	RPM1	8_10
Seita.8G167100	Pib	8_10
Seita.8G167100	Pi-b protein	8_10
Seita.8G182600	RGH1A	8_11
Seita.8G181000	RPM1	8_11
Seita.8G181000	LRR14	8_11
Seita.8G183100	RPM1	8_11
Seita.8G183100	LRR14	8_11
Seita.8G181900	RPP13	8_11
Seita.8G181900	Pib	8_11
Seita.8G184900	YNR1	8_14
Seita.8G184900	YNR2	8_14
Seita.8G184900	YNR3	8_14
Seita.8G184900	YNR4	8_14
Seita.8G184900	YNR5	8_14
Seita.8G192200	Yr10/Mla1	8_15
Seita.8G191900	Yr10/Mla1	8_15
Seita.8G199600	Yr10/Mla1	8_16
Seita.8G199500	Yr10/Mla1	8_16
Seita.8G201100	RGA4	8_17
Seita.8G200400	Yr10/Mla1	8_17
Seita.8G200100	Yr10/Mla1	8_17
Seita.8G050000	NBS-LRR disease resistance protein homologue	8_2
Seita.8G049900	RPH8A	8_2
Seita.8G049900	RPM1	8_2
Seita.8G049900	RPP13	8_2
Seita.7G043100	RGA4	
Seita.7G306800	rp3	
Seita.7G306800	rp3-1	
Seita.7G014800	H0215A08.1	
Seita.7G187000	RPM1	
Seita.7G311300	RGA1	
Seita.7G311300	RPP13	
Seita.7G044200	RGA4	
Seita.7G060600	RGA2	
Seita.7G288400	RGA1	
Seita.7G288400	RPP13	
Seita.6G220900	Pi-b protein	
Seita.6G232200	RGA1	
Seita.6G232200	NBS3-RDG2A	
Seita.6G232200	RDG2A	
Seita.6G235800	RGA1	
Seita.6G235800	NBS3-RDG2A	
Seita.6G021300	Lr21	
Seita.6G021300	rust resistance protein Rp1-dp8-like	
Seita.6G092200	RPM1	
Seita.6G092200	RPP13	
Seita.6G143000	MLA1	
Seita.1G011600	RGA-1	
Seita.1G011100	RGA-1	
Seita.1G053100	RPM1	
Seita.1G053100	RPP13	
Seita.1G006400	RPP13	
Seita.1G006400	Yr10	
Seita.1G072100	RPM1	
Seita.1G072100	RPP13	
Seita.1G191600	RPM1	
Seita.2G176700	RGA4	
Seita.2G176700	Pi15	
Seita.2G176700	Pi5-1	
Seita.2G378800	RPM1	
Seita.2G057000	BPH14-1	
Seita.2G057000	BPH14-2	
Seita.2G076700	Nbs1-Pi2	
Seita.2G078700	RPM1	
Seita.2G078700	MLA6	
Seita.2G075000	RPM1	
Seita.2G075000	MLA6	
Seita.2G009000	MLA6	
Seita.2G011700	Nitrate-induced NOI protein	
Seita.2G178600	blight resistance protein SH20	
Seita.2G178600	Pi5-2	
Seita.2G246400	putative ATPase	
Seita.2G246400	RPM1	
Seita.2G246400	PPR1	
Seita.2G246400	RXO1	
Seita.2G055500	BPH14-1	
Seita.2G055500	BPH14-2	
Seita.9G224800	RGA3	
Seita.9G392900	RGA4	
Seita.9G392900	XA1	
Seita.9G296900	blight resistance protein RGA4	
Seita.9G296900	blight resistance protein SH20	
Seita.9G296900	blight resistance protein T118	
Seita.9G296900	RGA3	
Seita.9G298900	RPM1	
Seita.9G466500	RPS2	
Seita.9G419800	disease resistance protein At3g14460	

Seita.6G01 4600	RGH1A	6_1	Seita.8G04 9900	RPP8	8_2	Seita.1G19 1600	RPP13	Seita.9G54 9300	RPP13
Seita.6G01 4500	RPM1	6_1	Seita.8G04 9900	Nitrate-induced NOI protein	8_2	Seita.1G01 0400	RGA-1	Seita.9G54 9300	blight resistance protein RGA3
Seita.6G01 4500	RGH1A	6_1	Seita.8G24 2600	Pm3b	8_20	Seita.J0257 00	RGA2	Seita.9G54 9300	RGA4
Seita.6G01 4800	RPM1	6_1	Seita.8G24 3100	RPM1	8_20	Seita.J0257 00	B0809H07.6	Seita.9G18 2800	RGA4
Seita.6G01 4800	RGH1A	6_1	Seita.8G24 2800	Pm3b	8_20	Seita.J0257 00	XA1		
Seita.6G01 4700	RPM1	6_1	Seita.8G24 9300	RGA2	8_21	Seita.3G27 4400	RPP13		
Seita.6G01 4700	RGH1A	6_1	Seita.8G24 9300	RGA4	8_21	Seita.3G38 8700	MLA1		
Seita.6G01 7100	Nbs1-ON	6_2	Seita.8G24 9100	RGA2	8_21	Seita.3G38 8700	Pi36		
Seita.6G01 7100	Nbs1-Pi2	6_2	Seita.8G24 9100	RGA4	8_21	Seita.3G33 8500	RPM1		
Seita.6G01 7500	NBS-LRR disease resistance protein homologue	6_2	Seita.8G06 4900	RPM1	8_3	Seita.3G31 7600	RPM1		
Seita.6G02 3600	RPM1	6_3	Seita.8G06 4400	RPM1	8_3	Seita.3G34 2000	RGC1B		
Seita.6G02 3600	RPP13	6_3	Seita.8G06 4700	RPM1	8_3	Seita.3G35 0300	RPM1		
Seita.6G02 3600	RGA1	6_3	Seita.8G06 4300	RPM1	8_3	Seita.3G27 8200	RGA3		
Seita.6G02 3500	RPM1	6_3	Seita.8G06 4300	LRR14	8_3	Seita.3G27 8200	putative blight resistance protein		
Seita.6G02 3500	RPP13	6_3	Seita.8G08 7200	MLA1	8_4	Seita.3G40 2500	powdery mildew resistance protein PM3b		
Seita.6G22 9200	RPM1	6_4	Seita.8G08 8100	RPM1	8_5	Seita.3G40 2500	Pm3b		
Seita.6G22 9200	Pib	6_4	Seita.8G08 8300	RPM1	8_5	Seita.3G22 1500	RPM1		
Seita.6G22 9200	Pi-b protein	6_4	Seita.8G08 8300	RPP8	8_5	Seita.8G18 4000	YNR1		
Seita.6G22 9600	Pi-b protein	6_4	Seita.8G08 9100	Nitrate-induced NOI protein	8_6	Seita.8G18 4000	YNR2		
Seita.6G22 9700	RPM1	6_4	Seita.8G08 8900	RPM1	8_6	Seita.8G18 4000	YNR3		
Seita.6G22 9700	Pi-b protein	6_4	Seita.8G08 9500	RPM1	8_6	Seita.8G18 4000	YNR4		
Seita.6G22 9300	Pi-b protein	6_4	Seita.8G08 9500	RPP13	8_6	Seita.8G18 4000	YNR5		
Seita.6G23 3300	RGA1	6_5	Seita.8G09 0100	RPM1	8_7	Seita.8G15 5500	RGA4		
Seita.6G23 3400	RGA1	6_5	Seita.8G09 0100	RPP13	8_7	Seita.8G20 2300	RPM1		
Seita.6G23 3400	NBS3-RDG2A	6_5	Seita.8G12 4300	RPM1	8_8	Seita.8G24 4000	Pm3b		
Seita.6G23 3400	RDG2A	6_5	Seita.8G16 2700	RGA4	8_9	Seita.8G18 3400	RPM1		
Seita.6G23 2800	RGA1	6_5	Seita.8G16 2600	RGA4	8_9	Seita.8G18 3400	LRR14		
Seita.6G23 2800	NBS3-RDG2A	6_5	Seita.9G37 4900	RGC1B	9_1	Seita.8G13 3400	Yr10		
Seita.6G23 2800	RDG2A	6_5	Seita.9G37 5200	RGC1B	9_1	Seita.8G04 8800	RGA4		

ACKNOWLEDGEMENTS

Support for this dissertation work came from the USDA-NIFA hatch Projects to M. Nepal (SD00H469-13 and SD00H659-18), South Dakota Agriculture Experiment Station, and Department of Biology & Microbiology at South Dakota State University.

REFERENCES

1. Yoshimura S, Yamanouchi U, Katayose Y, et al. Expression of Xa1, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. *Proceedings of the National Academy of Sciences.* 1998;95(4):1663-1668.
2. Wang X, Richards J, Gross T, et al. The rpg4-mediated resistance to wheat stem rust (Puccinia graminis) in barley (Hordeum vulgare) requires Rpg5, a second NBS-LRR gene, and an actin depolymerization factor. *Molecular Plant-Microbe Interactions.* 2013;26(4):407-418.
3. Sacco MA, Mansoor S, Moffett P. A RanGAP protein physically interacts with the NB‐LRR protein Rx, and is required for Rx‐mediated viral resistance. *The Plant Journal.* 2007;52(1):82-93.
4. Li J, Chen Q, Wang L, Liu J, Shang K, Hua H. Biological effects of rice harbouring Bph14 and Bph15 on brown planthopper, Nilaparvata lugens. *Pest management science.* 2011;67(5):528-534.
5. Liu S, Kandoth PK, Warren SD, et al. A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. *Nature.* 2012;492(7428):256-260.

6. Bromham L, Cowman PF, Lanfear R. Parasitic plants have increased rates of molecular evolution across all three genomes. *BMC evolutionary biology.* 2013;13(1):1.

7. Flor HH. Current status of the gene-for-gene concept. *Annual review of phytopathology.* 1971;9(1):275-296.

8. Jones JD, Dangl JL. The plant immune system. *Nature.* 2006;444(7117):323-329.

9. Van Der Biezen EA, Jones JD. Plant disease-resistance proteins and the gene-for-gene concept. *Trends in biochemical sciences.* 1998;23(12):454-456.

10. Shao F, Golstein C, Ade J, Stoutemyer M, Dixon JE, Innes RW. Cleavage of Arabidopsis PBS1 by a bacterial type III effector. *Science.* 2003;301(5637):1230-1233.

11. van der Hoorn RA, Kamoun S. From guard to decoy: a new model for perception of plant pathogen effectors. *The Plant Cell.* 2008;20(8):2009-2017.

12. Michelmore RW, Meyers BC. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Research.* 1998;8(11):1113-1130.

13. Bergelson J, Kreitman M, Stahl EA, Tian D. Evolutionary dynamics of plant R-genes. *Science.* 2001;292(5525):2281-2285.

14. McGrann GR, Stavrinides A, Russell J, et al. A trade off between mlo resistance to powdery mildew and increased susceptibility of barley to a newly important disease, Ramularia leaf spot. *Journal of experimental botany.* 2014:ert452.

15. Rivas S. Nuclear dynamics during plant innate immunity. *Plant physiology.* 2012;158(1):87-94.

16. Glazebrook J. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu Rev Phytopathol.* 2005;43:205-227.

17. Vleeshouwers VG, Oliver RP. Effectors as tools in disease resistance breeding against biotrophic, hemibiotrophic, and necrotrophic plant pathogens. *Molecular plant-microbe interactions.* 2014;27(3):196-206.

18. Lam E, Kato N, Lawton M. Programmed cell death, mitochondria and the plant hypersensitive response. *Nature.* 2001;411(6839):848-853.

19. Lamb C, Dixon RA. The oxidative burst in plant disease resistance. *Annual review of plant biology.* 1997;48(1):251-275.

20. Hématy K, Cherk C, Somerville S. Host–pathogen warfare at the plant cell wall. *Current opinion in plant biology.* 2009;12(4):406-413.

21. Barton L, Newsome SD, Chen F-H, Wang H, Guilderson TP, Bettinger RL. Agricultural origins and the isotopic identity of domestication in northern China. *Proceedings of the National Academy of Sciences.* 2009;106(14):5523-5528.

22. Zohary D, Hopf M, Weiss E. *Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin.* Oxford University Press on Demand; 2012.

23. Zhang G, Liu X, Quan Z, et al. Genome sequence of foxtail millet (Setaria italica) provides insights into grass evolution and biofuel potential. *Nature biotechnology.* 2012;30(6):549-554.

24. Li P, Brutnell TP. Setaria viridis and Setaria italica, model genetic systems for the Panicoid grasses. *Journal of Experimental Botany.* 2011;62(9):3031-3037.

25. Daverdin G, Bahri BA, Wu X, et al. Comparative relationships and chromosome evolution in switchgrass (Panicum virgatum) and its genomic model, foxtail millet (Setaria italica). *BioEnergy Research.* 2015;8(1):137-151.

26. Jia G, Huang X, Zhi H, et al. A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (Setaria italica). *Nature genetics.* 2013;45(8):957-961.

27. Vicentini A, Barber JC, Aliscioni SS, Giussani LM, Kellogg EA. The age of the grasses and clusters of origins of C4 photosynthesis. *Global Change Biology.* 2008;14(12):2963-2977.

28. Bennetzen JL, Schmutz J, Wang H, et al. Reference genome sequence of the model plant Setaria. *Nature biotechnology.* 2012;30(6):555-561.

29. Li P, Brutnell TP. Setaria viridis and Setaria italica, model genetic systems for the Panicoid grasses. *Journal of Experimental Botany.* 2011:err096.

30. Varshney RK, Koebner RM. *Model plants and crop improvement.* CRC Press; 2006.

31. Wolfe KH, Gouy M, Yang Y-W, Sharp PM, Li W-H. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proceedings of the National Academy of Sciences.* 1989;86(16):6201-6205.

32. Karthikeyan V, Gnanamanickam S. Biological control of Setaria blast (Magnaporthe grisea) with bacterial strains. *Crop Protection.* 2008;27(2):263-267.

33. Sesma A, Osbourn AE. The rice leaf blast pathogen undergoes developmental processes typical of root-infecting fungi. *Nature.* 2004;431(7008):582-586.

34. Dean RA, Talbot NJ, Ebbole DJ, et al. The genome sequence of the rice blast fungus Magnaporthe grisea. *Nature.* 2005;434(7036):980-986.

35. Wang ZX, Yano M, Yamanouchi U, et al. The Pib gene for rice blast resistance belongs to the nucleotide binding and leucine‐rich repeat class of plant disease resistance genes. *The Plant Journal.* 1999;19(1):55-64.

36. Jia Y, McAdams SA, Bryan GT, Hershey HP, Valent B. Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *The EMBO journal.* 2000;19(15):4004-4014.

37. Bryan GT, Wu K-S, Farrall L, et al. A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene Pi-ta. *The Plant Cell.* 2000;12(11):2033-2045.

38. Orbach MJ, Farrall L, Sweigard JA, Chumley FG, Valent B. A telomeric avirulence gene determines efficacy for the rice blast resistance gene Pi-ta. *The Plant Cell.* 2000;12(11):2019-2032.

39. Qu S, Liu G, Zhou B, et al. The broad-spectrum blast resistance gene Pi9 encodes a nucleotide-binding site–leucine-rich repeat protein and is a member of a multigene family in rice. *Genetics.* 2006;172(3):1901-1914.

40. Li J, Sun Y, Liu H, Wang Y, Jia Y, Xu M. Genetic variation and evolution of the Pit blast resistance locus in rice. *Genetic resources and crop evolution.* 2014;61(2):473-489.

41. Eizenga G, Agrama H, Lee F, Jia Y. Exploring genetic diversity and potential novel disease resistance genes in a collection of rice (Oryza spp.) wild relatives. *Genetic resources and crop evolution.* 2009;56(1):65-76.

42. Belay G. *Cereals and pulses.* Vol 1: PROTA; 2006.

43. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of NBS-LRR–encoding genes in Arabidopsis. *The Plant Cell Online.* 2003;15(4):809-834.

44. Nepal MP, Benson BV. CNL Disease Resistance Genes in Soybean and Their Evolutionary Divergence. *Evolutionary bioinformatics online.* 2015;11:49.

45. Andersen EJ, Ali S, Reese RN, Yen Y, Neupane S, Nepal MP. Diversity and Evolution of Disease Resistance Genes in Barley (Hordeum vulgare L.). *Evolutionary bioinformatics online.* 2016;12:99.

46. Kersey PJ, Allen JE, Christensen M, et al. Ensembl Genomes 2013: scaling up access to genome-wide data. *Nucleic acids research.* 2014;42(D1):D546-D552.

47. Goodstein DM, Shu S, Howson R, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic acids research.* 2012;40(D1):D1178-D1186.

48. Larkin MA, Blackshields G, Brown N, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23(21):2947-2948.

49. Finn RD, Clements J, Arndt W, et al. HMMER web server: 2015 update. *Nucleic acids research.* 2015:gkv397.

50. Finn RD, Bateman A, Clements J, et al. Pfam: the protein families database. *Nucleic acids research.* 2013:gkt1223.

51. Jones P, Binns D, Chang H-Y, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014:btu031.

52. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in bipolymers. Department of Computer Science and Engineering, University of California, San Diego; 1994.

53. Kearse M, Moir R, Wilson A, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012;28(12):1647-1649.

54. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and

maximum parsimony methods. *Molecular biology and evolution.* 2011;28(10):2731-2739.

55. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular biology and evolution.* 2016:msw054.

56. He J, Dai X, Zhao X. PLAN: a web platform for automating high-throughput BLAST searches and for managing and mining results. *Bmc Bioinformatics.* 2007;8(1):53.

57. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. *BMC bioinformatics.* 2009;10(1):421.

58. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. *Nature genetics.* 2000;25(1):25.

59. Yang X, Wang J. Genome-Wide Analysis of NBS-LRR Genes in Sorghum Genome Revealed Several Events Contributing to NBS-LRR Gene Evolution in Grass Species. *Evolutionary bioinformatics online.* 2016;12:9.

60. Frazier TP, Palmer NA, Xie F, et al. Identification, characterization, and gene expression analysis of nucleotide binding site (NB)-type resistance gene homologues in switchgrass. *BMC genomics.* 2016;17(1):892.

61. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics.* 2014:btu817.

62. Jupe F, Pritchard L, Etherington GJ, et al. Identification and localisation of the NB-LRR gene family within the potato genome. *BMC genomics.* 2012;13(1):75.

63. Venables WN, Smith DM, Team RC. An introduction to R-Notes on R: A programming environment for data analysis and graphics. 2018.

64. Rozas J. DNA sequence polymorphism analysis using DnaSP. *Bioinformatics for DNA sequence analysis*: Springer; 2009:337-350.

65. Soderlund C, Bomhoff M, Nelson WM. SyMAP v3. 4: a turnkey synteny system with application to plant genomes. *Nucleic acids research.* 2011:gkr123.

66. Zhao Y, Weng Q, Song J, et al. Bioinformatics Analysis of NBS-LRR Encoding Resistance Genes in Setaria italica. *Biochemical genetics.* 2016:1-17.

67. Andersen EJ, Shaw SR, Nepal MP. Identification of Disease Resistance Genes in Aegilops Tauschii Coss. (Poaceae). *Proceedings of the South Dakota Academy of Science.* 2015;94:273-287.

68. Michelmore RW, Christopoulou M, Caldwell KS. Impacts of resistance gene genetics, function, and evolution on a durable future. *Annual review of phytopathology.* 2013;51:291-319.

69. Williams SJ, Sornaraj P, deCourcy-Ireland E, et al. An autoactive mutant of the M flax rust resistance protein has a preference for binding ATP, whereas wild-type M protein binds ADP. *Molecular plant-microbe interactions.* 2011;24(8):897-906.

70. Dietrich RA, Richberg MH, Schmidt R, Dean C, Dangl JL. A novel zinc finger protein is encoded by the Arabidopsis LSD1 gene and functions as a negative regulator of plant cell death. *Cell.* 1997;88(5):685-694.

71. Rojas CM, Senthil-Kumar M, Tzin V, Mysore K. Regulation of primary plant metabolism during plant-pathogen interactions and its contribution to plant defense. *Frontiers in plant science.* 2014;5:17.

72. Jordan T, Schornack S, Lahaye T. Alternative splicing of transcripts encoding Toll-like plant resistance proteins–what's the functional relevance to innate immunity? *Trends in plant science.* 2002;7(9):392-398.

73. McHale L, Tan X, Koehl P, Michelmore RW. Plant NBS-LRR proteins: adaptable guards. *Genome biology.* 2006;7(4):1.

74. Bent AF, Kunkel BN, Dahlbeck D, et al. RPS2 of Arabidopsis thaliana: a leucine-rich repeat class of plant disease resistance genes. *SCIENCE-NEW YORK THEN WASHINGTON-*. 1994:1856-1856.

75. Kawasaki T, Nam J, Boyes DC, et al. A duplicated pair of arabidopsis RING‐finger E3 ligases contribute to the RPM1‐and RPS2‐mediated hypersensitive response. *The Plant Journal.* 2005;44(2):258-270.

76. Allen RL, Meitz JC, Baumber RE, et al. Natural variation reveals key amino acids in a downy mildew effector that alters recognition specificity by an Arabidopsis resistance gene. *Molecular plant pathology.* 2008;9(4):511-523.

77. Benson BV. *Disease Resistance Genes and their Evolutionary History in Six Plant Species*: Biology and Microbiology, South Dakota State University; 2014.

78. Devos K, Wang Z, Beales J, Sasaki T, Gale M. Comparative genetic maps of foxtail millet (Setaria italica) and rice (Oryza sativa). *Theoretical and Applied Genetics.* 1998;96(1):63-68.

79. Bouchenak‐Khelladi Y, Verboom GA, Savolainen V, Hodkinson TR. Biogeography of the grasses (Poaceae): a phylogenetic approach to reveal evolutionary history in geographical space and geological time. *Botanical Journal of the Linnean Society.* 2010;162(4):543-557.

80. Cloutier S, McCallum BD, Loutre C, et al. Leaf rust resistance gene Lr1, isolated from bread wheat (Triticum aestivum L.) is a member of the large psr567 gene family. *Plant molecular biology.* 2007;65(1-2):93-106.

81. Yahiaoui N, Srichumpa P, Dudler R, Keller B. Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene Pm3b from hexaploid wheat. *The Plant Journal.* 2004;37(4):528-538.

82. Wang G-L, Song W-Y, Ruan D-L, Sideris S, Ronald PC. The cloned gene, Xa21, confers resistance to multiple Xanthomonas oryzae pv. oryzae isolates in transgenic plants. *Mol Plant Microbe Interact.* 1996;9(9):850-855.

83. Kalia B, Wilson DL, Bowden RL, Singh RP, Gill BS. Adult plant resistance to Puccinia triticina in a geographically diverse collection of Aegilops tauschii. *Genetic Resources and Crop Evolution.* 2016:1-14.

84. Vikas V, Sivasamy M, Kumar J, et al. Stem and leaf rust resistance in wild relatives of wheat with D genome (Aegilops spp.). *Genetic resources and crop evolution.* 2014;61(4):861-874.

85. Assefa S, Fehrmann H. Resistance to wheat leaf rust in Aegilops tauschii Coss. and inheritance of resistance in hexaploid wheat. *Genetic Resources and Crop Evolution.* 2000;47(2):135-140.

86. Singrün C, Rauch P, Morgounov A, Hsam S, Zeller F. Identification of powdery mildew and leaf rust resistance genes in common wheat (Triticum aestivum L.). Wheat varieties from the Caucasus, Central and Inner Asia. *Genetic Resources and Crop Evolution.* 2004;51(4):355-370.

87. Cesari S, Thilliez G, Ribot C, et al. The rice resistance protein pair RGA4/RGA5 recognizes the Magnaporthe oryzae effectors AVR-Pia and AVR1-CO39 by direct binding. *The Plant Cell.* 2013;25(4):1463-1481.

88. Césari S, Kanzaki H, Fujiwara T, et al. The NB‐LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. *The EMBO journal.* 2014;33(17):1941-1959.

89. Bulgarelli D, Biselli C, Collins NC, et al. The CC-NB-LRR-type Rdg2a resistance gene confers immunity to the seed-borne barley leaf stripe pathogen in the absence of hypersensitive cell death. *PLoS One.* 2010;5(9):e12599.

90. Cannarozzi G, Plaza-Wüthrich S, Esfeld K, et al. Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (Eragrostis tef). *BMC genomics.* 2014;15(1):581.

91. Schlueter JA, Dixon P, Granger C, et al. Mining EST databases to resolve evolutionary events in major crop species. *Genome.* 2004;47(5):868-876.

92. Shindo C, Bernasconi G, Hardtke CS. Natural genetic variation in Arabidopsis: tools, traits and prospects for evolutionary ecology. *Annals of Botany.* 2007;99(6):1043-1054.

93. Bittner‐Eddy PD, Crute IR, Holub EB, Beynon JL. RPP13 is a simple locus in Arabidopsis thaliana for alleles that specify downy mildew resistance to different avirulence determinants in Peronospora parasitica. *The Plant Journal.* 2000;21(2):177-188.

94. Rose LE, Bittner-Eddy PD, Langley CH, Holub EB, Michelmore RW, Beynon JL. The maintenance of extreme amino acid diversity at the disease resistance gene, RPP13, in Arabidopsis thaliana. *Genetics.* 2004;166(3):1517-1527.

95. Ashfield T, Redditt T, Russell A, et al. Evolutionary relationship of disease resistance genes in soybean and Arabidopsis specific for the Pseudomonas syringae effectors AvrB and AvrRpm1. *Plant physiology.* 2014;166(1):235-251.

96. Rose L, Atwell S, Grant M, Holub EB. Parallel loss-of-function at the RPM1 bacterial resistance locus in Arabidopsis thaliana. *Frontiers in plant science.* 2012;3.

97. Nagy ED, Bennetzen JL. Pathogen corruption and site-directed recombination at a plant disease resistance gene cluster. *Genome research.* 2008;18(12):1918-1923.

98. Du B, Zhang W, Liu B, et al. Identification and characterization of Bph14, a gene conferring resistance to brown planthopper in rice. *Proceedings of the National Academy of Sciences.* 2009;106(52):22163-22168.

99. Lao S-H, Huang X-H, Huang H-J, Liu C-W, Zhang C-X, Bao Y-Y. Genomic and transcriptomic insights into the cytochrome P450 monooxygenase gene repertoire in the rice pest brown planthopper, Nilaparvata lugens. *Genomics.* 2015;106(5):301-309.

100. Aguda R, Rombach M, Im D, Shepard B. Suppression of populations of the brown planthopper, Nilaparvata lugens (Stål)(Hom.; Delphacidae) in field cages by entomogenous fungi (Deuteromycotina) on rice in Korea. *Journal of Applied Entomology.* 1987;104(1‐5):167-172.

101. Pattemore JA, Hane JK, Williams AH, Wilson BA, Stodart BJ, Ash GJ. The genome sequence of the biocontrol fungus Metarhizium anisopliae and comparative genomics of Metarhizium species. *BMC genomics.* 2014;15(1):1.

102. Andersen EJ, Nepal MP. Genetic diversity of disease resistance genes in foxtail millet (Setaria italica L.). *Plant Gene.* 2017;10:8-16.

103. Yang X, Wang J. Genome-wide analysis of NBS-LRR genes in Sorghum genome revealed several events contributing to NBS-LRR gene evolution in grass species. *Evolutionary Bioinformatics.* 2016;12:EBO. S36433.

104. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research.* 2016:gkw290.

CHAPTER 4: GENOME-WIDE IDENTIFICATION OF DISEASE RESISTANCE GENES IN *AEGILOPS TAUSCHII* COSS. (POACEAE)

This chapter is published in the journal *Proceeding of the South Dakota Academy of Sciences*:

Andersen, E.J., S.R. Shaw, and M.P. Nepal. 2015. Genome-wide Identification of Disease Resistance Genes in *Aegilops tauschii* Coss. (Poaceae). *PSDAOS*. 94:281-295

ABSTRACT

Identification of disease resistance genes (R-genes) and revealing their functions are important for understanding a plant's defense against pathogens. *Aegilops tauschii*, the contributor of wheat's D-genome, has a recently available complete genome sequence, and genome-wide identification of R-genes in this plant would give insight into the evolution of wheat resistance genes. The main objectives of this project were to identify CNL (Coiled-coil, Nucleotide-binding site, and Leucine-rich region) R-genes within the *A. tauschii* genome, and elucidate their evolutionary relationships within *Aegilops* and across the genome of two model plants--*Arabidopsis* and rice. *In silico* analyses were conductedin which known CNL genes of *Arabidopsis* and rice were used to search for their orthologs in *A. tauschii*. A total of 402 CNL resistance genes were identified within the *A. tauschii* genome and recovered three clades (A, B, and C) of *A. tauschii* CNL genes of which CNL C is the largest clade, a single member represents clade A, and clade D is entirely absent. Each of these clades was characterized by a consistent motif structure. The number of exons varied from 1 to 28 with an average number of 4.5. The majority of CNL genes were inferred to have originated by tandem

duplications, and the historical gene duplication events perhaps diversified the members in response to a unique pathogen pressure. Identification of *Aegilops* R-genes would help us understand the evolution of R-genes, particularly those located in the D-genome of wheat, and has a potential implication in creating a durable R-gene in Aegilops, wheat, and other crop species in future.

INTRODUCTION

Plant defense against pathogens involves complex signaling pathways that trigger resistance responses [1]. Such responses typically lead to a hypersensitive response, but can also include the production of anti-pathogen chemicals or cell wall fortification [2]. Hypersensitive response, in particular, is a general response that involves the programmed cell death of a section of tissue that has been infected by a pathogen to quarantine the affected area [2]. Disease resistance genes, or R-genes, encode proteins that are involved in the detection of pathogen attacks and activation of subsequent downstream plant response signaling. The R-genes occur as multigene families, and multiple models have been proposed to describe their mechanism of action. The Gene-for-Gene Model describes plants having specific dominant resistance genes that counter a corresponding pathogen avirulence genes in an evolutionary arms-race [3]. Introducing more molecular details, the Guard Model describes resistance genes bound to plant proteins and are activated when that protein is cleaved by a pathogen protein [4,5], while the Zig-Zag Model describes the pathogen evolving new avirulence genes that evade plant basal immunity [1]. Recently R-genes have been classified into eight specific groups [6]. Among them, the overwhelming majority of the R-genes fall under the NBS-LRR

type, the largest class of R-genes [7,8]. The NBS-LRR genes can be categorized into two major types based upon whether they start with a Toll Interleukin Receptor (TIR-NBS-LRR or TNL; absent in monocots) or a Coiled Coil (CC-NBS-LRR or CNL; present in all plants) [8]. Resistance genes evolve rapidly due to the high selection pressure put onto the plant population by pathogen load [9] that causes faster gene diversification [10]. This diversification is caused primarily by gene recombination and transposable elements' activities [11]. Their loss is also possible by deficient duplications and the loss of lineages, as evidenced in cucumber and watermelon genomes that contain many fewer resistance genes [12]. In addition, the evolution of R-genes occurs through a trade-off between physical, chemical, and molecular defenses in response to coevolving pathogens [2].

Increasing availability of complete genome sequences of plants at various taxonomic levels allows us to carry out comparative analyses for identification of R-genes and understanding of the evolutionary processes involved. CNL R-genes have been identified for various plant species such as papaya (6; [13], cucumber (18; [14], rice (159, 149; [15,16], *Arabidopsis* (55; [8], poplar (119; [17], *Medicago* (177; [18], soybean (188, [16,19], potato (370; [20], and are yet to be identified in *Aegilops tauschii* Coss. (Poaceae), the D-genome contributor of bread wheat (*Triticum aestivum* L.). *A. tauschii* underwent hybridization with *Triticum turgidum* several thousand years ago, forming bread wheat [21]. The objectives of this research were to identify *A. tauschii* CNL resistance genes and elucidate their evolutionary relationships within *A. tauschii* and across the genomes of *Arabidopsis* and rice, two model plant species.

MATERIALS AND METHODS

A. tauschii protein sequences were searched in the Ensembl Genomes site [22]. Previously identified *Arabidopsis* CNL resistance genes (Meyers et al. 2003) were obtained from the Phytozome database [23]. First, fifty CNL genes of *Arabidopsis* were aligned in the program ClustalW and used to construct a Hidden Markov Model to search for the entire set of *A. tauschii* protein sequences with a stringency of 0.05. The *A. tauschii* genes were uploaded into the program Geneious [24] and annotated with InterProScan [25] to identify NBARCs with the program Pfam (pfam.sanger.ac.uk) that allowed the exclusion of sequences with TIR motifs.

The protein sequences with NBARCs were used to construct a reiterative HMM to search the *A. tauschii* proteins for species-specific CNL genes at a stringency of 0.001. A total of 810 genes were identified through first HMM at a stringency of 0.05. Of these genes, 711 were determined to contain NBARCs through domain annotation with InterProScan. The reiterative HMM identified 779 genes and after removing gene duplicates, 711 of these 779 genes were determined to contain NBARCs, of which only 544 genes contained both NBARC and "DiseaseResist" domains. The NBARCs of these genes were then uploaded to MEME suite to preform MEME analysis [26] and annotate the three characteristic domains of the CNL genes, i.e. P-loop, Kinase-2, and GLPL motifs. All genes containing these three motifs were aligned using Muscle integrated in the program MEGA 6.0 [27]. *Arabidopsis* as well as rice sequences were also imported into MEGA 6.0 to make two phylogenetic trees (100 bootstrap replicates using the JTT+G Model for both trees) to look for evolutionary relationships between the genes. Exon

structure was also determined using exon information and scaffold location data from the Ensembl Genomes site. Gene exon coordinates were used in the program Fancygene v1.4 to visualize the exon-intron structure.

RESULTS AND DISCUSSION

Of the 33,928 *A. tauschii* protein sequences analyzed, 402 genes (1.2% of the genome) were identified as CNL genes. All these genes had P-loop, Kinase-2, and GLPL motifs, the characteristic domains of the CNL genes. Phylogenetic relationships of the identified CNL genes along with their orthologs in *Arabidopsis* and in rice are shown in **Figure 4.1 and 4.2**, respectively. The CNL genes were nested in three clades (A, B and C). The clade D found in *Arabidopsis* and other dicot species was completely absent. The CNL-A clade was severely reduced to one member in the *A. tauschii* genome whereas *Arabidopsis* has six CNL-A members. While *A. tauschii* has a substantially larger genome than rice, the number of coding genes for *A. tauschii* and rice are quite similar, at 33,929 and 35,679 genes, respectively [15,21]. The CNL gene-content in the two genomes is not highly divergent, despite a huge difference in genome size between the two species. Table 1 shows that the number of CNL genes does not necessarily correlate with genome size (G-value paradox; Michelmore et al. 2013). With the larger genome size (2.7Gb), however, the *A. tauschii* genome contains a higher number of CNL genes. The rice genome (420 Mb) contains approximately 150 CNL genes [15]. All CNL clade information for the 402 identified genes is summarized in **Table 4.2**.

This study confirmed through MEME analysis (**Figure 4.3**) the presence of characteristic motifs (P-loop, Kinase-2, and GLPL) in all 402 CNL genes in *Aegilops*. The motif compositions presented here are similar to that in *Arabidopsis* [28], corresponded to the phylogenetic clustering represented in the phylogenetic tree (**Figure 4.1, 4.3**). For instance, Motif 8 (CPxxL) was common in the CNL-C4 clade but only in a few genes in the rest of CNL-C (Clades CNL-C1, CNL-C2, and CNL-C3). Since only the most prevalent motifs were labeled, and few CNL-A and CNL-B genes were present, it is likely that motifs were present but not described by the MEME analysis.

Since *A. tauschii* genes have not been mapped onto their chromosomes, gene clustering analysis was not performed in the present study. It is highly likely that the genes exist in many clusters throughout the genome (Meyers et al. 2003), particularly in the extrapericentromeric regions of the chromosomes as documented in soybean [16,19]. Further analyses of NBS-LRR disease resistance gene clustering will need to be done once this information becomes available. Also not available yet are the alternate transcripts for each of the genes. This is evident because the number of protein sequences available is equal to the number of coding genes within the genome. In other genomes, such as the barley genome, many more protein sequences exist that give information on alternative splicing amongst the resistance genes. Alternative splicing would increase the possible resistance gene proteins, which would be highly useful while facing a quickly evolving pathogen. While information on alternate splicing is not available for the *Aegilops* CNL genes, exon/intron information is available (**Figure 4.4**). The average exon content of 4.45 exons per gene is higher than previously found in *Arabidopsis* and CNL-C genes in soybean [16,19]. The number of exons varied from 1 (F775_00002) to 28

(F775_52438). Thirty five CNL genes had one exon, 77 had two, 83 had three, 58 had four, 44 had five, 28 had six, 23 had seven, 9 had eight, 14 had nine, 10 had ten, six had 11, seven had 12, three had 13, two had 14, one had 22 and one gene had 28 exons (**Figure 4.5**). With the multitude of genes with many exons, it can be hypothesized that alternate splicing has a large impact on the protein structure of the resistance genes, since multiple exons allow for a higher number of combinations during splicing [29]. Alternative splicing has been shown to play an important role in resistance gene expression in *Arabidopsis* [29,30].

Phylogenetic analysis of *A. tauschii* CNL genes shows an expansion of the CNL-C group and a slight reduction of CNL-B members relative to *Arabidopsis* (Figure 1). There is a severe reduction of the CNL-A clade to a single member. These results in the *Aegilops* genome are consistent with the CNL genes in rice, another monocot species [16,19]. There was low interspecific nesting indicating the lower prevalence of segmental duplications. Since chromosome location and gene clustering information were not available, instances of tandem versus segmental duplications could not be determined with a high degree of certainty. Genes that are nested together within a clade and occurring within the same gene clusters are likely to have originated through tandem duplications. The current study presented several instances of tandem duplication: for example F775_14065 and F775_14066 which are sister members (Figure 1), and subsequently accessioned, it is highly likely that they originated by tandem duplication. Other examples of tandem duplications include F775_11136 and F775_11137, F775_02795 and F775_02796, F775_10498 and F775_10499, F775_10336 and F775_10337, and three genes F775_17386, F775_17388 and F775_17389.

Orthologs of some *A. tauschii* CNL genes have been previously characterized. For example, RPM1 of *Arabidopsis thaliana* is involved in the resistance response to *Pseudomonas syringae* [31]. As shown in Figure 1, the *Arabidopsis* RPM1 ortholog in *Aegilops* has three paralogs (F775_10347, F775_14260, and F775_13161) indicating an expansion of this particular gene. It could be hypothesized that *A. tauschii* evolved the three genes in response to diversifying *P. syringae* strains or similar pathogens since the split of common ancestors of *Arabidopsis* and *A. tauschii*. The diversification of RPM1 orthologs in *Aegilops* might have been because of the selection pressure imposed by different pathogens in *A. taushii*'s life history. Figure 2 shows expansions of several *Aegilops* CNL genes: for example, eleven *A. tauschii* paralogs (F775_10913, 12507, 12011, 05946, 06830, 13024, 33089, 06253, 11684, 09360, and 21401) are related to rice gene LOC_Os08g10260. This shows that *A. tauschii* might have evolved as many as 11 genes in response to the same pathogen as in rice, perhaps diversifying in the *Aegilops* niche.

Due to the growing problem of Ug99 stem rust in wheat production of East Africa and the Middle-East, the CNL resistance gene SR33 has been identified as a possible solution [32]. The result determines that accession F775_10122 represents the SR33 gene in *Aegilops*, which could be the gene of interest for developing a durable resistance in wheat. Other genes (F775_13548, F775_16813, and F775_18040) closely related to SR33 might contain valuable traits as well. Further investigation of these genes, along with the splice variants of F775_10122 is warranted if SR33 proves to be useful in agricultural production. *In silico* analyses of R-genes such as presented here are integral stepping-stones toward the use of these identified genes as weapons against evolving

pathogens. While further investigation of gene expression data and genomic composition is important for understanding functional characterization, the present study provides information on the diversity and evolutionary history of the CNL genes in *A. tauschii* genome, and has a potential implication in wheat crop improvement with durable resistant genes in the future.

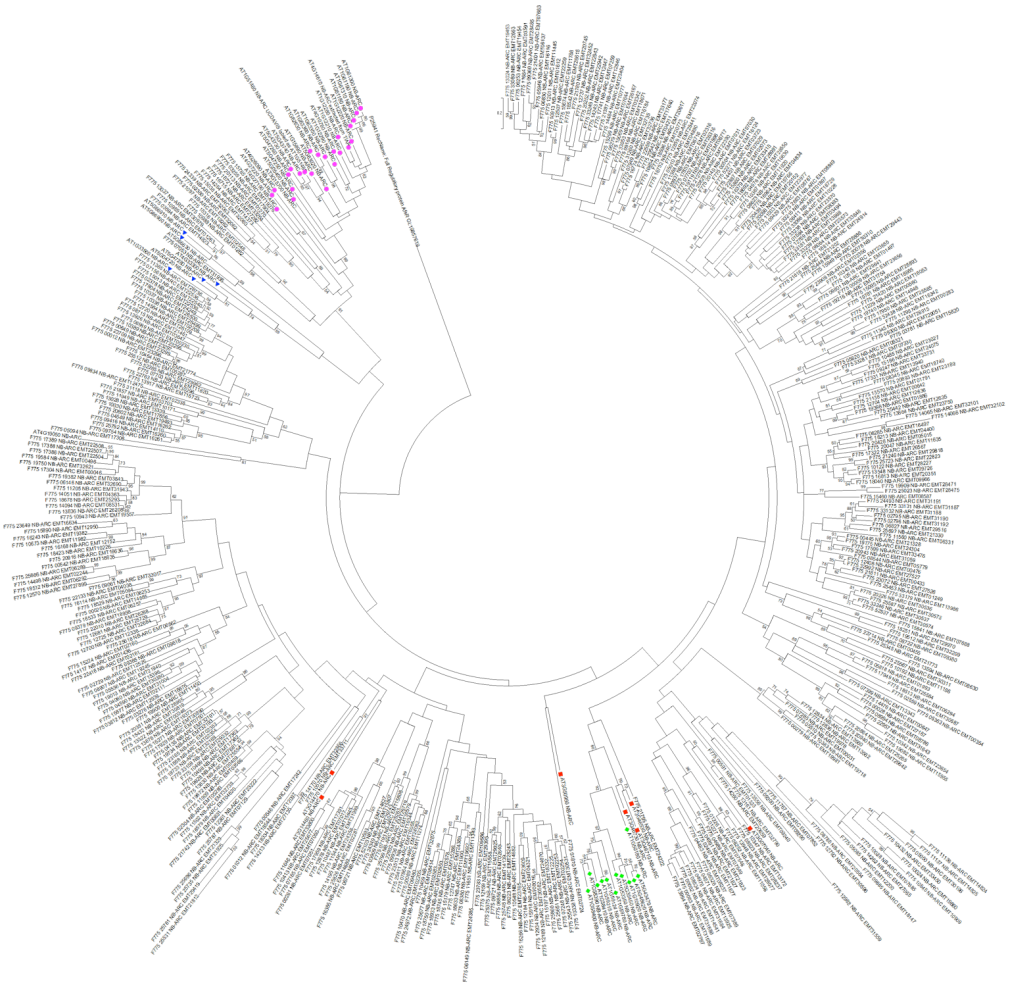


Figure 4.1 Phylogenetic analysis of the CNL genes of *A. tauschii* and their orthologs in *A. thaliana*. The tree was constructed using the JTT+G model with 100 bootstrap replicates. CNL clades A, B, C, and D are shown with blue, pink, red, and green markers, respectively.

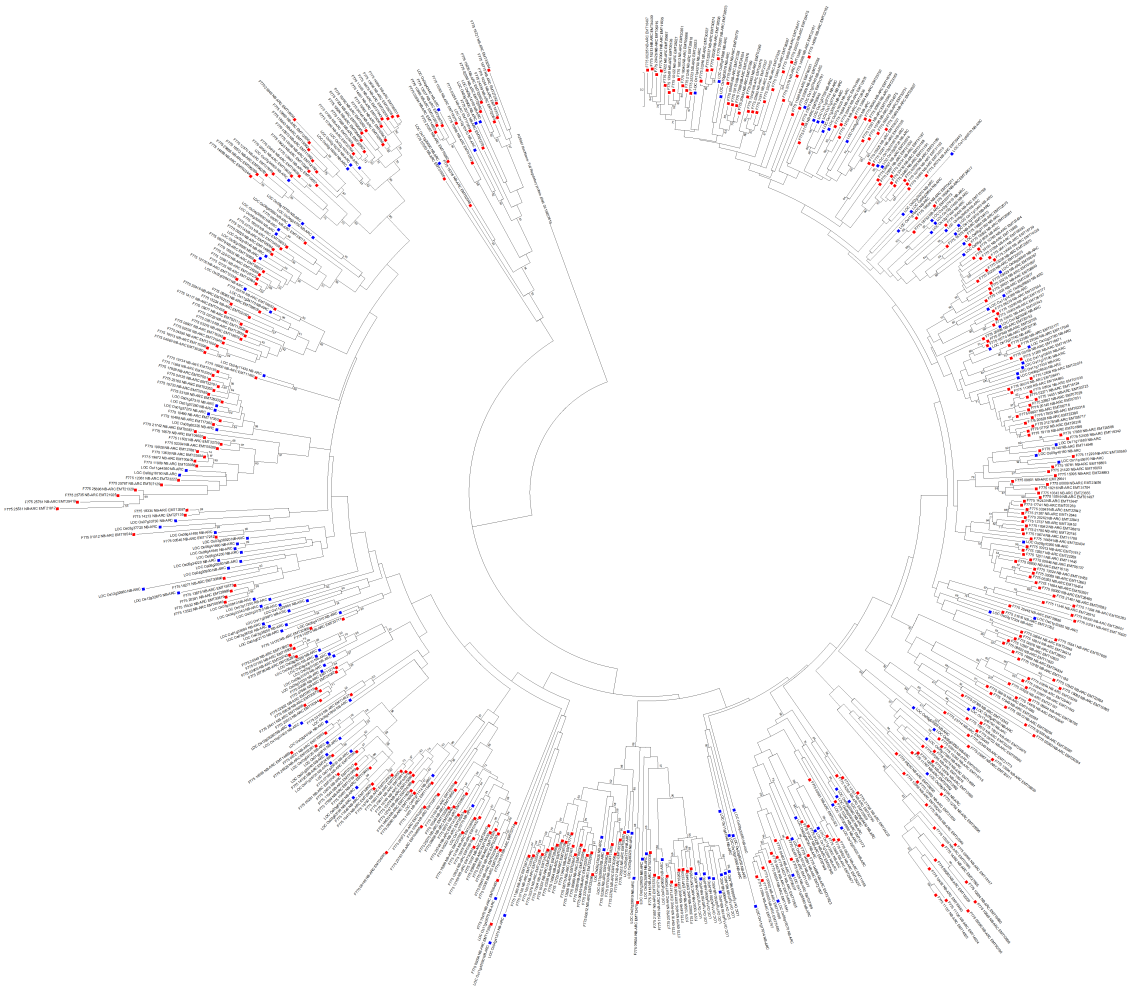


Figure 4.2 Phylogenetic analysis of the CNL genes of *A. tauschii* and their orthologs in rice. The tree was constructed using the JTT+G model with 100 bootstrap replicates. *A. tauschii* and rice genes are shown with red and blue markers, respectively.

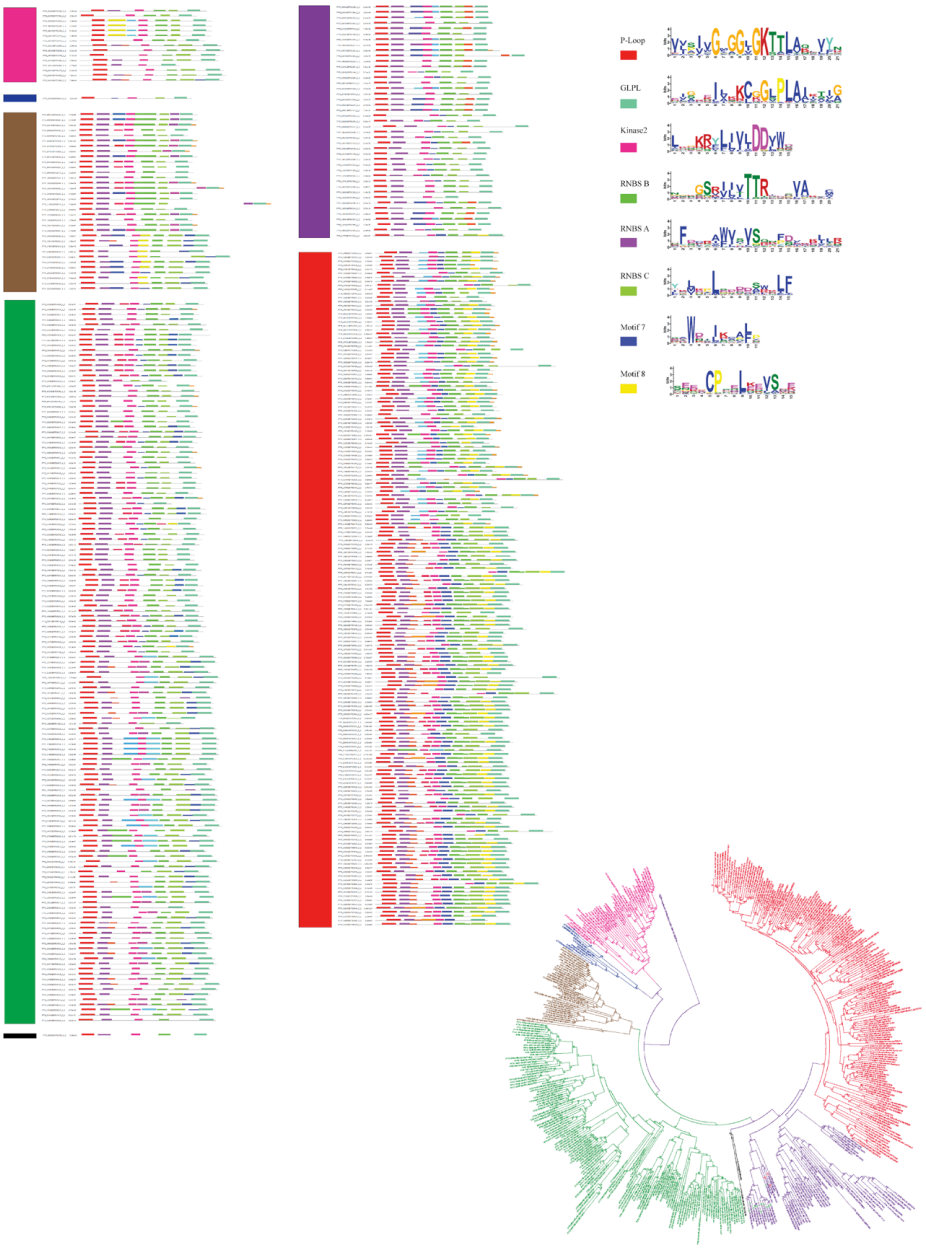


Figure 4.3 MEME analysis of the 402 *A. tauschii* genes. The block diagrams show the characteristics three motifs used to identify CNL genes (P-Loop, Kinase-2, and GLPL) along with other highly prevalent motifs, split according to clade as shown by the tree (lower right) color-coded to represent the domain compositions in Figure 1. CNL-B, A, C1, C2, C3, and C4 are colored pink, blue, brown, green purple, and red, respectively.

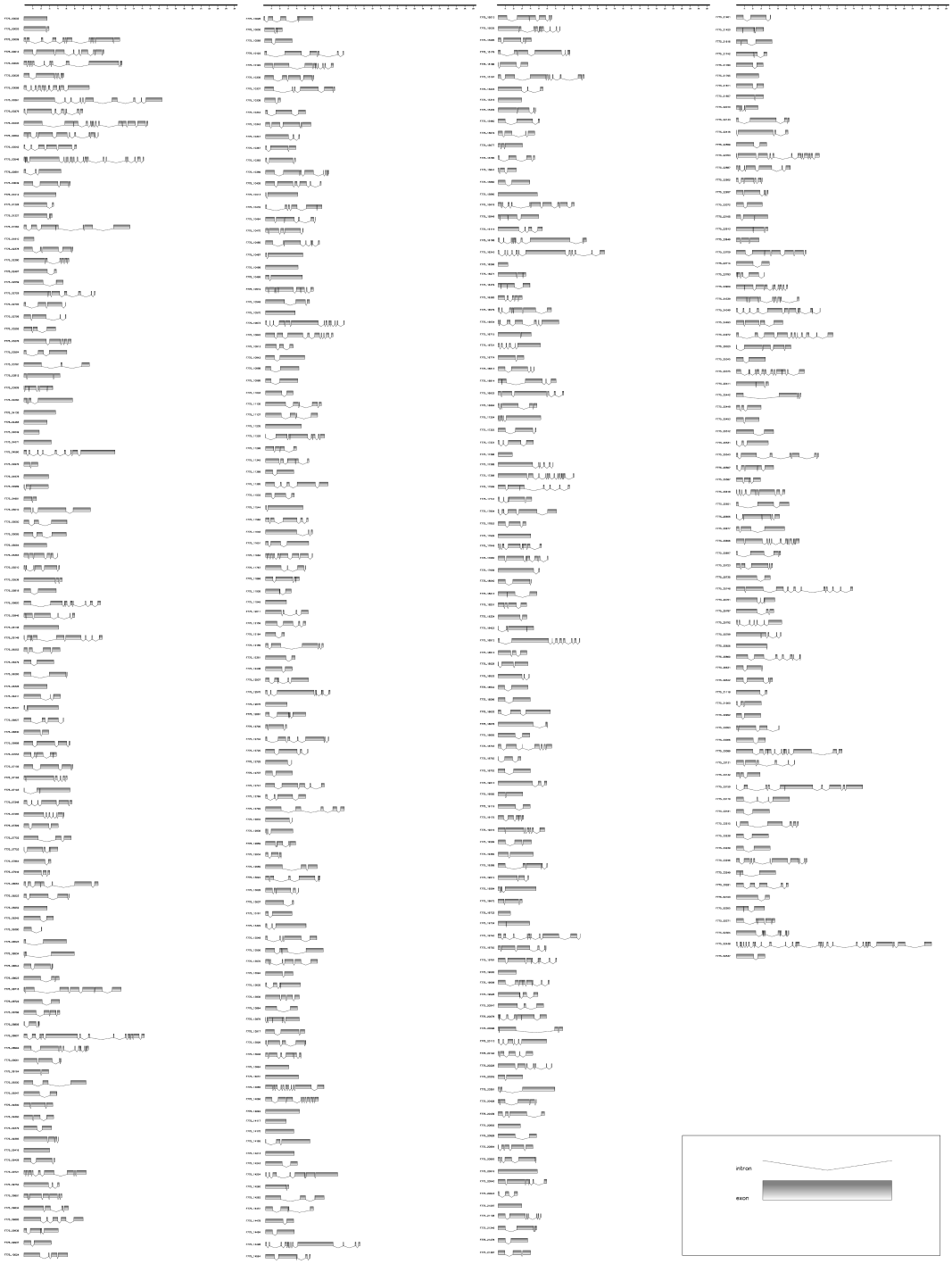


Figure 4.4 Exon content of the 402 *A. tauschii* genes showing splice locations between exons (gray bars) and introns (dashed lines). Genes are listed by accession.

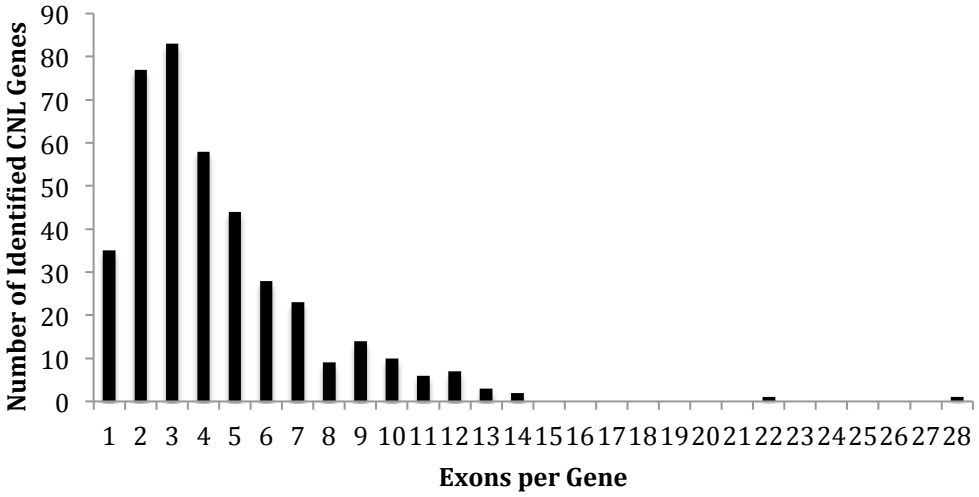


Figure 4.5 Number of CNL genes with specific number of exons in *A. tauschii*

Table 4.1 Genome size and CNL gene content of selected plant species. This table was modified from Marone et al [33]. Genome size and CNL gene references are both listed in the references column.

Species	Genome Size	Number of CNL genes	Reference
Aegilops tauschii	4.4 Gb	402	[21]
Glycine max	1.115 Gb	188	[16,19,34]
Solanum tuberosum	844 Mb	370	[20,35]
Phaseolus vulgaris	587 Mb	94	[16,36]
Vitis vinifera	487 Mb	203	[37,38]
Populus trichocarpa	423 Mb	119	[17,39]
Oryza sativa	420 Mb	159, 149	[15, 16,40]
Medicago truncatula	375 Mb	177	[18,41]
Carica papaya	372 Mb	6	[13,42]
Brassica rapa	284 Mb	30	[43,44]
Brachypodium distachyon	272 Mb	102	[45,46]
Cucumis sativus	244 Mb	18	[14,47]
Arabidopsis lyrata	207 Mb	21	[48,49]
Arabidopsis thaliana	125 Mb	55	[8,50]

Table 4.2 List of all *Aegilops tauschii* CNL genes according to clade.

Aegilops gene	Clade	*Aegilops* gene	Clade	*Aegilops* gene	Clade	*Aegilops* gene	Clade	*Aegilops* gene	Clade
F775_00002	C3	F775_09061	C2	F775_12934	C2	F775_18513	C4	F775_25567	C4
F775_00003	C3	F775_09164	C4	F775_12982	C3	F775_18529	C2	F775_25587	C4
F775_00009	C4	F775_09200	C3	F775_13024	C4	F775_18533	C2	F775_25618	C2
F775_00012	C1	F775_09247	C4	F775_13028	C1	F775_18542	C4	F775_25651	C2
F775_00020	C2	F775_09300	C4	F775_13037	B	F775_18596	C4	F775_25666	C2
F775_00028	C4	F775_09360	C4	F775_13161	C3	F775_18633	C2	F775_25677	C2
F775_00089	C2	F775_09379	C2	F775_13322	C2	F775_18678	C2	F775_25696	C2
F775_00261	C2	F775_09385	C2	F775_13548	C4	F775_18692	C4	F775_25697	C4
F775_00279	C3	F775_09416	C1	F775_13556	C3	F775_18745	C2	F775_25723	C4
F775_00445	C4	F775_09429	C4	F775_13570	C4	F775_18750	C2	F775_25735	C2
F775_00504	C4	F775_09721	C2	F775_13594	C4	F775_18752	C4	F775_25748	C2
F775_00542	C2	F775_09754	C1	F775_13630	C2	F775_19013	C2	F775_25761	C2
F775_00546	C2	F775_09801	C4	F775_13836	C2	F775_19082	C4	F775_25787	C2
F775_00591	C3	F775_09834	C1	F775_13864	C4	F775_19119	C4	F775_25792	C1
F775_00649	C1	F775_09885	C3	F775_13876	C2	F775_19175	C4	F775_25799	C2
F775_01012	C2	F775_09936	C2	F775_13917	C1	F775_19216	C4	F775_25826	C2
F775_01226	C3	F775_09937	C4	F775_13926	C1	F775_19299	C4	F775_25860	C2
F775_01227	C3	F775_10024	C3	F775_13948	C3	F775_19382	C2	F775_26631	C4
F775_01584	C4	F775_10028	C3	F775_13994	C3	F775_19398	C4	F775_29542	C4
F775_01810	C	F775_10030	C1	F775_14051	C2	F775_19512	C4	F775_31118	C1
F775_02378	C1	F775_10069	C4	F775_14065	C4	F775_19584	C2	F775_31260	C4
F775_02380	B	F775_10122	C4	F775_14066	C4	F775_19672	C2	F775_32992	C4
F775_02497	C3	F775_10192	C4	F775_14094	C2	F775_19733	C2	F775_33053	C4
F775_02559	C4	F775_10336	C1	F775_14117	C2	F775_19734	C2	F775_33066	C4
F775_02729	C2	F775_10337	C1	F775_14170	C2	F775_19740	C4	F775_33089	C4
F775_02795	C4	F775_10338	C2	F775_14195	C2	F775_19750	C2	F775_33131	C4
F775_02796	C4	F775_10342	C4	F775_14213	C2	F775_19781	C4	F775_33132	C4
F775_03255	B	F775_10343	C4	F775_14243	C4	F775_19900	C2	F775_33159	C2
F775_03276	C2	F775_10347	C3	F775_14254	B	F775_19909	C4	F775_33179	C4
F775_03594	B	F775_10367	C4	F775_14260	C3	F775_19928	C2	F775_33181	C4
F775_03781	C4	F775_10383	C3	F775_14262	C3	F775_20047	C4	F775_33215	C4
F775_03812	C2	F775_10389	C1	F775_14451	C4	F775_20078	C4	F775_33238	C4
F775_03909	C2	F775_10409	C2	F775_14478	C4	F775_20098	C4	F775_33239	C4
F775_04060	C2	F775_10413	C2	F775_14484	C4	F775_20113	B	F775_33246	C4
F775_04135	C2	F775_10432	C3	F775_14498	C2	F775_20140	C4	F775_33249	C4
F775_04483	C3	F775_10464	C1	F775_14564	C1	F775_20226	C4	F775_33281	C4
F775_04549	C1	F775_10470	C2	F775_15013	C1	F775_20252	C4	F775_52103	C2
F775_04571	C3	F775_10485	C4	F775_15035	B	F775_20381	C2	F775_52265	C1
F775_04590	C2	F775_10487	C2	F775_15095	C4	F775_20428	C4	F775_52271	C4
F775_04976	C4	F775_10498	C2	F775_15179	C2	F775_20439	C4	F775_52304	C2
F775_04978	C3	F775_10499	C2	F775_15186	C4	F775_20802	C1	F775_52483	C4
F775_04989	C3	F775_10519	C4	F775_15197	C2	F775_20828	C4	F775_52537	C4
F775_04991	C3	F775_10548	C2	F775_15224	C2	F775_20864	C4		
F775_05010	C4	F775_10570	C2	F775_15316	B	F775_20893	C4		
F775_05050	C3	F775_10673	C2	F775_15432	C2	F775_20916	C2		
F775_05085	C3	F775_10845	C3	F775_15460	C4	F775_20940	C4		
F775_05094	C1	F775_10913	C4	F775_15674	C4	F775_20943	C4		
F775_05363	C4	F775_10943	C2	F775_15677	C2	F775_21097	B		
F775_05510	C3	F775_10988	B	F775_15785	B	F775_21138	C4		
F775_05536	C2	F775_10989	B	F775_15841	C4	F775_21246	C4		
F775_05818	C4	F775_11003	C4	F775_15860	C4	F775_21278	C4		
F775_05820	C4	F775_11136	C3	F775_15890	C2	F775_21387	C4		
F775_05946	C4	F775_11137	C3	F775_15918	C2	F775_21401	C4		
F775_06146	C2	F775_11205	C2	F775_15949	C4	F775_21420	C4		
F775_06149	C2	F775_11229	C4	F775_16114	C2	F775_21616	C4		
F775_06253	C4	F775_11298	C4	F775_16168	C2	F775_21742	C2		
F775_06279	C4	F775_11345	C4	F775_16243	C2	F775_21780	C4		
F775_06285	C4	F775_11368	C4	F775_16266	C2	F775_21795	C		
F775_06326	C3	F775_11385	C1	F775_16271	C2	F775_21811	C4		
F775_06411	C4	F775_11502	C2	F775_16379	C4	F775_21857	C1		
F775_06721	C2	F775_11544	C2	F775_16385	C2	F775_22010	C2		
F775_06827	C4	F775_11560	C4	F775_16579	C2	F775_22133	C2		
F775_06830	C4	F775_11646	C2	F775_16654	C4	F775_22416	C2		
F775_06989	C1	F775_11651	C2	F775_16715	B	F775_22559	C4		
F775_07053	A	F775_11684	C4	F775_16721	B	F775_22763	C1		
F775_07156	C1	F775_11767	C3	F775_16774	C4	F775_22887	C4		
F775_07165	C2	F775_11868	C2	F775_16813	C4	F775_22902	C2		
F775_07193	C2	F775_11909	C2	F775_16814	C4	F775_22957	C4		
F775_07248	C2	F775_11949	C1	F775_16933	C2	F775_23072	C4		
F775_07285	C3	F775_12011	C4	F775_16964	C4	F775_23165	C2		
F775_07399	C4	F775_12159	C2	F775_17304	C2	F775_23513	C2		
F775_07702	C4	F775_12184	C2	F775_17322	C4	F775_23649	C2		
F775_07703	C1	F775_12189	C2	F775_17331	C4	F775_23709	C1		
F775_07864	C2	F775_12361	C2	F775_17386	C2	F775_23714	C4		
F775_07949	C4	F775_12408	C4	F775_17388	C2	F775_23783	C2		
F775_08064	C4	F775_12507	C4	F775_17389	C2	F775_23909	C4		
F775_08223	C2	F775_12570	C2	F775_17599	C4	F775_24339	B		
F775_08252	C3	F775_12676	C2	F775_17741	C4	F775_24349	C2		
F775_08345	C4	F775_12681	C2	F775_17804	C1	F775_24493	C4		
F775_08380	C2	F775_12700	C2	F775_17853	C4	F775_24972	C2		
F775_08523	C3	F775_12704	C4	F775_17929	C2	F775_25023	C4		
F775_08534	C3	F775_12720	C1	F775_17949	C4	F775_25345	C4		
F775_08544	C4	F775_12725	C2	F775_17950	C4	F775_25375	C2		
F775_08623	C4	F775_12737	C4	F775_17959	C2	F775_25411	C2		
F775_08715	C1	F775_12747	C2	F775_18040	C4	F775_25442	C4		
F775_08722	C4	F775_12769	C4	F775_18213	C4	F775_25448	C4		
F775_08786	C3	F775_12792	C3	F775_18251	C4	F775_25453	C4		
F775_08856	C2	F775_12834	C3	F775_18334	C2	F775_25512	C1		
F775_08907	C2	F775_12836	C3	F775_18423	C2	F775_25531	C2		
F775_08994	C4	F775_12859	C4	F775_18512	C2	F775_25543	C2		

ACKNOWLEDGEMENTS

Support for this dissertation work came from the USDA-NIFA hatch Projects to M. Nepal (SD00H469-13 and SD00H659-18), South Dakota Agriculture Experiment Station, and Department of Biology & Microbiology at South Dakota State University.

REFERENCES

1. Jones JD, Dangl JL. The plant immune system. *Nature.* 2006;444(7117):323-329.
2. Hammond-Kosack KE, Jones J. Resistance gene-dependent plant defense responses. *The Plant Cell.* 1996;8(10):1773.
3. Flor HH. Current status of the gene-for-gene concept. *Annual review of phytopathology.* 1971;9(1):275-296.
4. Van Der Biezen EA, Jones JD. Plant disease-resistance proteins and the gene-for-gene concept. *Trends in biochemical sciences.* 1998;23(12):454-456.
5. Shao F, Golstein C, Ade J, Stoutemyer M, Dixon JE, Innes RW. Cleavage of Arabidopsis PBS1 by a bacterial type III effector. *Science.* 2003;301(5637):1230-1233.
6. Gururani MA, Venkatesh J, Upadhyaya CP, Nookaraju A, Pandey SK, Park SW. Plant disease resistance genes: current status and future directions. *Physiological and Molecular Plant Pathology.* 2012;78:51-65.
7. Meyers BC, Kaushik S, Nandety RS. Evolving disease resistance genes. *Current opinion in plant biology.* 2005;8(2):129-134.
8. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of NBS-LRR–encoding genes in Arabidopsis. *The Plant Cell Online.* 2003;15(4):809-834.

9. Bergelson J, Kreitman M, Stahl EA, Tian D. Evolutionary dynamics of plant R-genes. *Science.* 2001;292(5525):2281-2285.

10. Michelmore RW, Meyers BC. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Research.* 1998;8(11):1113-1130.

11. McGrann GR, Stavrinides A, Russell J, et al. A trade off between mlo resistance to powdery mildew and increased susceptibility of barley to a newly important disease, Ramularia leaf spot. *Journal of experimental botany.* 2014:ert452.

12. Lin X, Zhang Y, Kuang H, Chen J. Frequent loss of lineages and deficient duplications accounted for low copy number of disease resistance genes in Cucurbitaceae. *BMC genomics.* 2013;14(1):335.

13. Porter BW, Paidi M, Ming R, Alam M, Nishijima WT, Zhu YJ. Genome-wide analysis of Carica papaya reveals a small NBS resistance gene family. *Molecular Genetics and Genomics.* 2009;281(6):609-626.

14. Wan H, Yuan W, Bo K, Shen J, Pang X, Chen J. Genome-wide analysis of NBS-encoding disease resistance genes in Cucumis sativus and phylogenetic study of NBS-encoding genes in Cucurbitaceae crops. *BMC genomics.* 2013;14(1):109.

15. Zhou T, Wang Y, Chen J-Q, et al. Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Molecular Genetics and Genomics.* 2004;271(4):402-415.

16. Benson BV. *Disease Resistance Genes and their Evolutionary History in Six Plant Species*: Biology and Microbiology, South Dakota State University; 2014.

17. Kohler A, Rinaldi C, Duplessis S, et al. Genome-wide identification of NBS resistance genes in Populus trichocarpa. *Plant molecular biology.* 2008;66(6):619-636.

18. Ameline-Torregrosa C, Wang B-B, O'Bleness MS, et al. Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant Medicago truncatula. *Plant physiology.* 2008;146(1):5-21.

19. Nepal MP, Benson BV. CNL Disease Resistance Genes in Soybean and Their Evolutionary Divergence. *Evolutionary Bioinformatics Online.* 2015;11:49-63.

20. Lozano R, Ponce O, Ramirez M, Mostajo N, Orjeda G. Genome-wide identification and mapping of NBS-encoding resistance genes in Solanum tuberosum group phureja. *PLoS One.* 2012;7(4):e34775.

21. Jia J, Zhao S, Kong X, et al. Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature.* 2013;496(7443):91-95.

22. Kersey PJ, Allen JE, Christensen M, et al. Ensembl Genomes 2013: scaling up access to genome-wide data. *Nucleic acids research.* 2014;42(D1):D546-D552.

23. Goodstein DM, Shu S, Howson R, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic acids research.* 2012;40(D1):D1178-D1186.

24. Kearse M, Moir R, Wilson A, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012;28(12):1647-1649.

25. Jones P, Binns D, Chang H-Y, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014:btu031.

26. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in bipolymers. Department of Computer Science and Engineering, University of California, San Diego; 1994.

27. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and

maximum parsimony methods. *Molecular Biology and Evolution.* 2011;28(10):2731-2739.

28. Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide‐binding superfamily. *The Plant Journal.* 1999;20(3):317-332.

29. Tan X, Meyers BC, Kozik A, et al. Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in Arabidopsis. *BMC plant biology.* 2007;7(1):56.

30. Dinesh-Kumar S, Baker BJ. Alternatively spliced N resistance gene transcripts: their possible role in tobacco mosaic virus resistance. *Proceedings of the National Academy of Sciences.* 2000;97(4):1908-1913.

31. Mackey D, Holt BF, Wiig A, Dangl JL. RIN4 interacts with Pseudomonas syringae type III effector molecules and is required for RPM1-mediated resistance in Arabidopsis. *Cell.* 2002;108(6):743-754.

32. Periyannan S, Moore J, Ayliffe M, et al. The gene Sr33, an ortholog of barley Mla genes, encodes resistance to wheat stem rust race Ug99. *Science.* 2013;341(6147):786-788.

33. Marone D, Russo MA, Laidò G, De Leonardis AM, Mastrangelo AM. Plant nucleotide binding site–leucine-rich repeat (NBS-LRR) genes: active guardians in host defense responses. *International journal of molecular sciences.* 2013;14(4):7302-7326.

34. Schmutz J, Cannon SB, Schlueter J, et al. Genome sequence of the palaeopolyploid soybean. *nature.* 2010;463(7278):178-183.

35. Consortium PGS. Genome sequence and analysis of the tuber crop potato. *Nature.* 2011;475(7355):189-195.

36. Schmutz J, McClean PE, Mamidi S, et al. A reference genome for common bean and genome-wide analysis of dual domestications. *Nature genetics.* 2014;46(7):707-713.

37. Yang S, Zhang X, Yue J-X, Tian D, Chen J-Q. Recent duplications dominate NBS-encoding gene expansion in two woody species. *Molecular Genetics and Genomics.* 2008;280(3):187-198.

38. Jaillon O, Aury J-M, Noel B, et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *nature.* 2007;449(7161):463-467.

39. Tuskan GA, Difazio S, Jansson S, et al. The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). *science.* 2006;313(5793):1596-1604.

40. Goff SA, Ricke D, Lan T-H, et al. A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). *Science.* 2002;296(5565):92-100.

41. Young ND, Debellé F, Oldroyd GE, et al. The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature.* 2011;480(7378):520-524.

42. Ming R, Hou S, Feng Y, et al. The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). *Nature.* 2008;452(7190):991-996.

43. Mun J-H, Yu H-J, Park S, Park B-S. Genome-wide identification of NBS-encoding resistance genes in Brassica rapa. *Molecular Genetics and Genomics.* 2009;282(6):617-631.

44. Wang X, Wang H, Wang J, et al. The genome of the mesopolyploid crop species Brassica rapa. *Nature genetics.* 2011;43(10):1035-1039.

45. Tan S, Wu S. Genome wide analysis of nucleotide-binding site disease resistance genes in Brachypodium distachyon. *Comparative and Functional Genomics.* 2012;2012.

46. Vogel JP, Garvin DF, Mockler TC, et al. Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature.* 2010;463(7282):763-768.

47. Huang S, Li R, Zhang Z, et al. The genome of the cucumber, Cucumis sativus L. *Nature genetics.* 2009;41(12):1275-1281.

48. Guo Y-L, Fitz J, Schneeberger K, Ossowski S, Cao J, Weigel D. Genome-wide comparison of nucleotide-binding site-leucine-rich repeat-encoding genes in Arabidopsis. *Plant physiology.* 2011;157(2):757-769.

49. Hu TT, Pattyn P, Bakker EG, et al. The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. *Nature genetics.* 2011;43(5):476-481.

50. Initiative AG. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *nature.* 2000;408(6814):796.

CHAPTER 5: RESISTANCE GENES IN WHEAT SHOW EVIDENCE OF TANDEM DUPLICATIONS AS A MECHANISM FOR FUNCTIONAL DIVERSIFICATION

ABSTRACT

Proteins encoded by plant resistance genes (R genes) detect pathogenic effectors and initiate immune responses. These proteins generally contain three domains essential to their function: Leucine-Rich Repeats, NB-ARCs, and Coiled-Coils. The major objectives of this project were to identify NB-ARC-encoding R genes in wheat (*Triticum aestivum* L.) and assess their evolutionary history. Since previous studies have shown that many R genes lack some domains associated with receptor function, analysis of all NB-ARC-encoding genes provides a more comprehensive view of how these genes evolve. Similarities between members of gene clusters were identified using the program Geneious, gene chromosomal locations were mapped using the program Circa, clustered genes were compared through the construction of distance trees using the program MEGA, and chromosomes were aligned to assess synteny using the program SyMAP. The results showed that the wheat genome contains a total of 2151 NB-ARC-encoding genes, of 1298 formed 547 gene clusters. Many of these clusters included highly similar genes likely formed through tandem duplications. Of the 2151 NB-ARC-encoding genes, 1552 encode Leucine-Rich Repeats (LRRs; approximately 72%), 802 of which are Coiled-Coil (CC) domain-encoding CC-NBS-LRR (CNL) genes and three are Resistance to Powdery mildew 8 (RPW8) domain-encoding RPW8-NBS-LRR (RNL) genes. Surprisingly, five of the NB-ARC-encoding genes encoded a Toll/Interleukin-1 Receptor (TIR), with no LRR, known as TN genes. CNL clades formed similar phylogenetic nesting patterns as those found in previous analysis of close wheat relatives, showing that

grasses posses a large expansion of CNL-C genes. Comparisons of the wheat genome with those of its close relatives, barley (*Hordeum vulgare* L.) and Tausch's goatgrass (*Aegilops tauschii* Coss.), found similar locations for homologous NB-ARC-encoding genes. These results indicate that R genes in wheat have diversified through duplication to encode receptors that recognize additional pathogenic effectors, which future research should seek to characterize.

INTRODUCTION

Wheat (*Triticum aestivum* L.) provides approximately 20% of the human population's caloric intake [1] and is afflicted by over 100 different diseases caused by hundreds of pathogen and pest species [2]. Diseases that substantially reduce yield, such as biotrophic rusts and necrotrophic leaf spotting diseases, impact global markets and food supply. Historically devastating pathogens continue to produce new strains, which overcome past sources of resistance (i.e. Ug99) [3-6]. Leaf spotting diseases similarly cause a high percentage of yield loss under favorable conditions and appear as several forms: *Septoria tritici* blotch, spot blotch, *Stagonospora nodorum* blotch, and tan spot [7]. Additional diseases like Fusarium head blight, wheat streak mosaic virus, and powdery mildew also plague wheat fields with symptoms that are difficult and costly to manage. Disease resistant wheat cultivars, such as those bred by Edgar McFadden and Norman Borlaug [8,9], have been a significant contribution to agriculture and the understanding of phytopathology. With the help of advanced genetic and genomic technologies, recent efforts identified resistance genes (R genes) that confer resistance to strains of various

pathogens, such as the recently discovered Ug99 R-genes *Sr33* and *Sr35* [10,11]. The hexaploid bread wheat genome (AABBDD), a draft of which has recently become available [10,11], formed through the hybridization of three separate species: *Triticum urartu* (A), an unknown relative of *Aegilops speltoides* (B), and *Aegilops tauschii* (D)[12-14]. This unique polyploidy may have resulted in novel regulatory mechanisms that were necessary due to the presence of multiple progenitor resistance signaling pathways. The large and redundant nature of wheat's hexaploid genome makes it a good candidate for studying R gene evolution with respect to the recent polyploidization events. Detailed understanding of wheat immune system components provides a framework for further attempts at the development of durable resistance in cereals necessary for reduced yield loss from biotic stress. R genes are generally receptors encoding a Nucleotide-Binding Site and a Leucine-Rich Repeat, giving them the alternative names NBS-LRR or NLR. The LRR is thought to interact with pathogen effectors, allowing the protein to initiate signaling mechanisms (i.e. kinases and transcription factors) that initiate defense responses. The primary domain found in the NBS is the Nucleotide-Binding site found in Apoptotic protease activating factor 1, R genes, and *Caenorhabditis elegans* death-4 protein (NB-ARC). The NB-ARC is associated with ATP/ADP binding, since the molecule uses this upon activation.

The objectives of this research were to conduct genome-wide identification of wheat R genes encoding NB-ARC domains and assess their genomic architecture and potential functional divergence with respect to those in the genomes of wheat relatives. Results showed that tandem duplication can explain many of the events that led to the diversification of these genes. Genes encoding NB-ARC domains are likely involved in

resistance responses and are of use to breeders as sources of resistance. This research will contribute to development of resistance by associating R genes with their homologs found in surrounding chromosomal locations.

MATERIALS AND METHODS

Wheat chromosome, gene, and protein sequences were downloaded from the Ensembl Genomes database through the Biomart tool [15]. InterProScan [16] was used to identify wheat protein sequences with NB-ARC domains (PF00931). Locations for genes encoding NB-ARCs were analyzed for clusters as described in Jupe et al. (2012) using the criteria that genes be within 200,000 bases of each other with fewer than eight addition genes between them [17]. NB-ARC domain motifs were also assessed using MEME software[18], identifying those with P-loop, Kinase-2, and GLPL motifs. Clustered genes with the aforementioned motifs were aligned and manually curated using ClustalW2 integrated within the program Geneious [19,20]. The program MEGA 7 was used to construct a neighbor-joining tree with 100 bootstraps [21]. The online program iTOL [22] was used to color tree leaves, allowing easy visualization of which chromosomes each gene was found on. Wheat and *Aegilops tauschii* R gene locations from Ensembl Genomes were used to construct a genomic map using the program Circa (http://omgenomics.com/circa). Clustered wheat R genes were also compared using the program Circoletto [23], creating images that display high similarities between genes. Synteny between barley and wheat was determined by aligning chromosome sequences in the program SyMAP [24].

RESULTS

Approximately half of wheat's NB-ARC-containing (NBAC) proteins also contained a CC domain, and approximately 75% possessed LRRs (**Figure 5.1** and **Table 5.1**). Several NBAC proteins contained AP2/ERF, B3, WRKY, and Zinc Finger transcription factor domains. The NBAC proteins also contained kinase, lectin, thioredoxin, and GRAS domains. Of the 2151 NB-ARC-encoding genes, 1505 had NB-ARCs with P-loop, Kinase-2, and GLPL motifs. Among the 1552 proteins that had Leucine-Rich Repeats (LRRs; approximately 72%), 802 contained Coiled-Coil (CC) domains (CNL) and three had Resistance to Powdery mildew 8 (RPW8) domains (RNL). Five of the NB-ARC-encoding genes encoded a Toll/Interleukin-1 Receptor (TIR), with no LRR (TN proteins). NBAC genes formed 547 gene clusters. Clustered genes from each chromosome were compared, resulting in similarity diagrams as shown in **Figure 5.2**, illustrating high similarity among clustered genes. Many of the clustered genes possessed greater than 75% similarity within each cluster. While chromosomes from each of the subgenomes were very similar (e.g. chromosomes 1A, 1B, and 1D), differences between the number of clustered genes could have been easily spotted. This was especially noticeable for chromosomes 4A, 4B, and 4D, where 4A contained 57 clusters (involving 119 genes) and 4B and 4D contained four clusters (of eight genes) and three clusters (of seven genes), respectively. It is unknown to whether this diversification of NBAC genes in 4A happened prior to the first or second hybridization event in wheat. Since *Triticum urartu* gene locations were not available, a clear understanding of the difference between *T. urartu* chromosome 4 and wheat chromosome 4A could not be

reached. **Figure 5.3** shows a neighbor-joining tree composed of NBAC sequences, specifically those with P-loop, Kinase-2, and GLPL motifs. Nesting patterns in the neighbor-joining tree highlighted additional relationships between clustered genes, displaying which groups arose through tandem duplication, where clustered genes nested together, or segmental duplication, where clustered genes nested separately. Many pairs and triplets of genes from the same cluster nest together on the tree. Due to the close relationship between barley and wheat, synteny between wheat and barley chromosomes was also assessed. Chromosomes of wheat progenitors, such as *Aegilops tauschii*, align with wheat, 1D of wheat matching 1D of *A. tauschii*, and so on. As is visible from the syntenic map (**Figure 5.4**), the same phenomenon takes place between barley chromosomes (1H-7H) and the three subgenomes of wheat. An earlier analysis of wheat R genes shows similar syntenic patterns (see **Appendix III**).

Figure 5.1 NBAC gene locations on wheat chromosomes as generated using the program Circa. The outer track of lines indicates the locations of NBAC genes. Within those are the subset of NB-ARC-containing genes that contain CC and LRR domains, respectively. Within the wheat circle are the locations of NBAC genes in *Aegilops tauschii* (contributor of the wheat D subgenome) and *Hordeum vulgare* (close relative of wheat progenitors).

Table 5.1 NB-ARC-containing genes and domains found within them. Of the 2151 genes found in wheat, many contained Coiled-Coil (CC) domains and LRR domains.

R Protein Domains	Genes	Proposed Function
NB-ARC	2151	ATP-Binding
CC	1126	Signaling
LRR	1552	Ligand-Binding
TIR	5	Signaling
AP2/ERF	3	Transcription Factor
B3	3	Transcription Factor
WRKY	3	Transcription Factor
Zinc Finger	25	Transcription Factor
Protein Kinase	52	Kinase
S/T Protein Kinase	46	Kinase
Jacalin-Like Lectin	7	Signaling
Kelch repeat	7	Protein Interaction
Thioredoxin	10	Resistance Response
GRAS	4	Signaling

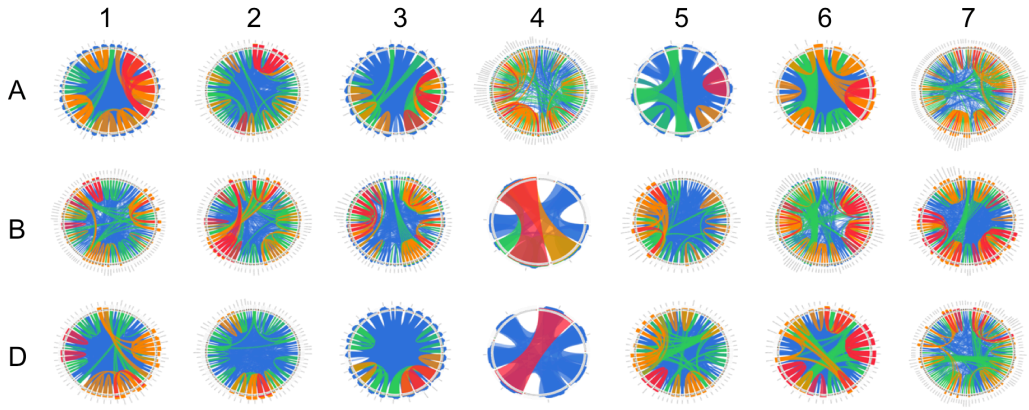


Figure 5.2 Similarity of clustered NBAC genes on each chromosome. Chromosomes are shown 1-7 with subgenomes A, B, and D from superior to inferior. Clustered genes are arranged in a circle, with lines describing <50% similarity, <75% similarity, <99.9999% similarity, and 100% similarity in blue, green, orange, and red, respectively.

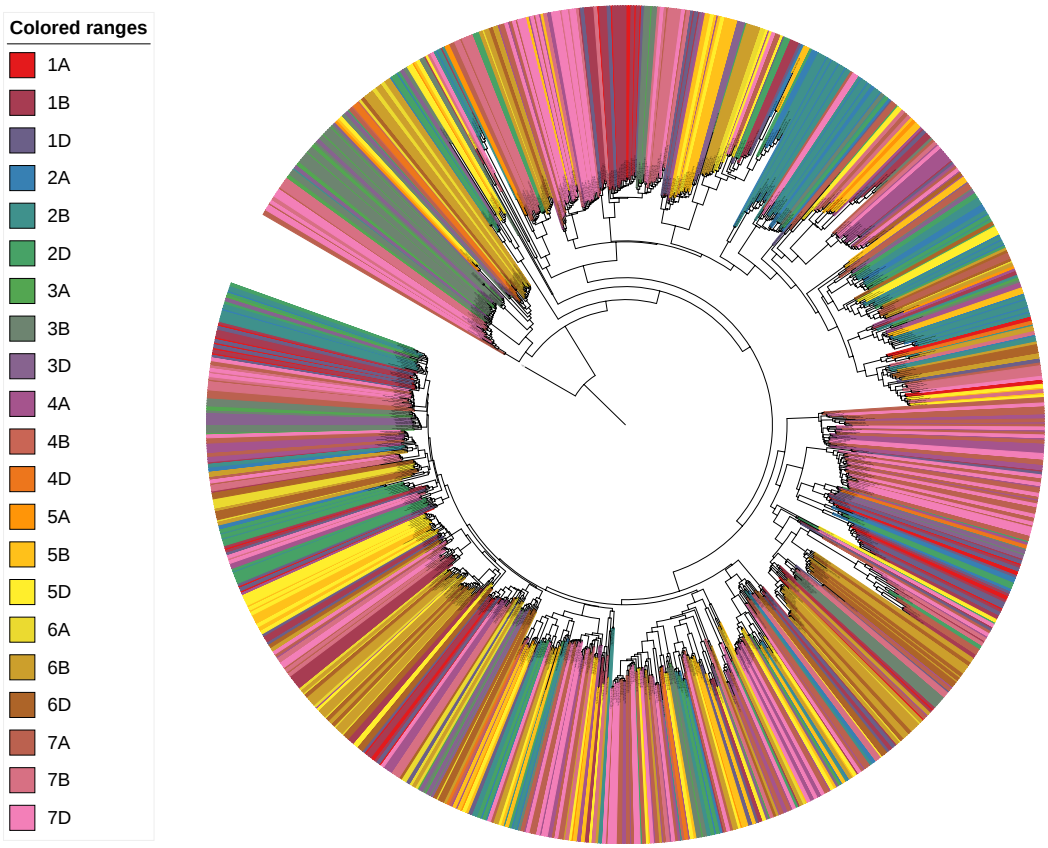


Figure 5.3 Distance relationships among clustered NBAC genes. Sequences were included if they contained P-loop, Kinase-2, and GLPL motifs as identified by MEME and are found in clusters. This neighbor-joining tree was generated using MEGA 7. Color profiles from the chromosomes in Figure 1 were used and can be seen the figure legend.

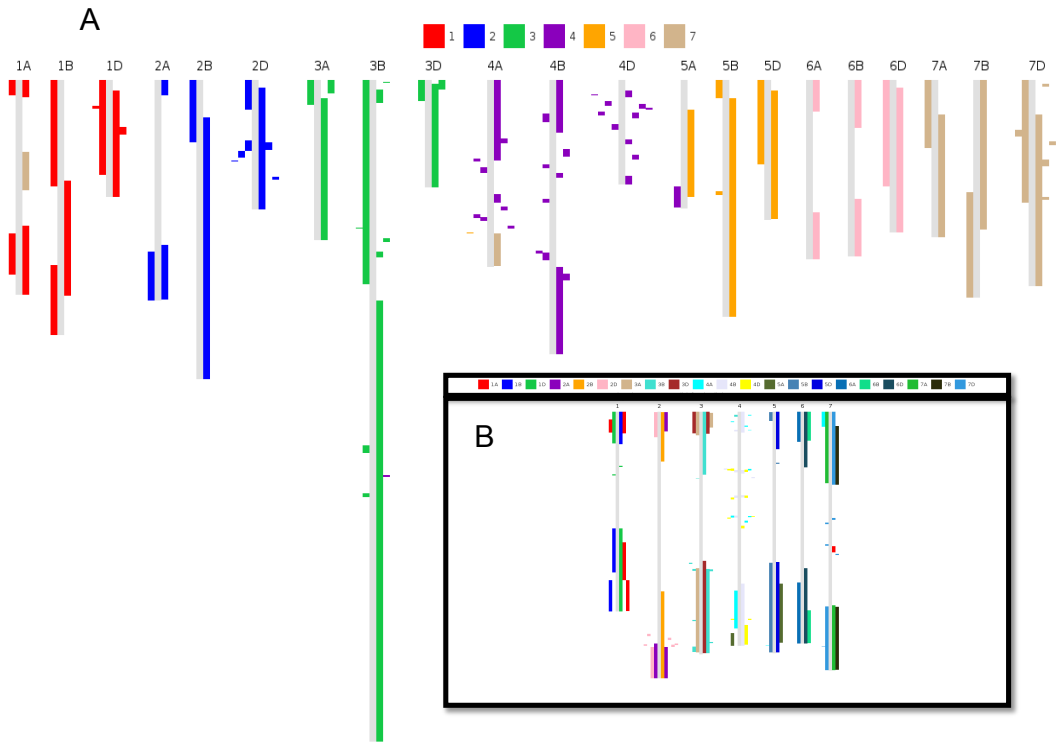


Figure 5.4 Synteny between chromosomes of wheat and barley. Barley chromosomes were aligned with wheat using the program SyMAP. Barley chromosomes are mapped onto the 21 chromosomes of wheat (A) and the inverse image of wheat chromosomes mapped on the seven chromosomes of barley (B).

DISCUSSION

This investigation focused on all genes with NB-ARCs because NLR systems have been found that use groups of sensor and helper NLRs to detect and initiate defense responses when pathogenic effectors are present [25]. Not all functional NLRs have all characteristic domains, such as Pb1 [26]. Excluding genes that do not contain particular domains or motifs may not include important genes that assist with resistance responses. Therefore, proteins lacking CC or LRR domains may also contribute to resistance responses, especially those with additional domains that are involved in signaling [27]. The

distribution of NBAC genes across wheat chromosomes concurs with previous studies in barley and foxtail millet, where R genes were also found in clusters in extra-pericentromeric regions of chromosomes [28,29]. Unequal crossing over between chromosomes as a mechanism for duplication likely explains the formation of these clusters, which then allows for their diversification. Previous studies have highlighted this explanation for the location of the quickly evolving genes [30]. *A. tauschii* and *H. vulgare* share a similar pattern as wheat (**Figure 5.1**), with a similar number of R genes located at the ends of chromosomes. Analyses of the clusters of NLR genes revealed many genes with high sequence similarity (**Figure 5.2**), indicating their origin by tandem duplications, especially those that are only a few hundred nucleotides away from each other and share >90% similarity. Through tandem duplication, wheat NLR genes may have diversified to respond to rapidly evolving and perhaps closely related pathogens. Many pathogens possess diverse races, such as the pathogen *Pyrenophora tritici-repentis* with eight races [31], and race-specific (vertical) resistance has been identified to wheat pathogens, such as powdery mildew [32]. This type of resistance involves one or two genes, as opposed to horizontal resistance, which involves multiple genes (quantitative) and provides resistance to many pathogens. Horizontal resistance may include other signaling factors and types of receptors, relying only partially on NLRs [33].

While it can be inferred that similar genes that are close to each other on a chromosome are likely due to tandem duplications, several genes were dissimilar and close together. This can be seen in **Figure 5.3**, where, while many clustered genes nest together, providing visual evidence of tandem duplications with closely clustered genes nesting together, many do not nest together. This phenomenon has two major

explanations: 1) the tandem duplication took place long ago in evolutionary history and has had time to diversify greatly, or 2) a segmental duplication took place, causing the gene to become located next to another R gene or R gene cluster. R genes are highly diversified in plants, many plants possessing hundreds of them, indicating that these R genes originate from very ancient precursors. Ancient tandem duplications would have much time to diversify, especially if particular selective pressures are put upon the ancestors of modern species. However, segmental duplications cannot be discounted due to the presence of genes in some clusters that are highly similar to genes in other clusters [34]. In these cases, transposable elements may play some role in movement of these genes around to other chromosomes or distant locations on the same chromosome [35].

Barley and the progenitors of wheat diverged only approximately 8-9 million years ago [36]. Therefore, barley provides an excellent partner for wheat synergistic comparison. Both barley and wheat experienced artificial selection since both have been grown for food production since the agricultural revolution approximately 10,000 years ago. Wheat differs from barley in that it is an allohexaploid resulting from hybridization of three species, each containing seven pairs of chromosomes to total 21 pairs, while barley remained diploid with only seven pairs of chromosomes. The wheat genome, consisting of A, B, and D subgenomes, maps to the barley genome (H), with wheat chromosomes 1A, 1B, and 1D containing much synteny to 1H of barley. The syntenic map (**Figure 5.4**) displays this similarity between genomes. NLR gene clusters in these syntenic blocks were investigated to see which duplications took place before the wheat-barley divergence. Instances exist where barley possesses duplicated genes that remained individuals in wheat, and vice versa. The similarity between wheat and *A. tauschii* is

much closer, since *A. tauschii* contributed wheat's D subgenome only a few thousand years ago. While the A subgenome progenitor, *Triticum urartu*, has limited genomic availability, future studies may be able to assess differences between the NLR gene architecture in the two genomes. *A. tauschii* and barley provide excellent comparisons with wheat due to the relatively short period of time since their divergence. The similarities between R genes in wheat relatives show that the highly diverse family of R genes is necessary for survival, whereas the differences in number and phylogeny point to differences in selection pressure that these species each face.

CONCLUSION

In this study, clustering of R genes in wheat has been described, as compared to its progenitors and barley, a close relative. Gene similarities within clusters were assessed, showing that tandem duplication explains some of the diversification among R genes, along with segmental duplication and possible action by transposable elements. Wheat possesses 2151 NB-ARC-encoding genes, with many of those encoding the domains associated with functional NLRs that likely function as receptors, detecting pathogenic effectors. In wheat's 21 chromosomes, 547 clusters were found, with many of them containing highly similar genes. Future research should seek to functionally classify which pathogens these proteins trigger responses to and if duplications can be associated with known events in wheat's evolutionary history. Additional genomic data on *Aegilops speltoides*, a relative to the contributor of wheat's B subgenome, as well as availability of

data for *Triticum urartu*, contributor of wheat's A subgenome, will allow for more thorough analysis of the evolution of disease resistance genes in wheat.

ACKNOWLEDGEMENTS

Support for this dissertation work came from the USDA-NIFA hatch Projects to M. Nepal (SD00H469-13 and SD00H659-18), South Dakota Agriculture Experiment Station, and Department of Biology & Microbiology at South Dakota State University.

REFERENCES

1. FAOSTAT. FAO Database. 2013; http://www.fao.org/faostat/en/ - data. Accessed August 3, 2017.
2. Murray TD, Bockus, W.W., Bowden, R.L., Hunger, R.M., and Smiley, R.W., collators. Diseases of Wheat (Triticum spp. L.). *Common Names of Plant Diseases* 2015; http://www.apsnet.org/publications/commonnames/Pages/Wheat.aspx. Accessed July 28, 2017.
3. Pretorius Z, Singh R, Wagoire W, Payne T. Detection of virulence to wheat stem rust resistance gene Sr31 in Puccinia graminis. f. sp. tritici in Uganda. *Plant Disease.* 2000;84(2):203-203.
4. Singh RP, Hodson DP, Huerta-Espino J, et al. The emergence of Ug99 races of the stem rust fungus is a threat to world wheat production. *Annual review of phytopathology.* 2011;49:465-481.

5. Saari EE, Prescott J. World distribution in relation to economic losses. *The cereal rusts.* 1985;2(1):259-298.

6. Huerta-Espino J, Singh R, Germán S, et al. Global status of wheat leaf rust caused by Puccinia triticina. *Euphytica.* 2011;179(1):143-160.

7. Singh RP, Singh PK, Rutkoski J, et al. Disease Impact on Wheat Yield Potential and Prospects of Genetic Control. *Annual review of phytopathology.* 2016;54:303-322.

8. McFadden ES. A successful transfer of emmer characters to vulgare wheat. *J Am Soc Agron.* 1930;22(12):1020-1034.

9. Borlaug NE. Contributions of conventional plant breeding to food production. *Science.* 1983;219(4585):689-693.

10. Periyannan S, Moore J, Ayliffe M, et al. The gene Sr33, an ortholog of barley Mla genes, encodes resistance to wheat stem rust race Ug99. *Science.* 2013;341(6147):786-788.

11. Saintenac C, Zhang W, Salcedo A, et al. Identification of wheat gene Sr35 that confers resistance to Ug99 stem rust race group. *Science.* 2013;341(6147):783-786.

12. Ling H-Q, Zhao S, Liu D, et al. Draft genome of the wheat A-genome progenitor Triticum urartu. *Nature.* 2013;496(7443):87-90.

13. Jia J, Zhao S, Kong X, et al. Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature.* 2013;496(7443):91-95.

14. Marcussen T, Sandve SR, Heier L, et al. Ancient hybridizations among the ancestral genomes of bread wheat. *Science.* 2014;345(6194):1250092.

15. Kersey PJ, Allen JE, Christensen M, et al. Ensembl Genomes 2013: scaling up access to genome-wide data. *Nucleic acids research.* 2014;42(D1):D546-D552.

16. Jones P, Binns D, Chang H-Y, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014:btu031.

17. Jupe F, Pritchard L, Etherington GJ, et al. Identification and localisation of the NB-LRR gene family within the potato genome. *BMC genomics.* 2012;13(1):75.

18. Bailey TL, Boden M, Buske FA, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic acids research.* 2009;37(suppl_2):W202-W208.

19. Larkin MA, Blackshields G, Brown N, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23(21):2947-2948.

20. Kearse M, Moir R, Wilson A, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012;28(12):1647-1649.

21. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular biology and evolution.* 2016:msw054.

22. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research.* 2016:gkw290.

23. Darzentas N. Circoletto: visualizing sequence similarity with Circos. *Bioinformatics.* 2010;26(20).

24. Soderlund C, Bomhoff M, Nelson WM. SyMAP v3. 4: a turnkey synteny system with application to plant genomes. *Nucleic acids research.* 2011:gkr123.

25. Jones JD, Vance RE, Dangl JL. Intracellular innate immune surveillance devices in plants and animals. *Science.* 2016;354(6316):aaf6395.

26. Hayashi N, Inoue H, Kato T, et al. Durable panicle blast-resistance gene Pb1 encodes an atypical CC-NBS-LRR protein and was generated by acquiring a promoter through local genome duplication. *The Plant Journal.* 2010;64(3):498-510.

27. Baggs E, Dagdas G, Krasileva K. NLR diversity, helpers and integrated domains: making sense of the NLR IDentity. *Current opinion in plant biology.* 2017;38:59-67.

28. Andersen EJ, Ali S, Reese RN, Yen Y, Neupane S, Nepal MP. Diversity and Evolution of Disease Resistance Genes in Barley (Hordeum vulgare L.). *Evolutionary bioinformatics online.* 2016;12:99.

29. Andersen EJ, Nepal MP. Genetic diversity of disease resistance genes in foxtail millet (Setaria italica L.). *Plant Gene.* 2017;10:8-16.

30. Marone D, Russo MA, Laidò G, De Leonardis AM, Mastrangelo AM. Plant nucleotide binding site–leucine-rich repeat (NBS-LRR) genes: active guardians in host defense responses. *International journal of molecular sciences.* 2013;14(4):7302-7326.

31. Abdullah S. Characterization of Pyrenophora Tritici-Repentis in Wheat and Rye to Study Tan Spot Susceptibility and Insights Into It's Relationship with Stem Rust Resistance. 2017.

32. Bourras S, McNally KE, Ben-David R, et al. Multiple avirulence loci and allele-specific effector recognition control the Pm3 race-specific resistance of wheat to powdery mildew. *The Plant Cell.* 2015;27(10):2991-3012.

33. Kushalappa AC, Yogendra KN, Karre S. Plant innate immune response: qualitative and quantitative resistance. *Critical Reviews in Plant Sciences.* 2016;35(1):38-55.

34. Leister D. Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. *Trends in genetics.* 2004;20(3):116-122.

35. Kim S, Park J, Yeom S-I, et al. New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome biology.* 2017;18(1):210.

36. Middleton CP, Senerchia N, Stein N, et al. Sequencing of chloroplast genomes from wheat, barley, rye and their relatives provides a detailed insight into the evolution of the Triticeae tribe. *PLoS One.* 2014;9(3):e85761.

CHAPTER 6: DIVERSIFICATION OF DISEASE RESISTANCE RECEPTORS THROUGH INTEGRATED DOMAIN FUSIONS IN WHEAT AND ITS PROGENITORS

This chapter has been accepted for publication in the *Journal of Botanical Research*: Andersen, E.J. and M.P. Nepal. In Press. Diversification of Disease Resistance Receptors Through Integrated Domain Fusions in Wheat and its Progenitors. *Journal of Botanical Research.*

ABSTRACT

Pathogenic effectors inhibit plant resistance responses by interfering with intracellular signaling mechanisms. To deal with rapidly evolving effectors, plant Nucleotide-binding, Leucine-rich repeat Receptors (NLRs) have evolved highly variable effector-recognition sites. While many NLRs utilize variable Leucine-Rich Repeats (LRRs) to bind to effectors, some have gained Integrated Domains (IDs) necessary for receptor activation or downstream signaling. While a few studies have identified IDs within NLRs, the homology and regulation of these genes have yet to be elucidated. A diverse set of wheat NLR-ID fusion proteins were identified as candidates for NLR functional diversification through ID effector recognition or signal transduction. IDs were compared with homologs in wheat progenitors and other grasses, revealing evolutionary conservation of >80% amino acid sequence similarity. ID homology indicates that these domains originated as functional, non-NLR-encoding genes and were incorporated into NLR-encoding genes through duplication. Multiple NLR-ID genes encode alternative transcripts that include or exclude IDs. Expression of these alternative transcripts can be experimentally verified in wheat tissue. This indicates that plants employ alternative

splicing to regulate IDs, possibly using them as baits, decoys, and functional signaling components. NLR-ID diversity corresponds directly with the various signaling components essential to defense responses. NLR-ID fusion proteins greatly expand the potential functions for immune receptors, possibly removing the need for intermediate signaling factors that are often targeted by effectors. Genomic and expression data support the hypothesis that wheat uses alternative splicing to include and exclude IDs from NLR proteins. Based on their homology, IDs are evolutionarily conserved in non-NLR sequences of distantly related species, indicating retention of functionality. Future studies should aim to characterize NLR-ID fusion protein structures, demonstrate ID function (e.g. kinase activity), and associate alternative splicing with specific conditions.

INTRODUCTION

Plant innate immune systems utilize specialized receptor proteins to detect pathogens [1]. Nucleotide-binding, Leucine-rich repeat Receptor (NLR) proteins detect pathogen effectors that would otherwise inhibit host resistance responses [2]. In order to detect the hundreds of pathogens and pests, immune receptors must be able to respond to many elicitors. To accomplish this, NLRs have radiated to form a diverse family of resistant genes in plants [3]. Much of this diversity is accomplished by gene duplication and variation in NLR Leucine-Rich Repeats (LRRs), which allow NLRs to bind to new effectors [4]. Diversification has led to the formation of networks of sensor and helper NLRs, with some NLRs dimerizing to initiate signaling [5-8]. As another form of diversification, some NLRs have gained extra domains that may assist the receptor in

pathogen recognition or in resistance signaling. These domains, called Integrated Domains (IDs), resulted from a fusion of NLR and functional domains also involved in resistance, as outlined by the integrated decoy/sensor model [9,10]. ID diversity has been probed across many plant genomes, revealing a diversity of domains associated with potential roles in resistance [11].

The major objectives of this study are to identify wheat NLR-ID fusion proteins, infer their function, and assess their homology in wheat relatives and other selected monocot species. A method is also proposed to describe how plants utilize for NLR-ID regulation, which became apparent while manually assessing the transcript and protein sequences. Understanding the evolution of NLR-ID fusions provides a unique perspective on NLR diversification, which often gets ignored by large-scale analyses of NLR gene family evolution. These findings will improve the understanding of how NLRs diversify to oppose various pathogenic molecular weapons.

METHODS

Triticum aestivum protein sequences were downloaded using the Biomart application within the Ensembl Genomes [40] and Phytozome [41] databases. InterProScan annotations [42] were compiled and proteins containing NB-ARC domains (PF00931) were investigated. Pfam annotations not inherently part of NLR structure were assembled. Amino acid sequences and corresponding annotations were uploaded to the program Geneious [43] for sequence alignment, homology assessment and motif visualization. The IDs were manually investigated to assess protein location, potential function, homology

to proteins other species, and presence in variant transcripts. Function was assessed partially through domain descriptions available through the Pfam database [44], allowing for inferences about domain activity. Genomes investigated for homology include: *Aegilops tauschii*, *Amborella trichopoda*, *Arabidopsis thaliana*, *Brachypodium distachyon*, *Hordeum vulgare*, *Musa acuminata*, *Oryza sativa*, *Setaria italica*, *Triticum urartu*, and *Zea mays*. Genomic data was not available for *Aegilops speltoides*, which is believed to be the contributor of wheat's B genome. The Gene Structure Display Server 2.0 [45] was used to visualize alternative splicing of NLR-IDs. Wheat expression data was generated from datasets in NCBI and Wheat Gene Expression Atlas data [38].

RESULTS

IDs in Wheat

I have identified wheat NLR proteins with IDs that potentially function as molecular baits, decoys, or signal transduction factors. Wheat NLRs possess a diverse set of IDs, the most commonly occurring are kinase and DNA-binding domains. **Figure 6.1** shows the average location of these domains relative to protein length, averages calculated from every NLR-ID occurrence of that domain. Kinase domains are generally located in the N-terminal half of the protein with tyrosine kinase domains generally in the middle of the protein sequence. The more diverse class of DNA-binding domains vary by domain type, present in both the N- as well as C-termini. For example, Myb-like and BED zinc finger domains are generally at the N-terminus, while B3 and WRKY domains are at the C-terminus. Many other IDs located at the C-terminus have potential roles in

signaling, such as calmodulin-binding, jacalin-like lectin, thioredoxin, and ubiquitin-conjugating.

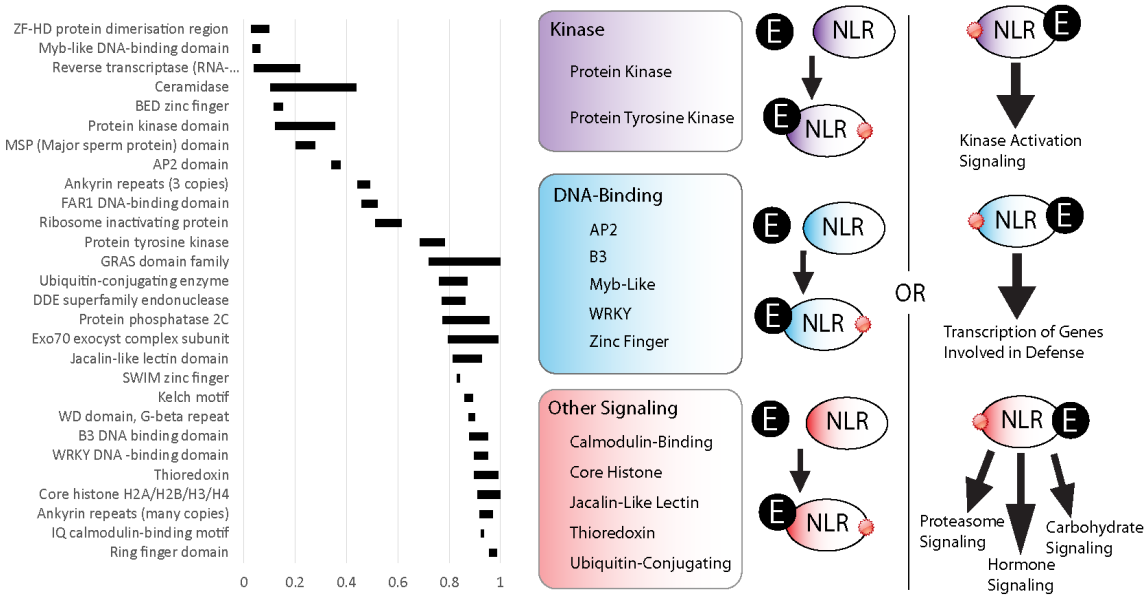
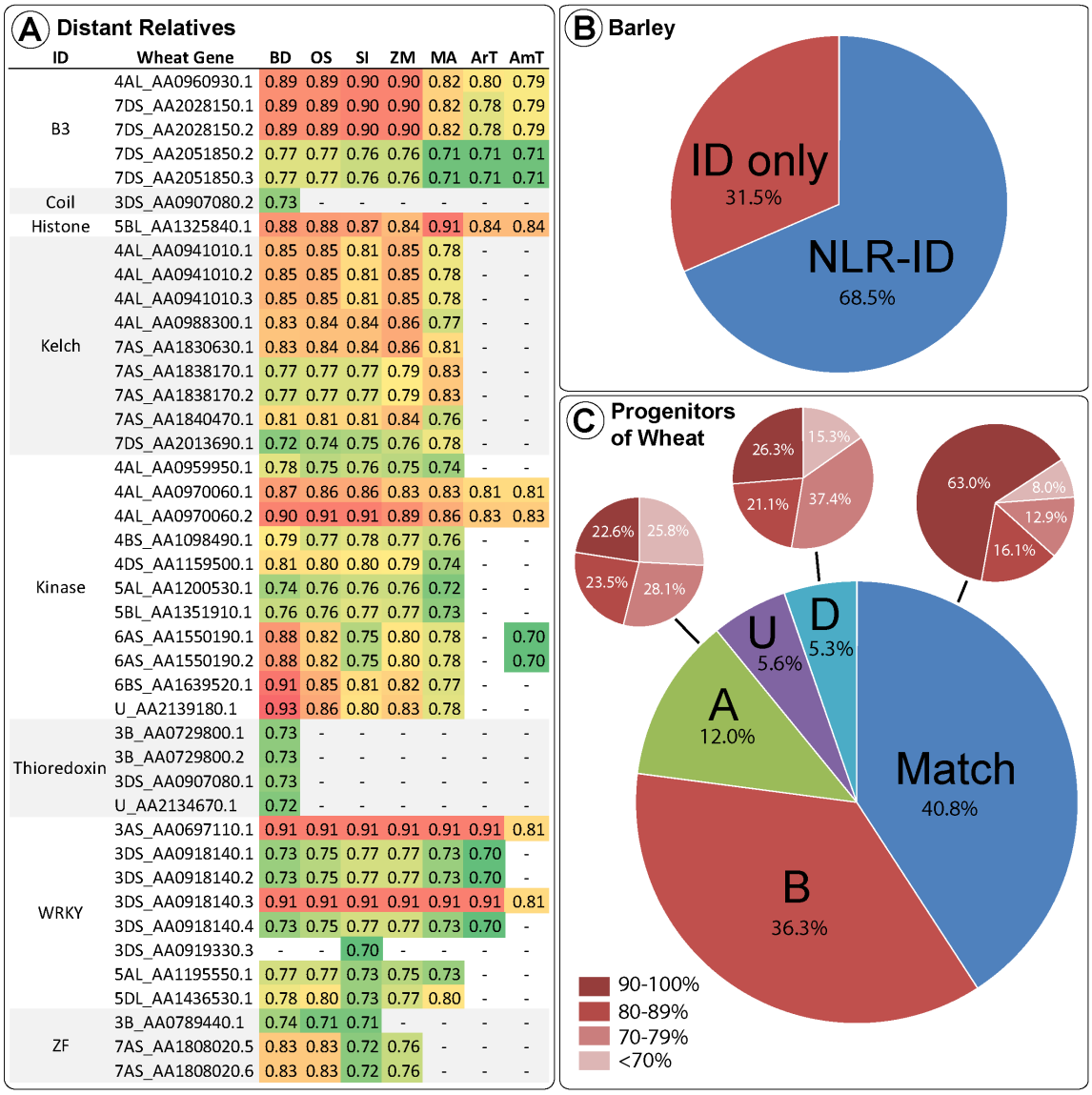


Figure 6.1 Integrated Domain (ID) locations, indicated by black rectangles, are shown within NLRs relative to protein length (0-1). IDs were grouped into functional categories, based on their potential involvement in kinase, DNA-binding, or other signaling activities (shown in purple, blue, and red, respectively). Schematic diagrams representing potential functions for these NLR-IDs are included with pathogenic effectors represented by black circles (labeled as 'E') and NLR-ID proteins as ovals color coded by ID type (i.e. 'Kinase', 'DNA-Binding', or 'Other Signaling'). The diagram includes representations of both effector-bait interaction (left) and NLR-ID signaling (right) that these domains may be involved in. Red circles at the sides of NLRs indicate activated NLR proteins.

ID Homology

Many *Triticum aestivum* (TA) IDs share homology with proteins in distantly related monocots. **Figure 6.2** shows the wheat accessions with high percent identity (above 70%), grouped by ID type and homolog species. The vast majority of these homologs do not contain NB-ARC domains, indicating a recent fusion. ID homologs in distant relatives generally were not NLR proteins. While other plants also possess NLR-ID fusions, many are lineage specific and are not conserved across diverse species. Barley, a close relative of wheat, possessed many of the same NLR-ID fusion proteins as in wheat, with 68.5% of ID homologs in barley also possessing NLRs. The two progenitors of wheat with sequenced genomes, *T. urartu* and *A. tauschii*, also possess wheat's NLR-ID fusions. Of these progenitor homologs, 40.8% matched the subgenome expected based upon the known progenitor-subgenome relationships. Genomes investigated for homology include: *Aegilops tauschii* (AT), *Amborella trichopoda* (AmT), *Arabidopsis thaliana* (ArT), *Brachypodium distachyon* (BD), *Hordeum vulgare* (HV), *Musa acuminata* (MA), *Oryza sativa* (OS), *Setaria italica* (SI), *Triticum urartu* (TU), and *Zea mays* (ZM). Genomic data was not available for *Aegilops speltoides* (AS).



ID	Wheat Gene	BD	OS	SI	ZM	MA	ArT	AmT
B3	4AL_AA0960930.1	0.89	0.89	0.90	0.90	0.82	0.80	0.79
	7DS_AA2028150.1	0.89	0.89	0.90	0.90	0.82	0.78	0.79
	7DS_AA2028150.2	0.89	0.89	0.90	0.90	0.82	0.78	0.79
	7DS_AA2051850.2	0.77	0.77	0.76	0.76	0.71	0.71	0.71
	7DS_AA2051850.3	0.77	0.77	0.76	0.76	0.71	0.71	0.71
Coil	3DS_AA0907080.2	0.73	-	-	-	-	-	-
Histone	5BL_AA1325840.1	0.88	0.88	0.87	0.84	0.91	0.84	0.84
Kelch	4AL_AA0941010.1	0.85	0.85	0.81	0.85	0.78	-	-
	4AL_AA0941010.2	0.85	0.85	0.81	0.85	0.78	-	-
	4AL_AA0941010.3	0.85	0.85	0.81	0.85	0.78	-	-
	4AL_AA0988300.1	0.83	0.84	0.84	0.86	0.77	-	-
	7AS_AA1830630.1	0.83	0.84	0.84	0.86	0.81	-	-
	7AS_AA1838170.1	0.77	0.77	0.77	0.79	0.83	-	-
	7AS_AA1838170.2	0.77	0.77	0.77	0.79	0.83	-	-
	7AS_AA1840470.1	0.81	0.81	0.81	0.84	0.76	-	-
	7DS_AA2013690.1	0.72	0.74	0.75	0.76	0.78	-	-
Kinase	4AL_AA0959950.1	0.78	0.75	0.76	0.75	0.74	-	-
	4AL_AA0970060.1	0.87	0.86	0.86	0.83	0.83	0.81	0.81
	4AL_AA0970060.2	0.90	0.91	0.91	0.89	0.86	0.83	0.83
	4BS_AA1098490.1	0.79	0.77	0.78	0.77	0.76	-	-
	4DS_AA1159500.1	0.81	0.80	0.80	0.79	0.74	-	-
	5AL_AA1200530.1	0.74	0.76	0.76	0.76	0.72	-	-
	5BL_AA1351910.1	0.76	0.76	0.77	0.77	0.73	-	-
	6AS_AA1550190.1	0.88	0.82	0.75	0.80	0.78	-	0.70
	6AS_AA1550190.2	0.88	0.82	0.75	0.80	0.78	-	0.70
	6BS_AA1639520.1	0.91	0.85	0.81	0.82	0.77	-	-
	U_AA2139180.1	0.93	0.86	0.80	0.83	0.78	-	-
Thioredoxin	3B_AA0729800.1	0.73	-	-	-	-	-	-
	3B_AA0729800.2	0.73	-	-	-	-	-	-
	3DS_AA0907080.1	0.73	-	-	-	-	-	-
	U_AA2134670.1	0.72	-	-	-	-	-	-
WRKY	3AS_AA0697110.1	0.91	0.91	0.91	0.91	0.91	0.91	0.81
	3DS_AA0918140.1	0.73	0.75	0.77	0.77	0.73	0.70	-
	3DS_AA0918140.2	0.73	0.75	0.77	0.77	0.73	0.70	-
	3DS_AA0918140.3	0.91	0.91	0.91	0.91	0.91	0.91	0.81
	3DS_AA0918140.4	0.73	0.75	0.77	0.77	0.73	0.70	-
	3DS_AA0919330.3	-	-	0.70	-	-	-	-
	5AL_AA1195550.1	0.77	0.77	0.73	0.75	0.73	-	-
	5DL_AA1436530.1	0.78	0.80	0.73	0.77	0.80	-	-
ZF	3B_AA0789440.1	0.74	0.71	0.71	-	-	-	-
	7AS_AA1808020.5	0.83	0.83	0.72	0.76	-	-	-
	7AS_AA1808020.6	0.83	0.83	0.72	0.76	-	-	-

Figure 6.2 Wheat IDs and their homologs in wheat progenitors and other divergent monocot species are shown, including *Arabdopsis* (model plant species) and *Amborella* (basal known Angiosperm). (A) **Sequence similarities above 70% are shown between wheat IDs and their homologs in Brachypodium (BD), rice (OS), foxtail millet (SI), maize (ZM), banana (MA), *Arabidopsis* (Art), and *Amborella* (AmT). Wheat accession names are shortened to only include the chromosome arm and the last digits unique to each transcript.** (B) **Barley ID homologs possessing and lacking NLR domains are compared.** (C) **Mapping of homologs among wheat and**

wheat progenitors is displayed – a match between the progenitor and subgenome (labeled 'Match'); subgenome A protein was more similar to an *Aegilops tauschii* sequence (labeled 'A'); subgenome D protein was more similar to a TU sequence (labeled 'D'); sequence was from the B subgenome with the unavailable AS progenitor (labeled 'B'), or the accession subgenome is unknown (labeled 'U'). Also, the level of homology between the pairs is demonstrated for 'Match', 'A', and 'D', with dark red corresponding to the proportion of sequences with high similarity (>90%) and lighter red corresponding to lower similarity (<70%).

NLR-ID Regulation

Some NLR-ID genes encode alternative transcripts that omit IDs or other domains, such as transmembrane helices. **Figure 6.3** illustrates a consolidation of all wheat NLR-IDs in which another transcript of the same gene excluded the ID. Alternative splicing of this kind would allow plants to regulate the use of IDs by including or excluding exons containing them. Similar characteristics were also observed in barley transcripts, indicating a conserved use of alternative splicing. Alternative transcripts may also be found in wheat progenitors, which currently lack available data. Expression data from the Wheat Gene Expression Atlas and NCBI shows differential expression between these alternative transcripts, which are shown in **Supplementary Figure 6.1** and **Supplementary Table 6.1**.

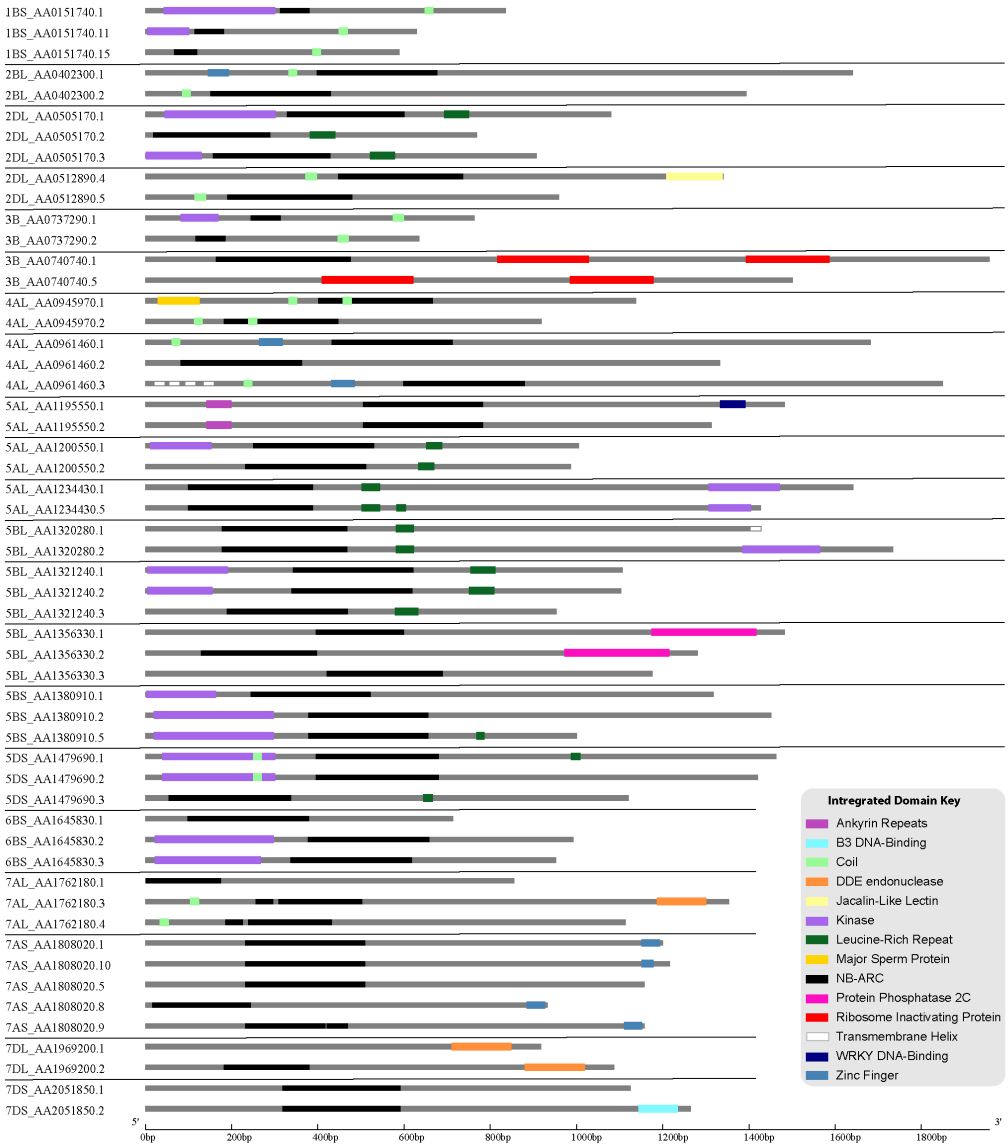


Figure 6.3 Wheat NLR-ID genes that encode alternative transcripts excluding or truncating IDs or other NLR domains are shown. Grey bars span the exon length in base pairs, black bars represent NB-ARC domains. Color segments annotate ID locations by domain type, which are defined in the integrated domain key. The scale, in base pairs, is shown along the bottom of the figure and black bars separate each set of transcripts. Wheat accession names are shortened to only show the chromosome arm, last code of gene name, and transcript number.

DISCUSSION

IDs Augment NLR Function Through Signaling and Recognition

Kinase and DNA-binding IDs likely function as signaling domains that help NLRs initiate defense responses. Current models of NLR function describe a conformational shift triggered when pathogenic effectors bind to the C-terminal LRR, causing the NB-ARC to exchange ADP for ATP, opening the protein up for the N-terminus to initiate further signaling [12-14]. LRRs, as highly variable domains of repeating Lxx amino acid residues, allow defense receptors to bind to diverse elicitors. The NB-ARC, as a P-loop-containing nucleoside triphosphate hydrolase, functions in hydrolysis of beta-gamma phosphate bonds in ATP, binding to phosphates using the Walker A (P-loop) motif and to magnesium ions necessary for catalysis by Walker B motifs [15]. This release of energy from ATP hydrolysis drives protein conformational change, allowing N-terminal domains (i.e. TIR or CC) to trigger downstream signaling. Kinase IDs found in wheat NLRs could initiate signaling through phosphorylation of transcription factors or other kinases (i.e. MAPK). Sarris et al. (2016) also found an abundance of NLR-kinase fusions, which possibly retain their biochemical activity [11]. DNA-binding domains could move directly to the nucleus upon activation, binding to promoters of pathogenesis-related (PR) genes to recruit transcription machinery. IDs that likely bind to DNA include: AP2, B3, Zinc Finger, Myb, and WRKY domains, which have been shown to play roles in pathogen resistance [16,17]. The *Arabidopsis* NLR gene AT4G12020 has been identified both as MAPKKK11 [18] and a TNL resistance gene [19], containing WRKY DNA-binding sites and a protein kinase domain. This gene is a homolog of SLH1, which

has been associated with hypersensitive response, possibly guarding a pathogen effector target [20]. Many NLR-ID fusion proteins contain transmembrane (TM) domains or nuclear localization signals (NLSs). Several proteins have multiple transmembrane domains, with proteins like 3B_AA0787000 containing seven, characteristic of other transmembrane proteins. NLSs indicate that DNA-binding domains may functionally interact with DNA as transcription factors.

In addition to signaling, some IDs may play direct roles in effector recognition as effector-binding domains or bait domains that mimic effector targets (**Figure 6.1**). Jacalin-like lectin domains, for example, bind to carbohydrates and can recognize carbohydrates that originate directly from pathogens or from damage incurred during infection [21-23]. Mannose-binding lectin domains were also found in NLRs, associated with disease resistance [24], along with "Wall-associated receptor kinase galacturonan-binding" and "Cleavage site for pathogenic type III effector avirulence factor Avr" domains. Lectin domains may distinguish proteins as helper NLRs, with carbohydrates acting as signals to initiate NLR activation. Other domains may play roles in effector recognition as bait domains that resemble effector targets. The resistance protein RRS1 becomes activated when an integrated WRKY domain interacts with *Ralstonia solanacearum* effector PopP2 and *Pseudomonas syringae* pv. *pisi* effector AvrRps4, effectors that otherwise target WRKY transcription factors [25,26]. Wheat NLR-WRKY fusions share homology with WRKY16, WRKY19, WRKY46, and WRKY54/70, with potential roles as targets, especially WRKY46, which is associated with bacterial resistance [11]. Accession 5DL_AA1436530 contains two variants of WRKY domains (WRKY and WSKY), possibly providing diverse baits for effectors. Accession 2BL_AA0441310

encodes a protein with separate AP2 and BED zinc finger domains, either allowing it to bind to separate promoters or as bait for more than one effector. Some bait proteins, such as PBS1, are kinases that pathogen effectors target for degradation, increasing the utility of NLR-kinase fusions. The Rosetta stone theory describes this association between fusions and linkage between protein function [27]. Several proteins with IDs and NB-ARCs do not contain LRRs, which would not be required for activation since baits have replaced LRRs in function.

The activity of IDs as baits is further supported by ID diversity, which corresponds to the diversity of defense regulatory components. IDs found in NLRs are also found in proteins that effectors target to interfere with defense. Several domains correspond to proteins involved in resistance signaling: calmodulin-binding (calcium signaling), Gibberellic acid insensitive (GAI) Repressor of GAI And Scarecrow (GRAS; gibberellin signaling), and ethylene responsive element binding (ethylene signaling). Several different domains contain IDs associated with the proteasome or ubiquitin, also involved in regulating resistance: protease subunit, proteasome component signature, cullin-repeat, RING/U-box, ubiquitin conjugating enzyme, and WD domains. Some IDs contain domains associated with regulation of DNA expression: core histone and chromatin organization modifier. Other IDs correspond to proteins involved in resistance responses: ribosome inactivating and ricin domains (disrupt ribosome activity), thioredoxin and kelch (oxidase activity in reactive oxygen species production), alpha subunit of tryptophan synthesis (synthesis of antiherbivory and antimicrobial compounds), Exo70 exocyst complex subunit (transport of antimicrobial compounds out of the cell), and DDE endonuclease (apoptosis). A few other domains are likely

associated with pathogen components: major sperm protein (nematode sperm function, targeted by plant RNA interference), FNIP (found in *Dictyostelium discoideum*), and reverse transcriptase (inhibition of viral infection). Additional viral IDs include: RNA-binding/recognition, retrovirus zinc finger-like domain, and integrase domains. IDs may also be associated with pathogen-derived resistance and RNAi that plants use to inhibit viruses and other pathogens. Other studies support this diversity in IDs, showing similar results in other plant species [11,28].

IDs Originate as Functional Domains and Close Relatives Share NLR-IDs

Bread wheat is an allohexaploid species resulting from two separate hybridization events and substantial artificial selection [29]. As a monocot, wheat shares distant relationships with other members of the family Poaceae (i.e. BD, ZM, SI, and OS). Naturally, much strong similarities exist between wheat and other members of the Triticeae tribe, including HV and wheat progenitors TU, AS, as well as AT (**Figure 6.2**). ID homologs in distant relatives generally do not contain NB-ARCs, indicating relatively recent origin of NLR-ID fusions. IDs with high percent similarity to homologs, indicative of functional retention, include: proteasome subunit, B3 DNA-binding, WRKY DNA-binding, core histone, protein kinase, and kelch motif. Other domains with moderate similarity include: jacalin-like lectins, ribosome inactivating protein, BED zinc finger, SWIM zinc finger, ZF-HD protein dimerization region, zinc knuckle, protein phosphatase 2C, tyrosine kinase, thioredoxin, major sperm protein, reverse transcriptase, and DDE endonuclease. Homologs may be obscured since mutations accumulate in regions not essential for function or effector-bait interaction, causing divergence from the original

sequences and making homology difficult to assess. Some mutations may increase functionality of NLR-IDs, since the original ID sequence was functionally optimized within a different protein. Some IDs may serve as baits for multiple targets, if targets possess similar modification/cleavage sites (e.g. similar WRKY domains).

Many IDs showed high homology in distant relatives. Kinase domains of up to 300 amino acids in length were over 80% similar to homologs. DNA-binding domains also had high homology in distant relatives. WRKY DNA binding domains present in wheat and progenitors have 90.5% similarity to several non-NLR genes in AT, BD, MA, OS, SI, and ZM. Many other IDs in wheat and its progenitors share >80% similarity with homologs in SI, ZM, BD, OS, AT, MA, and AmT. These results concur with previous investigations into IDs, where conserved IDs were identified in diverse plant species [11,30]. Some wheat proteins are very similar to their homologs in TU and AT, whereas others provide examples of one species diversifying from the other two. The histone ID in wheat protein 5BL_AA1325840 (approximately 100 amino acids) shares strong homology (>80%) with proteins in MA, BD, OS, SI, ArT, AmT, and ZM, a recent fusion not present in wheat relatives. Greater than 90% similarity was observed between the 182 amino acid long F775_12304|EMT01588 proteasome subunit domain and proteins in BD, OS, SI, and ZM. While this indicates that these accessions are close homologs, none of the other accessions have NB-ARCs, only peptidase, proteasome subunit, and nucleophile aminohydrolase domains. HV, TU, and TA homologs to this domain, while matching the sequence 100%, do not have NB-ARCs, indicating a very recent duplication and then fusion, after the hybridization of hexaploid wheat. Kelch motif IDs were found in one TU, three AT, and six TA proteins. Interestingly, only one of the TA proteins is in

the D subgenome and 5 in the A subgenome, when the opposite would be expected based upon subgenome origin.

While domain homology in distantly related species indicates functional origins of IDs, homologs identified in close relatives (i.e. HV) and wheat progenitors (TU and AT) indicate recent fusions and duplications. Unlike distant relatives, barley shares many NLR-ID fusions with wheat. This indicates that many of wheat's NLR-IDs happened before the divergence of barley and wheat progenitors. Since this divergence, wheat and barley have independently gained and lost NLR-ID proteins. Many TU and AT proteins are almost identical to proteins encoded by TA genes within the A and D subgenomes. IDs within wheat's B subgenome often originate from AS and do not have 100% homologs in TU and AT. In select cases, similarity was found between NLR-IDs and functional domains from non-NLR proteins, indicating potential NLR-ID fusions since the formation of wheat. Conversely, some close wheat relatives share homology with distantly related NLR-ID fusions, such as F775_00546|EMT17242 and Si008625m, with an 84.6% similarity (991 identical sites) between their whole sequences, both with NB-ARCs and kinase domains. Unexpectedly, many proteins were found in wheat but not a relative, or vice versa. These domains include: histone, ribosome inactivating protein, calcium signaling, cleavage for type III effectors, RNase H type, P450, antibiotic synthesis, and ubiquitin conjugating enzyme. Other domains were found in greater or lower numbers in wheat compared to its progenitors, such as DDE endonuclease and reverse transcriptase, indicating loss or duplication in one genome.

Plants Use Alternative Splicing to Regulate NLR-IDs

Many NLR-ID protein-encoding genes possess multiple transcripts, some of which lack IDs or truncate domains within protein. Examples found in wheat are displayed in **Figure 6.3**. This indicates that plants may use alternative splicing to regulate the use of IDs within a network of NLR proteins. Previous studies have shown that resistance to some pathogens requires alternative splicing [31,32], such as *RPS4* in *Arabidopsis* [33]; and splicing is used to truncate proteins like *RCT1* in *Medicago truncatula* [34]. Wheat has also shown evidence of alternative splicing of important resistance genes Lr10 and Sr35 [35,36]. Splicing patterns between wheat paralogs resulting from duplication also appear to be conserved. Stop codon-containing inter-exon regions can be included in the transcript to force a truncation of the protein. Truncated NB-ARCs may result in decoy proteins, where signaling function is lost but IDs 'distract' pathogenic effectors from functional target proteins. The potential involvement of alternative splicing is shown in **Figure 6.4**, integrated into the varied use of IDs in NLR proteins. In concurrence with the results, Yang et al. (2014) describe NLR alternative splicing as useful for regulating NLR autoinhibition or function in signal transduction, also detailing potential transcription factor activity [31]. Genes that have multiple copies of an ID can also regulate the number of copies in the protein through alternative splicing. Accession 7BL_AA1863630 encodes two transcripts, one with three WD domains (G-beta repeat) and the other with two, along with Coils and NB-ARCs. Accession 3B_AA0740740 encodes five transcripts that all contain two separate ribosome inactivating protein domains, four of which also have NB-ARCs. Alternative splicing may also allow the plant to select different localization for a gene product. For example,

4AL_AA0961460, encodes three transcripts, one with multiple transmembrane helices and a BED zinc finger ID, one lacking the transmembrane helices, and another lacking the transmembrane helices and the ID. The transmembrane-containing transcript is encoded by one and a half extra exons at the beginning of the transcript, with a total of five exons making up the gene.

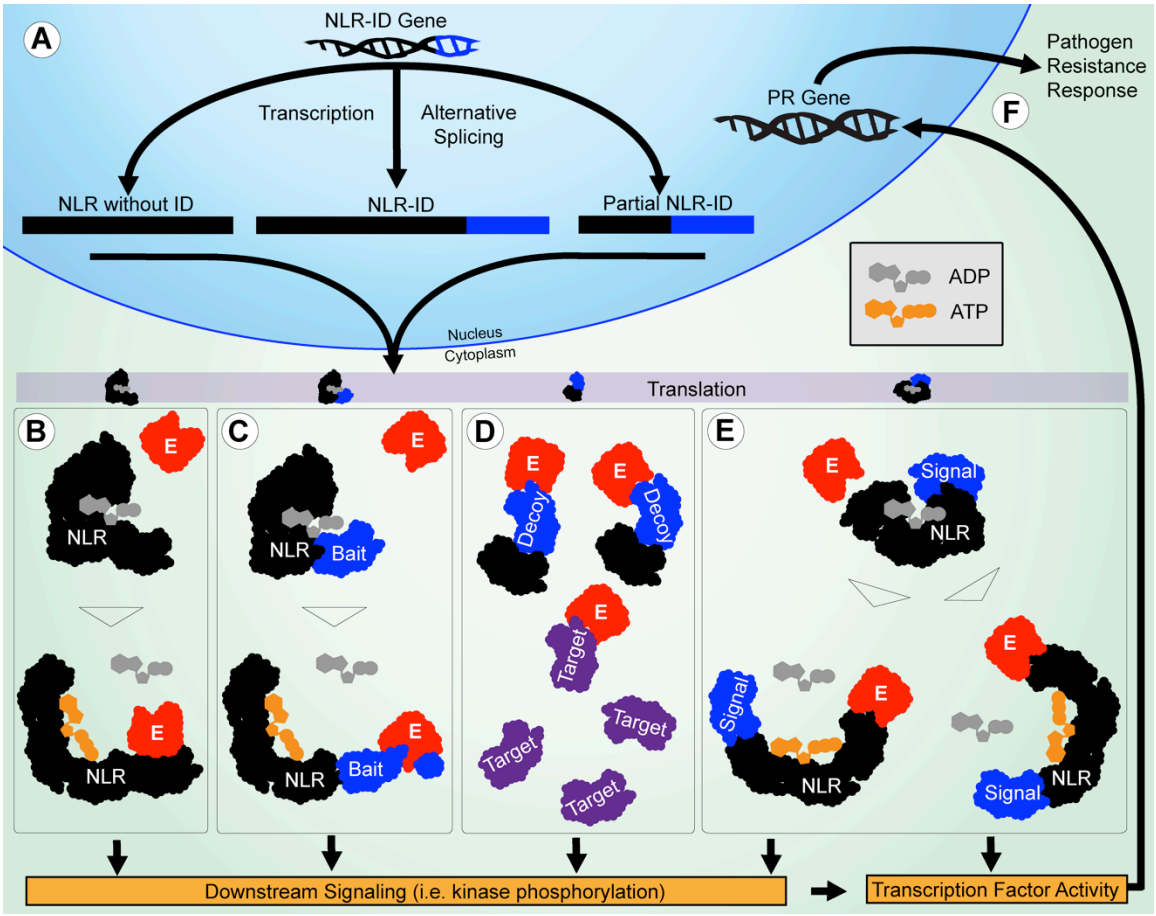


Figure 6.4 Potential roles of IDs in functional diversification of NLRs in pathogen resistance are illustrated. (A) **The NLR-ID gene is alternatively spliced during transcription to include or exclude IDs. The NLR and ID sequences are shown in black and blue, respectively.** (B) **The NLRs without IDs function through effector-LRR binding to activate the protein, and trigger downstream signaling. Effectors are shown in red.** (C) **When IDs are used as baits, they mimic pathogen targets, and**

cause NLR activation after they are modified. (D) **When IDs are used as decoys, they mimic pathogen targets to reduce effector interference in resistance signaling. The targets are shown in purple.** (E) **When IDs are used in signaling, they allow NLRs to act as signal transduction factors, less reliant on downstream signaling utilized by other NLRs.** (F) **Finally, transcription factor activity directly involving or triggered by NLRs causes PR genes to be expressed, leading to a resistance response.**

Wheat expression data shows that there are differences in the expression of these alternative transcripts shown in **Figure 6.3**. The expression values for the 54 transcripts present in **Figure 6.3** were mined from datasets present in the Wheat Gene Expression Atlas and NCBI databases. Expression data from Salcedo et al. (2017) shows that alternative splicing may result from different conditions [37]. At very least, these data provide support for **Figure 6.3** accessions as actual alternative transcripts that can be measured experimentally. This data is present in **Supplementary Figure 6.1** and **Supplementary Table 6.1**. In the Wheat Gene Expression Atlas data [38], several transcripts with different ID contents show differential expression in wheat tissues. For instance, both transcripts of 5AL_AA1195550 were expressed, much more for the WRKY-containing first transcript. Similar examples include 5AL_AA1200550 (kinase), 5BL_AA1321240 (kinase), 5BS_AA1380910 (LRR), 5DS_AA1479690 (kinase and coil), 7AS_AA1808020 (zinc finger), and 7DS_AA2051850 (B3). Accession 3B_AA0740740 shows evidence of higher expression of a truncated NLR-ID, lacking an NB-ARC. Several of these genes were more strongly expressed in the leaves, shoots, and spikes. One exception is 7DS_AA2051850, where the second B3-containing transcript is

expressed much more in the roots than in any other tissue, and much higher than the other B3-lacking transcript in the roots. This gene may be involved in resistance to soil-borne pathogens. More data is required to conclusively show differential expression based upon certain treatments and conditions.

While wheat shows evidence of NLR-ID alternative splicing, barley may have evolved a more diverse set of transcripts for NLR-IDs. Several barley NLR-ID proteins have dozens of transcripts, with several of those allowing for alternative use of IDs in NLR proteins. Many barley genes have alternative transcripts that encode NLR-ID, just NLR, just ID, or lack both. For example, barley gene HORVU3Hr1G010980 encodes 16 transcripts, many of which contain NB-ARCs and two separate ribosome inactivating domains, one with just the NB-ARC, one just the ribosome inactivating domain, and one short transcript without any domains. Previous studies have identified barley *Mla* genes as utilizing alternative splicing for resistance [39]. Barley genes can also encode multiple IDs. HORVU3Hr1G037800 and HORVU7Hr1G000320 can encode NB-ARC transcripts, Kelch motif transcripts, a Glutaredoxin transcripts, or a short transcript with none of those domains. HORVU1Hr1G079170 contains dozens of transcripts that can either encode NB-ARC and LRR protein or proteins with Glycoside hydrolase family 28 and TM helix proteins, potentially involved in two different regulatory pathways. HORVU7Hr1G120020 encodes 17 transcripts with NB-ARC, Coil, and Cleavage site for pathogenic type III effector avirulence factor Avr, with some transcripts only encoding the cleavage site with no NB-ARC, a possible decoy for pathogenic effectors. Barley gene HORVU5Hr1G085900 encodes 20 transcripts, some as NB-ARC-WRKY, just WSKY, and NB-ARC-WRKY-WSKY. Since some transcripts just contain the

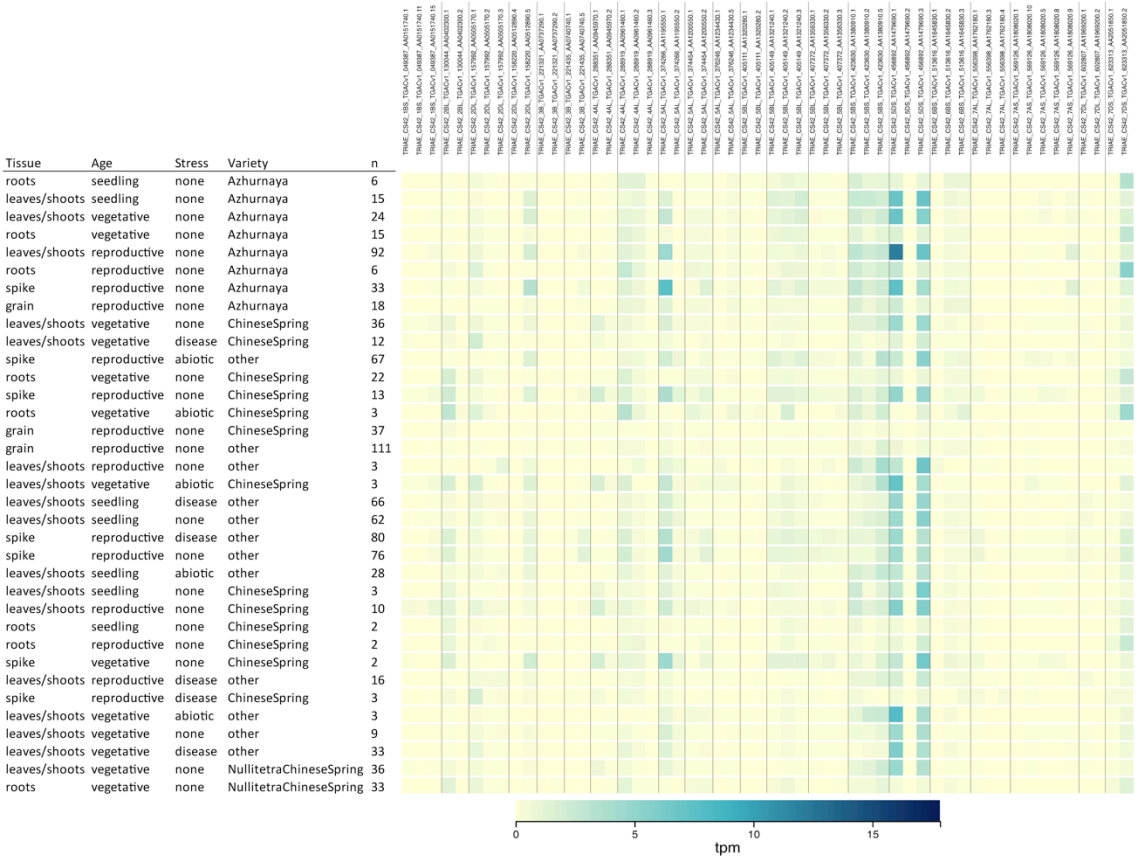
transcription factor domain, either this is functioning as a transcription factor, or it is a non-receptor decoy that reduces effector interference in WRKYs necessary for resistance response. This protein also contains a probable nuclear localization signal between the NB-ARC and the WRKY domains. This data supports a previous prediction that alternative splicing may allow for differential cellular localization [31].

CONCLUSIONS

I found that the diversity of integrated domains in NLRs corresponds directly to the multiple components utilized by plant cells to initiate resistance responses, such as kinases, transcription factors, hormone signaling receptors, and proteins involved in antimicrobial compound production. NLR-ID fusions give these immune receptors the potential to function as effector baits, decoys, and signal transduction factors. This functional diversification would allow plants to stop using intermediate signaling factors that effectors often target to inhibit resistance. Sequence homology both indicates that some IDs retain functionality and provides an explanation for ID origins as functional non-NLR proteins before their integration into NLRs. Using both genomic and expression data, these results showed that plants likely utilize alternative splicing to regulate the inclusion or exclusion of IDs in NLR proteins and have compiled a list of experientially verified alternative transcripts that truncate or exclude these domains. This optional deployment of NLR-IDs constitutes an important defense strategy to deal with rapidly evolving pathogen effectors. Future studies should aim to characterize the

structure of NLR-ID fusion proteins, demonstrate which IDs have retained enzymatic activity, and associate the expression of alternative transcripts with specific conditions.

SUPPLEMENTARY DATA



Tissue	Age	Stress	Variety	n
roots	seedling	none	Azhurnaya	6
leaves/shoots	seedling	none	Azhurnaya	15
leaves/shoots	vegetative	none	Azhurnaya	24
roots	vegetative	none	Azhurnaya	15
leaves/shoots	reproductive	none	Azhurnaya	92
roots	reproductive	none	Azhurnaya	6
spike	reproductive	none	Azhurnaya	33
grain	reproductive	none	Azhurnaya	18
leaves/shoots	vegetative	none	ChineseSpring	36
leaves/shoots	vegetative	disease	ChineseSpring	12
spike	reproductive	abiotic	other	67
roots	vegetative	none	ChineseSpring	22
spike	reproductive	none	ChineseSpring	13
roots	vegetative	abiotic	ChineseSpring	3
grain	reproductive	none	ChineseSpring	37
grain	reproductive	none	other	111
leaves/shoots	reproductive	none	other	3
leaves/shoots	vegetative	abiotic	ChineseSpring	3
leaves/shoots	seedling	disease	other	66
leaves/shoots	seedling	none	other	62
spike	reproductive	disease	other	80
spike	reproductive	none	other	76
leaves/shoots	seedling	abiotic	other	28
leaves/shoots	seedling	none	ChineseSpring	3
leaves/shoots	reproductive	none	ChineseSpring	10
roots	seedling	none	ChineseSpring	2
roots	reproductive	none	ChineseSpring	2
spike	vegetative	none	ChineseSpring	2
leaves/shoots	reproductive	disease	other	16
spike	reproductive	disease	ChineseSpring	3
leaves/shoots	vegetative	abiotic	other	3
leaves/shoots	vegetative	none	other	9
leaves/shoots	vegetative	disease	other	33
leaves/shoots	vegetative	none	NullitetraChineseSpring	36
roots	vegetative	none	NullitetraChineseSpring	33

Supplementary Figure 6.1 Expression of wheat NLR-ID alternative splicing candidate genes from Wheat Gene Expression Atlas. The 54 transcripts shown in Figure 3 **were used to generate this heatmap visualization within the Wheat Gene Expression Atlas database. Visualization layout was made based upon expVIP within the database.**

Supplementary Table 6.1 Expression of wheat NLR-ID alternative splicing candidate genes available in Wheat Gene Expression Atlas and NCBI databases. The 54 transcripts shown in Figure 3 **were acquired from the GSE106397 dataset in NCBI and the Wheat Gene Expression Atlas. Cells are colored by expression level, showing differential expression between alternative transcripts of the same gene.**

Supplementary Table 6.1 can be accessed from the following link:

https://figshare.com/s/0bdc5fd44ff6a5e94bcd

ACKNOWLEDGEMENTS

Support for this dissertation work came from the USDA-NIFA hatch Projects to M. Nepal (SD00H469-13 and SD00H659-18), South Dakota Agriculture Experiment Station, and Department of Biology & Microbiology at South Dakota State University.

REFERENCES

1. Jones JD, Dangl JL. The plant immune system. *Nature.* 2006;444(7117):323-329.
2. Jones JD, Vance RE, Dangl JL. Intracellular innate immune surveillance devices in plants and animals. *Science.* 2016;354(6316):aaf6395.
3. Shao ZQ, Xue JY, Wu P, et al. Large-Scale Analyses of Angiosperm Nucleotide-Binding Site-Leucine-Rich Repeat Genes Reveal Three Anciently Diverged Classes with Distinct Evolutionary Patterns. *Plant Physiol.* 2016;170(4):2095-2109.
4. DeYoung BJ, Innes RW. Plant NBS-LRR proteins in pathogen sensing and host defense. *Nature immunology.* 2006;7(12):1243.

5. Hayashi N, Inoue H, Kato T, et al. Durable panicle blast‐resistance gene Pb1 encodes an atypical CC‐NBS‐LRR protein and was generated by acquiring a promoter through local genome duplication. *The Plant Journal.* 2010;64(3):498-510.

6. Kofoed EM, Vance RE. Innate immune recognition of bacterial ligands by NAIPs determines inflammasome specificity. *Nature.* 2011;477(7366):592.

7. Bonardi V, Cherkis K, Nishimura MT, Dangl JL. A new eye on NLR proteins: focused on clarity or diffused by complexity? *Current opinion in immunology.* 2012;24(1):41-50.

8. Bonardi V, Dangl JL. How complex are intracellular immune receptor signaling complexes? *Frontiers in plant science.* 2012;3:237.

9. Cesari S, Bernoux M, Moncuquet P, Kroj T, Dodds PN. A novel conserved mechanism for plant NLR protein pairs: the “integrated decoy” hypothesis. *Frontiers in plant science.* 2014;5.

10. Wu C-H, Krasileva KV, Banfield MJ, Terauchi R, Kamoun S. The “sensor domains” of plant NLR proteins: more than decoys? *Frontiers in plant science.* 2015;6.

11. Sarris PF, Cevik V, Dagdas G, Jones JD, Krasileva KV. Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC biology.* 2016;14(1):8.

12. Takken FL, Goverse A. How to build a pathogen detector: structural basis of NB-LRR function. *Current opinion in plant biology.* 2012;15(4):375-384.

13. Michelmore RW, Christopoulou M, Caldwell KS. Impacts of resistance gene genetics, function, and evolution on a durable future. *Annual review of phytopathology.* 2013;51:291-319.

14. Cui H, Tsuda K, Parker JE. Effector-triggered immunity: from pathogen perception to robust defense. *Annual review of plant biology.* 2015;66:487-511.

15. Walker JE, Saraste M, Runswick MJ, Gay NJ. Distantly related sequences in the alpha‐ and beta‐subunits of ATP synthase, myosin, kinases and other ATP‐requiring enzymes and a common nucleotide binding fold. *The EMBO journal.* 1982;1(8):945-951.

16. Gutterson N, Reuber TL. Regulation of disease resistance pathways by AP2/ERF transcription factors. *Current opinion in plant biology.* 2004;7(4):465-471.

17. Buscaill P, Rivas S. Transcriptional control of plant defence responses. *Current opinion in plant biology.* 2014;20:35-46.

18. Jonak C, Ökrész L, Bögre L, Hirt H. Complexity, cross talk and integration of plant MAP kinase signalling. *Current opinion in plant biology.* 2002;5(5):415-424.

19. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of NBS-LRR–encoding genes in Arabidopsis. *The Plant Cell Online.* 2003;15(4):809-834.

20. Noutoshi Y, Ito T, Seki M, et al. A single amino acid insertion in the WRKY domain of the Arabidopsis TIR‐NBS‐LRR‐WRKY‐type disease resistance protein SLH1 (sensitive to low humidity 1) causes activation of defense responses and hypersensitive cell death. *The Plant Journal.* 2005;43(6):873-888.

21. Lannoo N, Van Damme EJ. Lectin domains at the frontiers of plant defense. 2014.

22. Xiang Y, Song M, Wei Z, et al. A jacalin-related lectin-like gene in wheat is a component of the plant defence system. *Journal of experimental botany.* 2011;62(15):5471-5483.

23. Esch L, Schaffrath U. An update on jacalin-like lectins and their role in plant defense. *International journal of molecular sciences.* 2017;18(7):1592.

24. Hwang IS, Hwang BK. The pepper mannose-binding lectin gene CaMBL1 is required to regulate cell death and defense responses to microbial pathogens. *Plant physiology.* 2011;155(1):447-463.

25. Sarris PF, Duxbury Z, Huh SU, et al. A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell.* 2015;161(5):1089-1100.

26. Le Roux C, Huet G, Jauneau A, et al. A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. *Cell.* 2015;161(5):1074-1088.

27. Date SV. The Rosetta stone method. *Bioinformatics*: Springer; 2008:169-180.

28. Baggs E, Dagdas G, Krasileva K. NLR diversity, helpers and integrated domains: making sense of the NLR IDentity. *Current opinion in plant biology.* 2017;38:59-67.

29. Consortium IWGS. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science.* 2014;345(6194):1251788.

30. Kroj T, Chanclud E, Michel‐Romiti C, Grand X, Morel JB. Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread. *New Phytologist.* 2016;210(2):618-626.

31. Yang S, Tang F, Zhu H. Alternative splicing in plant immunity. *International journal of molecular sciences.* 2014;15(6):10424-10445.

32. Shang X, Cao Y, Ma L. Alternative splicing in plant genes: a means of regulating the environmental fitness of plants. *International journal of molecular sciences.* 2017;18(2):432.

33. Zhang X-C, Gassmann W. RPS4-mediated disease resistance requires the combined presence of RPS4 transcripts with full-length and truncated open reading frames. *The Plant Cell.* 2003;15(10):2333-2342.

34. Tang F, Yang S, Gao M, Zhu H. Alternative splicing is required for RCT1-mediated disease resistance in Medicago truncatula. *Plant molecular biology.* 2013;82(4-5):367-374.

35. Sela H, Spiridon LN, PETRESCU AJ, et al. Ancient diversity of splicing motifs and protein surfaces in the wild emmer wheat (Triticum dicoccoides) LR10 coiled coil (CC) and leucine‐rich repeat (LRR) domains. *Molecular plant pathology.* 2012;13(3):276-287.

36. Saintenac C, Zhang W, Salcedo A, et al. Identification of wheat gene Sr35 that confers resistance to Ug99 stem rust race group. *Science.* 2013;341(6147):783-786.

37. Salcedo A, Rutter W, Wang S, et al. Variation in the AvrSr35 gene determines Sr35 resistance against wheat stem rust race Ug99. *Science.* 2017;358(6370):1604-1606.

38. Borrill P, Ramirez-Gonzalez R, Uauy C. expVIP: a customisable RNA-seq data analysis and visualisation platform opens up gene expression analysis. *Plant Physiology.* 2016:pp. 01667.02015.

39. Halterman DA, Wei F, Wise RP. Powdery mildew-induced Mla mRNAs are alternatively spliced and contain multiple upstream open reading frames. *Plant physiology.* 2003;131(2):558-567.

40. Kersey PJ, Allen JE, Armean I, et al. Ensembl Genomes 2016: more genomes, more complexity. *Nucleic acids research.* 2015;44(D1):D574-D580.

41. Goodstein DM, Shu S, Howson R, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic acids research.* 2011;40(D1):D1178-D1186.

42. Jones P, Binns D, Chang H-Y, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014;30(9):1236-1240.

43. Kearse M, Moir R, Wilson A, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012;28(12):1647-1649.

44. Finn RD, Coggill P, Eberhardt RY, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic acids research.* 2015;44(D1):D279-D285.

45. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics.* 2014;31(8):1296-1297.

CHAPTER 7: NECROTROPHIC FUNGUS *PYRENOPHORA TRITICI-REPENTIS* TRIGGERS EXPRESSION OF MULTIPLE RESISTANCE COMPONENTS IN RESISTANCE AND SUSCEPTIBLE WHEAT CULTIVARS

ABSTRACT

Tan spot (TS) of wheat, caused by the pathogen *Pyrenophora tritici-repentis* (*Ptr*), results in a yield loss by causing chlorosis and necrosis of healthy leaf tissue. Recent advancement in molecular tools and techniques has allowed wheat pathologists to elucidate genetics of *Ptr* infection, revealing three main host-selective toxins produced by *Ptr*, and three susceptibility genes in wheat. The major objective of this study was to compare gene expression in resistant and susceptible wheat cultivars after infection with Ptr ToxA-producing race 2 and direct infiltration with Ptr ToxA proteins. Greenhouse experiments included exposure of susceptible Glenlea and resistant Salamouni wheat cultivars to pathogen inoculum or direct infiltration of leaf tissue of Ptr ToxA protein isolate. Samples from the experiments were utilized for RNA sequencing to infer gene expression among the samples through RNA sequencing. Results showed that, upon *Ptr* contact with wheat tissue, *Ptr* started expressing *ToxA*. Resistant wheat, in response to *Ptr* inoculum, expressed genes associated with plant resistance responses, genes of interest including five chitinases, eight transporters, five pathogen-detecting receptors, and multiple classes of signaling factors. Resistant and susceptible wheat cultivars differed in expression among several groups of genes. Plants exposed to *Ptr* inoculum expressed transcription factors, kinases, receptors, and peroxidases, which are not expressed as highly as in the control samples or samples infiltrated with ToxA. Several of

these genes can be found in tan spot resistance QTLs on chromosomes 1A, 2D, 3B, and 5A. These results show that resistance to *Ptr* is likely due to the expression of many individual genes. Future studies should elucidate the specific roles these genes play in the wheat response to *Ptr*.

INTRODUCTION

Pathogenic fungi negatively impact crop production through qualitative and quantitative reduction in yield, through mycotoxin production and tissue necrosis [1]. Necrotic leaf spotting diseases are especially damaging to wheat crops [1], causing leaf tissue death reducing photosynthetic capacity and carbohydrate production. Wheat supplies a large portion of the calories for human nutrition [2], and fungal pathogens like *Pyrenophora tritici-repentis* (*Ptr*), cause some of the most devastating yield-limiting crop diseases [3]. Tan spot (TS) of wheat is caused by *Ptr* and results in up to 50% yield loss in wheat [4]. Pathogens like *Ptr* have evolved ways to hijack plant resistance signaling pathways, facilitating their infection and reproduction [5]. *Ptr* possesses a unique repertoire of toxins (namely Ptr ToxA, Ptr ToxB and Ptr ToxC) that rely on host susceptibility genes for infection [6]. Ptr ToxA moves into wheat mesophyll cells to disrupt chloroplast activity and Ptr ToxB acts extracellularly [7]. Recognized by brown necrotic lesions surrounded by yellow chlorotic halos as symptoms, *Ptr* overwinters as pseudothecia on wheat residue, dispersal of spores facilitated by wind and periods of prolonged moisture [8]. *Ptr* host-selective toxins (HSTs) cause necrosis and chlorosis in susceptible varieties [9,10]. Variation in *Ptr* races and wheat cultivar susceptibility has been documented [9,10], with findings shown in **Table 7.1** (shown below). Instead of resistance genes (R genes), wheat variation has been attributed to the presence of susceptibility genes present in

many cultivars [6]. Key details regarding the molecular interactions between HST and wheat proteins have emerged in recent years [6,11-14], stimulating the further elucidation of this host-pathogen interaction.

Table 7.1 Toxin production and sensitivity information for selected *Ptr* isolates and wheat cultivars, respectively (S=Sensitive, I=Insensitive).

Ptr Race	Ptr Effectors	Reference	Wheat Cultivar	ToxA	ToxB	ToxC	Reference
1	A, C	[15]	Glenlea	S	I	I	[9]
2	A		Salamouni	I	I	I	
3	C		Katepwa	S	S	I	
4	-		6B365	I	I	S	
5	B	[16]	ND495	S	I	I	
6	B, C	[17]	6B662	I	S	I	[10]
7	A, B	[18]					
8	A, B, C						

Host-selective toxins (HSTs) are fungal effectors that facilitate infection and access to nutrients. Many HSTs have been identified in species of the *Alternaria* and *Cochliobolus* genera, with additional HSTs likely to be found in future research [19]. To facilitate infection and nutrient extraction, *Ptr* produces three HSTs: Ptr ToxA, Ptr ToxB, and Ptr ToxC (see Faris et al. 2013 for review [6]). Ptr ToxA causes necrosis in wheat tissue and evidence supports the hypothesis that the Ptr *ToxA* gene was horizontally transferred from *Parastagonspora nodorum*, another pathogenic fungus of wheat that causes *Septoria/Stagonospora nodorum* blotch [20]. Ptr ToxA has been shown to move into wheat mesophyll cells, possibly as a homodimer, where it localizes to chloroplasts and binds to the ToxA-binding protein-1 (ToxABP1) and a plastocyanin (PCN), leading to chloroplast disruption, production of reactive oxygen species (ROS), and cell death [13,21,22]. Ptr ToxA contains an RGD-motif that is involved in target-binding and movement

into wheat mesophyll cells, likely by receptor-mediated endocytosis[23]. In contrast to Ptr ToxA, Ptr ToxB and Ptr ToxC are both native to *Ptr* and associated with tissue chlorosis [24]. Ptr ToxB acts extracellularly [7], likely through a ligand-receptor interaction that triggers plant signaling and leads to chlorosis. Ptr ToxC is a small, nonpolar, non-proteinaceous secondary metabolite [25] that has yet to be characterized. Both Ptr ToxA and Ptr ToxB lead to interference with photosynthesis and ROS accumulation, indicating similar mechanisms of action even though they differ in structure [14]. Several *Ptr* races have been identified to possess various combinations of the three toxins, and various wheat cultivars have shown sensitivity to individual toxin isolates. Integrated disease management of tan spot may include the use of resistant varieties, fungicide applications, crop rotation, and residue removal to limit overwintering of inoculum [8]. Fungicide applications and crop rotations, while effective, are expensive and inconvenient for farmers. Determination of how tan spot causes disease symptoms, followed by deployment of cultivars with durable resistance will provide an economical solution to yield loss [4].

The main objective of this study was to compare gene expression of resistant and susceptible wheat cultivars after infection with Ptr ToxA-producing race 2 and direct infiltration with Ptr ToxA. Identification of genes triggered by *Ptr* HSTs will provide candidates for assessing why some wheat cultivars resist TS symptoms and others develop necrotic lesions. These results will impact the development of tan spot-resistant wheat by informing breeding decisions that ultimately influence farmer variety selection. This will allow us to assess the molecular mechanism that involves *Ptr* effectors in wheat cells, and understand which signaling mechanism(s) may lead to the disease. In-depth

understanding of the molecular basis of the wheat-*Ptr* pathosystem will be useful in the development of wheat cultivars that possess durable resistance to *Ptr* and other necrotic leaf spotting diseases caused by similar pathogens, such as *Parastagonspora nodorum*.

MATERIALS AND METHODS

Two cultivars of wheat were grown, TS resistant Salamouni and TS susceptible Glenlea, selected due to their well-established insensitivity and sensitivity to Ptr ToxA, respectively. Seeds were germinated in petri dishes on damp filter paper and transferred to 3 x 9 cm plastic containers (Stuewe & Sons Inc., Tangent, OR, USA) filled with Sunshine Mix 1 soil (Sun Gro Horticulture, Agawam, MA, USA). Plants were uniformly watered daily with greenhouse conditions of 16 hour light and 8 hour dark cycles at 22°C for 14 days prior to inoculation/infiltration. Four treatments were used: spray *Ptr* race 2 inoculum, spray water (control), Ptr ToxA injection, and water injection (control). Inoculum was prepared following methods described in Abdullah et al. (2017) [26]. A culture of *Ptr* Race 2 isolate 86-124 was initiated using an agar plug and V8-PDA medium plates (agar = 10 g; potato dextrose agar (PDA) = 10 g; $CaCO_3$ = 3 g; V8 Juice = 150 mL and 850 mL distilled water) [27]. Plates were incubated in darkness for 5 days and colonies were flooded with distilled water and disrupted. Plates were then incubated for 24 hours at 22C in light and then for 24 hours at 16C to cause production of conidia. The spore suspension at 3000 spores/mL was used to inoculate plants with a Preval CO_2 pressurized sprayer (www.preval.com) [28,29]. A sprayer containing only sterile water was used as a control. Toxin infiltration was accomplished using a 10 µg/mL solution of Ptr ToxA, kindly provided by Dr. Timothy Friesen (USDA-ARS, North Dakota State

University), and directly infiltrating leaf tissue using a needle-less syringe [30], similarly using sterile water infiltration as a control. Inoculated plants were then placed in an automated humidity chamber at 100% humidity (misting 16 seconds/10 minutes). Samples were collected at three timepoints: 0 hr, 8 hr, and 16 hr. Liquid nitrogen was used to freeze the samples, stored at -80°C until RNA extraction could be completed using the Ambion PureLink RNA extraction kit with Trizol reagent. RNA was then purified using DNase to remove any DNA contamination and sequenced at Iowa State University using Illumina HiSEQ 3000 set at 100 base pairs with single reads.

Resulting FASTQ files were checked for quality using FASTQC [31], trimmed using Btrim [32], and mapped to the reference wheat genome using Hisat [33] and Htseq [34]. Genomic data was accessed from the Ensembl Genomes database [35]. The *Pyrenophora tritici-repentis* genome was also accessed from Ensembl Genomes in order to map any pathogen RNA reads that would not map to the wheat genome. The program R [36] with the Bioconductor package and DESeq2 [37] were used to assess differential expression of genes between samples. Genes differentially expressed between cultivars, treatments, or time points were compiled and assessed based on their sequence annotations and homology in Gene Ontology (GO), InterProScan [38], and BLAST [39] , used to predict the general functions and roles in stress response, metabolism, development, and various other cellular processes. The iDEP program was also used to assess differentially expressed genes in wheat tissue [40].

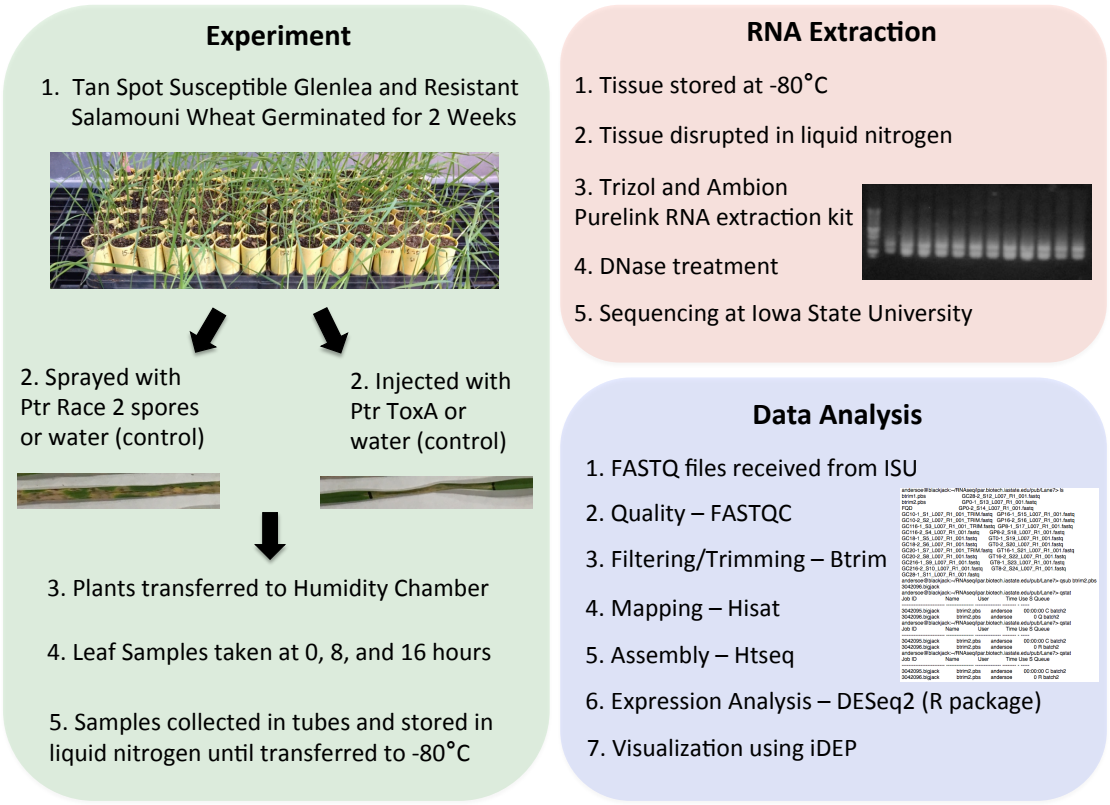


Figure 7.1 Workflow diagram of methods for greenhouse RNA sequencing experiment. Experiment was conducted at South Dakota State University Young Brothers Seed Technology Laboratory.

RESULTS

Pathogen Expression

ToxA encoding RNA sequences, gathered using the workflow in **Figure 7.1** mapped to the *Ptr* genome, showed that *Ptr* genes were found in the expression data in inoculated samples confirming that *Ptr* produces ToxA in wheat tissue upon the pathogen exposure (**Figure 7.2**). RNA sequencing data files are available from NCBI

(https://dataview.ncbi.nlm.nih.gov/object/PRJNA529906?reviewer=3qsuo5i095ic5v1k7fupmmmsbj) and are listed in **Supplementary Table 7.1**. Many of the other transcripts found from *Ptr* do not have a direct role in pathogenesis and are likely involved general cellular function. ToxA was not expressed in 0 hr samples, with the highest level of expression in a 0 hr sample being 3 counts. The 8 and 16 hr samples showed Tox A levels between 65 and 356 counts per sample. This indicates that *Ptr* does not express Tox A until it comes in contact with wheat tissue. No *Ptr* genes were significantly differentially expressed between the two cultivars, indicating that the difference in disease response resides mainly in wheat's response to *Ptr* and not *Ptr*'s response to different cultivars of wheat. ToxA expression was found to be higher in Salamouni samples than in Glenlea. Other *Ptr* genes may possess involvement in pathogenesis. Elongation factor is likely initiated once contact with wheat is sensed, along with peroxisomal enzymes. Proteins related to damage, such as heat shock proteins, could be expressed in response to enzymes wheat uses to defend itself during infection. Reads mapped to the wheat genome resulted in many differentially expressed genes, in wheat, as visualized in the heat map (**Figure 7.3**) where many genes either increased or decreased in expression, shown in red and green, respectively. The Weighted Correlation Network Analysis (WGCNA) and Principal Component Analysis (PCA) of the data (**Figure 7.4**) show that several groups of differentially expressed genes associate with particular functions and expression differs between the two cultivars. Wheat samples at 8 and 16 hours expressed many genes associated with response to stress and biosynthetic processes not expressed at 0 hours.

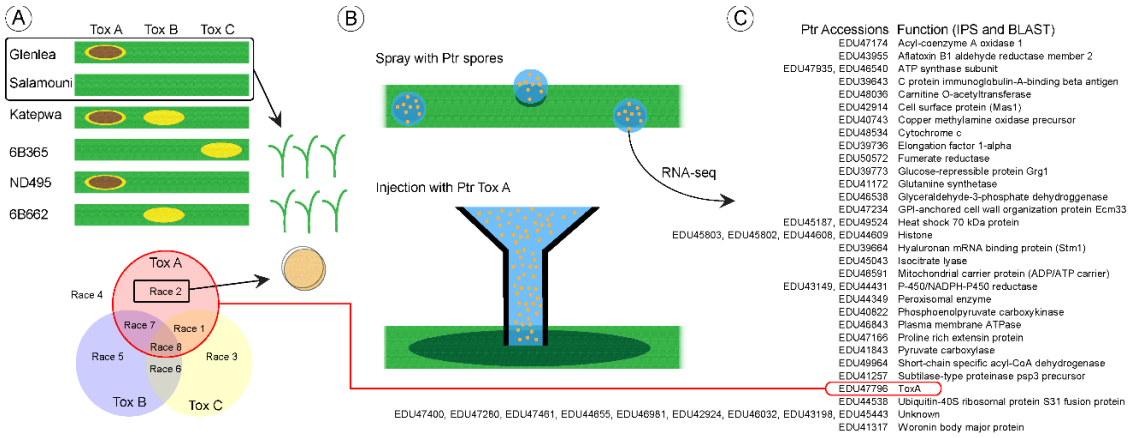


Figure 7.2 Experimental design to show that wheat *Ptr* genes are expressed upon *Ptr* exposure. (A) Wheat cultivars and *Ptr* races possess various susceptibility genes and effectors, respectively. *Ptr* Race 2 resistant and susceptible cultivars were selected to investigate the effect of Ptr ToxA on wheat tissue. (B) Samples were inoculated with *Ptr* Race 2 or infiltrated with ToxA. (C) *Ptr* genes expressed in inoculation samples included ToxA, a major cause of necrosis, showing that *Ptr* was expressing *ToxA* after inoculation.

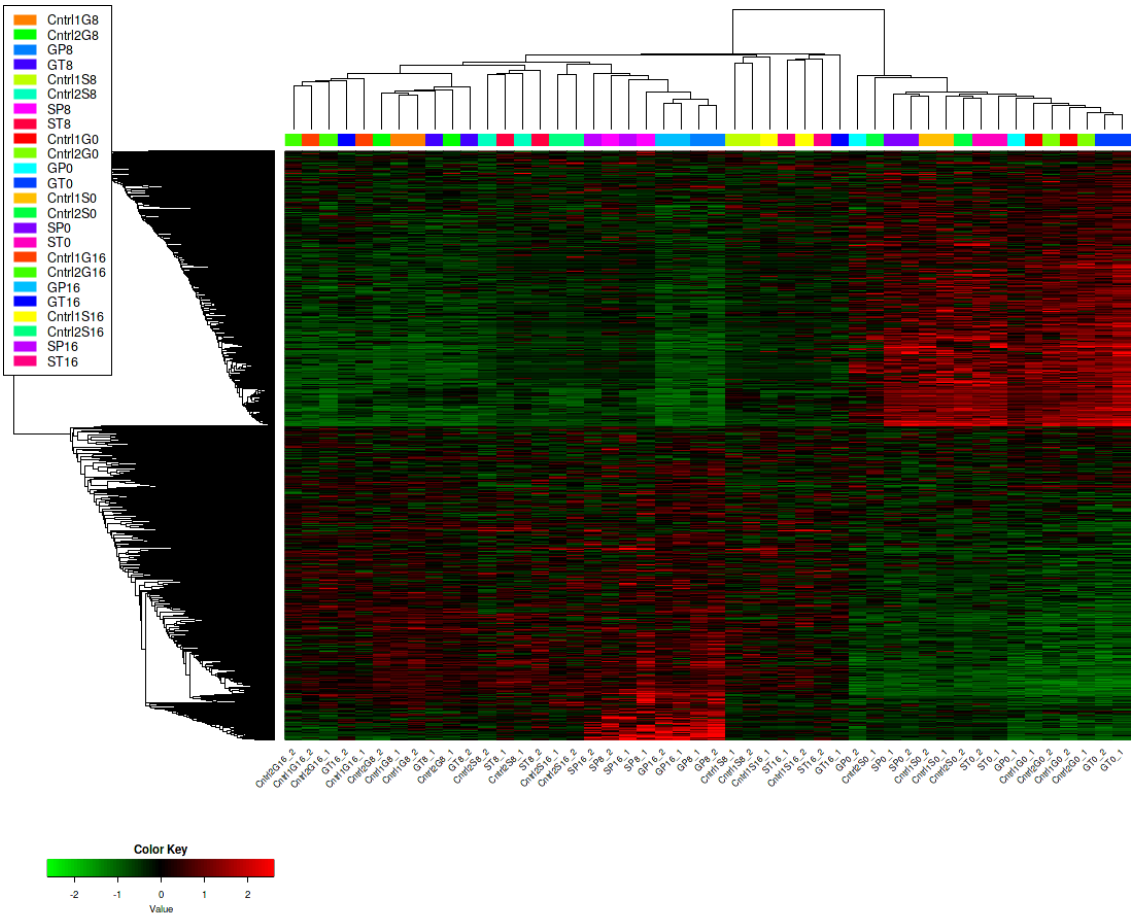
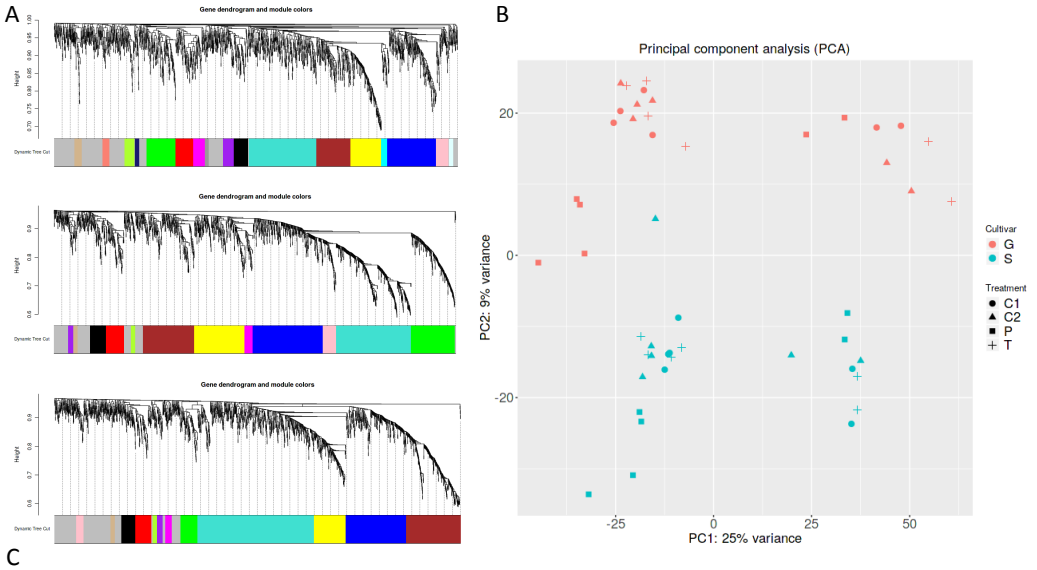


Figure 7.3 Heatmap visualization of expression data was constructed using integrated Differential Expression and Pathway analysis (iDEP). Sample names include cultivar (G and S for Glenlea and Salamouni, respectively), treatment (P, T, Cntrl1, Cntrl2 for Pathogen inoculation, Toxin infiltration, Inoculation Control, and Infiltration Control, respectively), and timepoint (0 hrs, 8 hrs, and 16 hrs). Samples that increased or decreased in expression were colored red and green, respectively, with little change shown in black.



C

0 hrs			8 hrs			16 hrs		
adj. P-value	Genes	Pathways	adj. P-value	Genes	Pathways	adj. P-value	Genes	Pathways
2.40E-12	18	Photosynthesis	6.20E-12	34	Carboxylic acid metabolic process	1.80E-08	18	Photosynthesis
2.70E-11	25	Generation of precursor metabolites and energy	1.90E-11	42	Small molecule metabolic process	9.40E-08	27	Generation of precursor metabolites and energy
3.40E-11	13	Photosynthesis, light reaction	3.80E-11	30	Drug metabolic process	2.20E-07	47	Organonitrogen compound biosynthetic process
1.40E-07	8	Photosynthesis, light harvesting	5.80E-11	13	Aromatic amino acid family metabolic process	9.00E-07	44	Small molecule metabolic process
1.10E-06	5	Protein repair	6.50E-10	17	Alpha-amino acid metabolic process	3.90E-06	42	Response to stress
9.00E-05	15	Response to abiotic stimulus	4.30E-09	24	Small molecule biosynthetic process	8.50E-06	31	Organic acid metabolic process
8.40E-04	3	Photosystem II repair	5.30E-09	21	Cellular amino acid metabolic process	5.10E-05	29	Carboxylic acid metabolic process
8.40E-04	5	Protein-chromophore linkage	8.70E-09	21	Organic acid biosynthetic process	5.70E-05	20	Response to abiotic stimulus
8.40E-04	2	Aldonate transmembrane transport	3.10E-08	32	Catabolic process	6.90E-05	16	Electron transport chain
8.40E-04	6	Chloroplast organization	4.20E-08	23	Cofactor metabolic process	1.10E-04	9	Photosynthesis, light reaction
1.00E-03	3	Response to red light	5.30E-08	18	Drug catabolic process	2.90E-04	15	Cellular carbohydrate metabolic process
1.00E-03	8	Response to light stimulus	1.70E-07	34	Response to stress	3.10E-04	18	Cellular amino acid metabolic process
						3.10E-04	11	Purine nucleoside monophosphate metabolic process

Figure 7.4 Analysis of expression data, including (A) Weighted Correlation Network Analysis (WGCNA) of the three timepoints (0, 8, and 16 hours); (B) principal component analysis comparing the two cultivars; and (C) Functional pathways in which genes from 0, 8, and 16 hours were associated as determined from the WGCNA.

Resistant and Susceptible Cultivar-Specific Expression

Many genes were differentially expressed between Glenlea and Salamouni wheat, with many possessing potential involvement in defense response (**Figure 7.5**). Several of these genes that were expressed highly in either Glenlea or Salamouni were expressed not at all or only slightly in the other. For example, TraesCS5B02G417500.1 encodes an NB-

ARC-containing protein and hundreds of reads from each of the Salamouni samples mapped to this gene, whereas not a single read was found in all Glenlea samples. For genes that are not expressed at all in one of the two cultivars, questions can be raised about if that cultivar even contains a functional version of that gene. Other genes were only associated with a single read in one cultivar. This could be a false positive. Groups of genes that are differentially expressed between the two cultivars fit into two categories: 1) groups that are mainly expressed in one cultivar, and 2) groups where approximately half of the genes are expressed in one cultivar and half expressed in the other.

Glenlea and Salamouni do not show the same expression of chitinase/chitin-binding, major sperm protein, MLO, permease, ubiquitin, WAK, wound-induced, and xylanase inhibitors. Each of these, with the exception of major sperm protein, were expressed almost exclusively in Salamouni. Salamouni also expressed a homolog of Snn1, which is a SnTox1 sensitivity gene that contains the domains of a calcium-binding and galacturonan-binding wall-associated kinase. While many genes were only expressed in Glenlea, groups of genes with similar domains were not exclusively expressed in Glenlea like they were in Salamouni. Glenlea and Salamouni express different NLRs, kinases, LRRs, LRR-Kinases, cytochrome P450 enzymes, F-box proteins, Lectins, Peptidases, Peptidase inhibitors, Peroxidases, UDP-glucosyltransferases, Thaumatins, and Transcription Factors. For each of these groups of genes, Glenlea or Salamouni express all or the vast majority of each gene, indicating that breeders should target those only expressed in Salamouni for breeding resistance into wheat lines. These are the categories that possess functions associated with pathogen resistance, not including the

many other genes differentially expressed between the cultivars that do not have possible resistance functions.

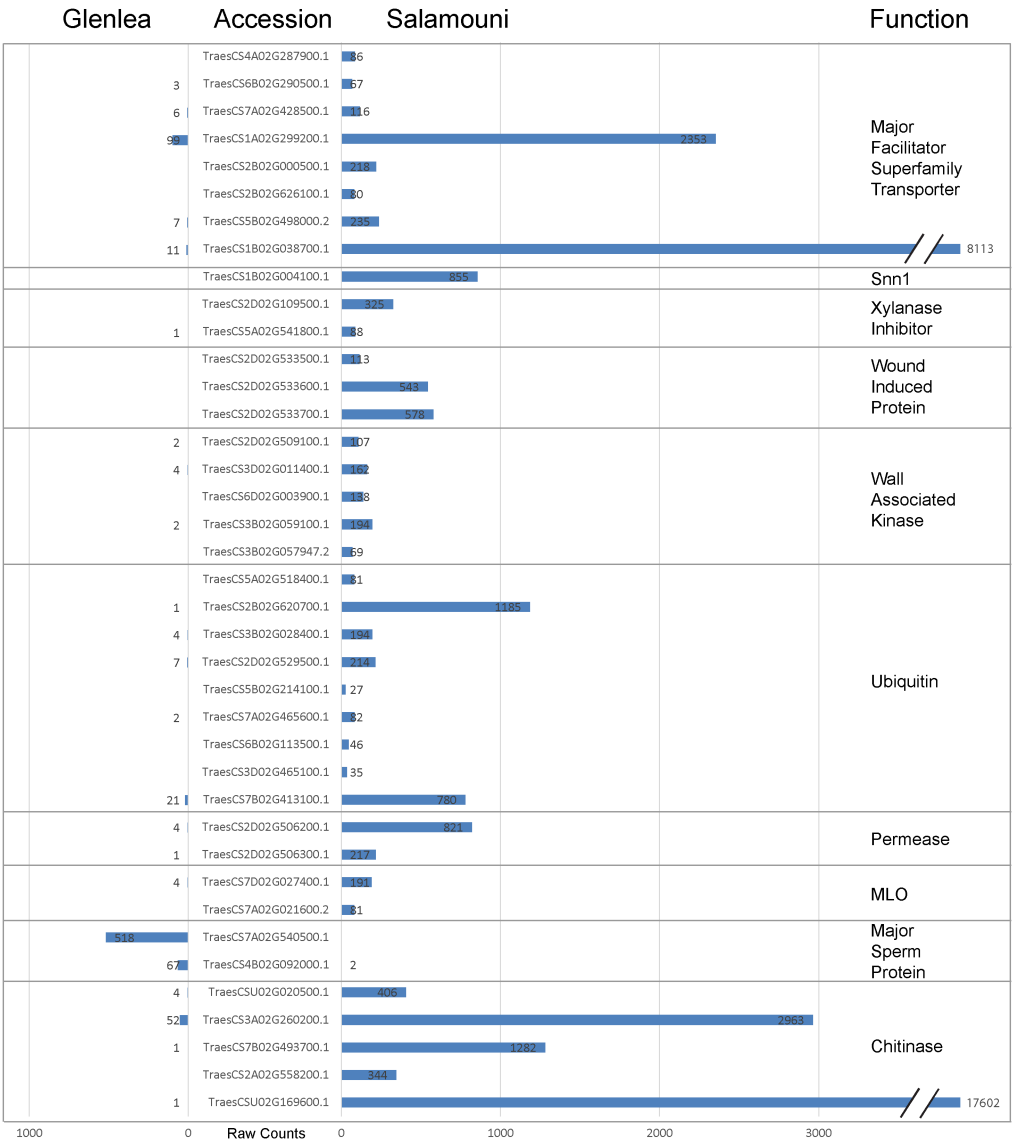


Figure 7.5 Gene expression distinction between Glenlea and Salamouni cutivars. Several genes were only expressed in Glenlea or Salamouni, with a few more showing only one read in one of the two cultivars.

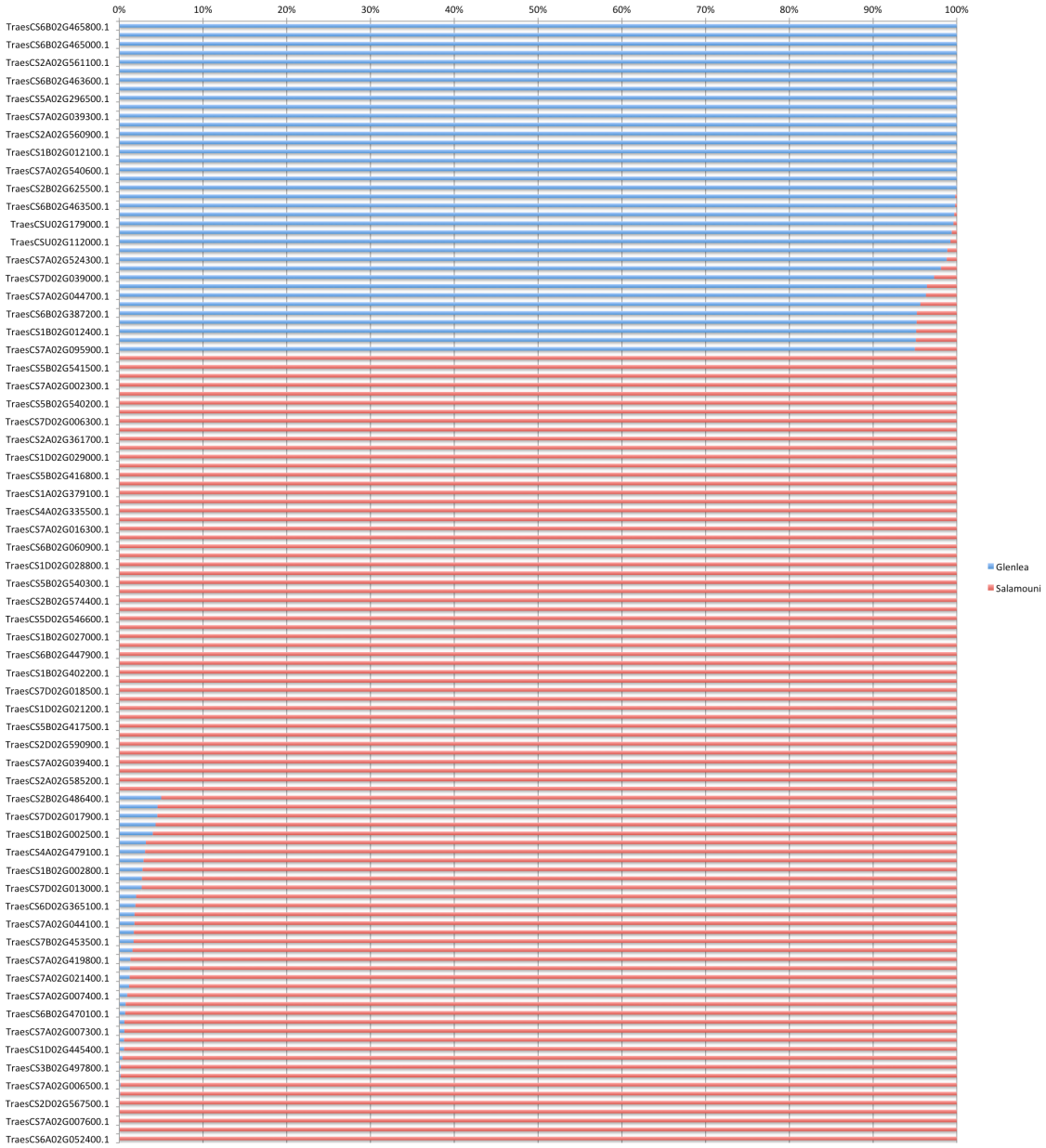


Figure 7.6 NLR-encoding genes that were significantly differentially expressed between cultivars. NLR accession names are listed on the left and the proportion of reads found in either of the two cultivars are shown to the right of each name. Glenlea and Salamouni expression data are indicated by blue and red lines, respectively.

Of the significantly differentially expressed genes, 125 NLRs were expressed mostly or entirely in one of the two cultivars (**Figure 7.6**). Unlike genes listed in **Figure 7.5,** where all or most of the genes of one type were expressed by one cultivar, NLRs were divided in expression between the two cultivars, with many expressed exclusively in one cultivar. Salamouni did show expression of more NLRs than Glenlea. However, the number of NLR-encoding genes expressed does not necessarily correlate with resistance, especially since it is not currently possible to associate individual NLRs with particular pathogens on a large scale.

Pathogen Inoculation and Toxin Infiltration Expression

Many groups of genes were expressed in both cultivars, but differed greatly between control samples and those treated with *Ptr* inoculum or ToxA (see **Figure 7.7** gene groups). While this figure also highlights some differences between the two cultivars, the primary focus of this data was the differences between treatments with similar patterns in both cultivars. Many genes expressed much more in samples treated with *Ptr* inoculum that were not highly expressed in controls or ToxA infiltrated samples (**Figure 7.7**). These included osmotin/thaumatin-like genes (e.g. TraesCS7B02G483400.1, TraesCS7D02G551400.1, and TraesCS7A02G558500.1) and cell wall synthesis genes such as TraesCS2B02G040600.1, TraesCS2B02G040500.1, both of which are associated with disease resistance. Other genes (see **Figure 7.7**) were expressed minimally or not at all in 0 hour samples and controls at 8 hrs, but highly expressed in 8hr pathogen or toxin exposure (**Figure 7.7**). These genes are likely expressed in response to Ptr or ToxA. Several genes (e.g. TraesCS2B02G553700.1, TraesCS6B02G170000.1, and TraesCS7D02G058600.1) showed significant differential

expression between treatments and cultivars, but did not share homology with any characterized genes, nor did they possess InterProScan domains. These genes are likely to play a role in pathogenesis, therefore their functional characterization is warranted and many were found in locations associated with QTLs that likely play a role in TS resistance (**Figure 7.8**).

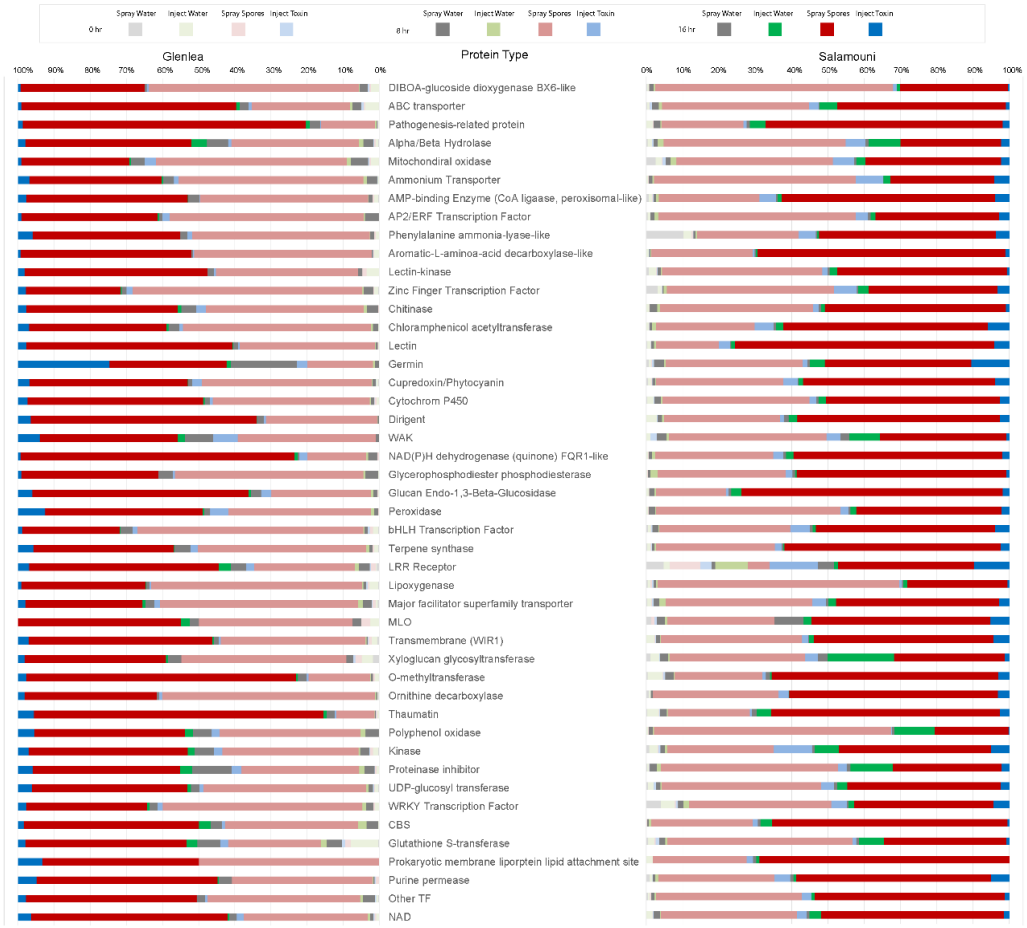


Figure 7.7 Genes expressed highly when exposed to *Ptr* or ToxA, but minimally in 0hr samples and other control samples. *Ptr* triggered expression of several genes that were minimally expressed at 0 hrs, regardless of treatment. Grey, green, red, and blue indicate control inoculation, control infiltration, *Ptr* inoculation, and Ptr ToxA infiltration, respectively, as shown in the legend at the top of the figure.

Domains	1A	2D	3B	5A
Alpha/Beta hydrolase	2	6	2	1
AP2/ERF Transcription Factor	1	2	2	-
bHLH Transcription Factor	-	2	-	1
Cullin repeat/Exocyst complex Exo70	-	4	-	-
Cytochrome P450	1	5	2	1
EGF calcium-binding/Kinase	1	2	-	-
F-box/LRR	5	-	-	-
Haem peroxidase	-	7	2	-
Jacalin-like Lectin	-	-	1	-
Kinase	3	14	2	5
Lectin/Kinase	4	6	1	-
LRR	4	1	4	1
LRR-Kinase	6	6	-	-
Major facilitator superfamily (transporter)	-	2	-	2
NAC Transcription Factor	-	4	-	-
NB-ARC/CC	3	2	3	-
NLR	8	5	1	-
Thaumatin-like	-	-	-	1
Thioredoxin	-	3	-	-
UDP-glucosyltransferase	2	-	9	-
WRKY Transcription Factor	-	1	2	-
Zinc Finger	6	10	5	5

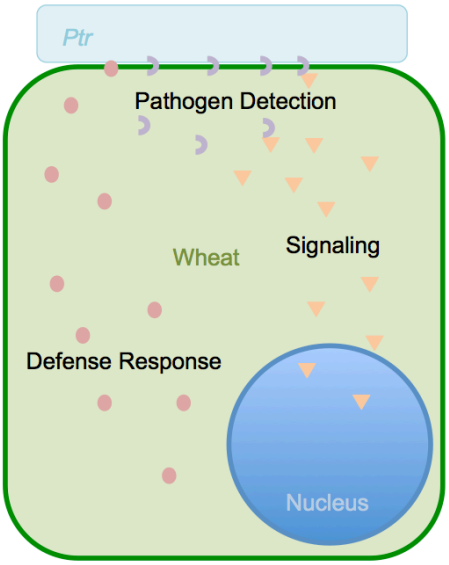


Figure 7.8 Differentially expressed genes located within QTLs associated with resistance to tan spot, as described in Kariyawasam et al. (2016), and their proposed biological function. Functions are grouped by the following categories: pathogen detection (purple arcs), signaling (orange triangles), and defense response (red ovals).

DISCUSSION

Ptr Triggers Expression of Resistance Components in Wheat

The data indicates that *Ptr* Race 2 produces Ptr ToxA after contact with wheat tissue (**Figure 7.2**). Even though Race 2 does not produce ToxB and ToxC, other factors likely elicit response to *Ptr*, such as chitin and other Pathogen-Associated Molecular Patterns (PAMPs) [41]. PAMPs and catalysis of the plant cell wall by fungal enzymes trigger the activation of Pattern-Recognition Receptors (PRRs) and Wall-Associated Kinases (WAKs), respectively [42]. PRRs generally possess Leucine-Rich Repeat (LRR)

and Kinase domains, with an extracellular LRR initiating intracellular kinase activity [42]. WAKs possess galacturonan-binding domains and kinase domains, with kinase activity initiated by cell wall fragments. Several kinases, including LRR-kinases and WAKs, were expressed more highly in plants sprayed with inoculum or injected with the toxin. Kinases then phosphorylate additional signaling molecules, eventually triggering transcription factor activation. Several transcription factors were expressed more highly when exposed to the pathogen. Expression of transcription factors of the AP2/ERF, bHLH, and WRKY classes were triggered by *Ptr*. While further functional characterization is required to know how these kinases and transcription factors are directly involved in response to *Ptr*, this analysis highlights which signaling factors may be involved. These kinases and transcription factors would not necessarily be the same signaling factors initiated by receptor detection of *Ptr*, since RNA-seq only measures changes in expression. These would be the kinases and transcription factors that are expressed as the result of initial receptor/signaling activity. Therefore, the initial detection is not captured in this experiment, only the genes expressed as a result of the initial detection of *Ptr*. As a direct result of pathogen detection, the first line of defense detectable through RNA-seq would be the defense response genes expressed more highly in the hours following inoculation. The heatmap (**Figure 7.3**), shows that many genes increased or decreased in expression through the course of the experiment with many genes differing between the two cultivars, as shown in the principal component analysis (**Figure 7.4**). The WGCNA (Figure 7.4) shows that many genes expressed at 8 and 16 hours may play roles in defense, particularly those in the pathways associated with stress

response (34 and 42 genes at 8 and 16 hours, respectively), small molecule metabolic process (42 and 44 genes), and small molecule biosynthesis process (24 at 8 hours).

Plants respond to fungal pathogens by producing several types of proteins. Some response proteins possess more definite functions in response to *Ptr*, such as chitinases, thaumatins, terpene synthases, purine permeases, and peroxidases. Each of these possesses likely functions in the breakdown of pathogen components or the production of compounds that deter pathogen growth and development. Some proteins break down components of fungal cell walls. Chitinases, for example break down chitin in fungal cell walls and inhibit fungal development and reproduction. Chitinases were expressed more by Salamouni than by Glenlea (**Figure 7.5**). Given that *Ptr* is a fungal pathogen, this production of chitinase may play a significant role in the resistance Salamouni possesses. WAK-encoding genes were also found to be expressed more in Salamouni, possibly initiating defense responses when damage to the cell wall takes place. Pandelova et al. (2009) also found differential expression of several signaling genes (i.e. MAPKs and transcription factors) along with response protein-encoding genes like chitinase [43]. NLRs, commonly described as major resistance factors, were largely or entirely either expressed in one cultivar or the other, with the majority of them expressed in Salamouni (**Figure 7.6**). This data does not necessarily indicate that NLRs not expressed in one cultivar are not in that cultivar's genome. However, this data does indicate that cultivars do not equally express NLRs. As with signaling factors, these are not necessarily the same receptors that initiated resistance mechanisms, since those receptors cannot be measured through RNA-seq.

Thuamatin-like proteins are permatins that make the fungal membranes more permeable, disrupting cellular activity. These defensive proteins were among the many types that expressed more highly in plants inoculated with *Ptr* spores (**Figure 7.7**). Other enzymes do not directly act against fungal cells, but catalyze the production of chemicals that inhibit fungal development. Benzoxazinoids, such as DIMBOA, are produced as a defense against pests and are synthesized through the activity of several BX enzymes. BX1 catalyzes the production of indole from indole-3-glycerolphosphate. Cytochrome P450 enzymes Bx2-Bx5 convert indole to DIBOA. BX6 catalyzes DIBOA-glucoside into TRIBOA-glucoside, which then is acted upon by BX7, converting it to DIMBOA-glucoside, the main benzoxazinoid for wheat [44]. UDP-glucosyltrasferases like Bx8 and Bx9 reduce autotoxicity of benzoxazinoids. This data indicates that several Cytochome P450, DIBOA-glucoside dioxygenase (BX6)-like, and UDP-glucosyltransferase proteins were more heavily expressed in race 2-inoculated samples. Since DIMBOA is known to possess antifungal properties, up-regulation of enzymes in its synthetic pathway fit expectations. Indole-3-glycerolphosphate can be converted into either DIMBOA-glucoside or tryptophan, with DIMBOA-glucoside then being coverted to 6-methoxy-2-benzoxazolinone (MBOA) and tryptophan being converted into serotonin. Fall and Solomon (2013) showed that in response to SnToxA from *Stagonospora nodorum*, wheat produces secondary metabolites serotonin and MBOA to inhibit *S. nodorum* sporulation and disrupt metabolism, respectively [45]. Therefore, the up-regulation of DIMBOA pathway enzymes may be to use serotonin and MBOA to inhibit *Ptr*. Fall and Solomon suggest that *S. nodorum* uses other effectors to suppress the response triggered by SnToxA.

Response proteins may work together in other ways to defend wheat from *Ptr* infection. Four groups of genes uniquely expressed in race 2-inoculated samples share a common defense pathway: UDP-glucosyltransferases, cytochrome P450s, beta-glucosidases, and alpha/beta hydrolases. These proteins lead to the production of hydrogen cyanide as a defense mechanism in *Sorghum bicolor*. The cyanogenic glucoside dhurrin is synthesized by UDP-glucosyltransferase and cytochrome P450. Dhurrinase, a beta-glucosidase, releases p-hydroxymandelonitrile from dhurrin, which is acted upon by p-(S)-hydroxymandelonitrile lyase, an alpha/beta hydrolase, to make hydrogen cyanide, as reviewed in Ordonio et al. (2016) [46]. Like *S. bicolor*, wheat also produces cyanogenic glucosides [47]. Genes encoding all four of these proteins were expressed more highly in samples inoculated with *Ptr* than other samples.

Many of the genes expressed more highly when wheat was exposed to *Ptr* inoculum possess additional functions in resistance systems. Pyridoxal phosphate-dependent decarboxylases (aromatic-L-amino-acid decarboxylases) were differentially expressed and play roles in GABA signaling. Wang et al. have proposed a self-defense model in *Fusarium graminearum*, where MFS transporters export DON and GABA promotes production of DON and self-defense [48]. Other response proteins counter *Ptr* proteins, such as proteinase inhibitors, which can be expected, since pathogens produce several enzymes that degrade host barriers to infection. Polyphenol oxidases are defense compounds known to reduce the digestibility of plant tissue. Phenylalanine ammonia-lyase (PAL) is another common defense protein plants use. Wheat dirigents have been shown to increase pathogen resistance and lignin biosynthesis [49], which would be used to made infection more difficult. Ornithine decarboxylase catalyzes the conversion of

ornithine to putrescine, which can then be converted to spermidine and spermine, a response to biotic stress caused by a broad range of pathogens [50]. Lipoxygenases aid in the synthesis of signaling molecules, such as jasmonates, which are important signaling factors in defense responses [51]. Signaling and response genes that were found to be mostly expressed in the samples inoculated with *Ptr* race 2 indicate that *Ptr* is initiating responses based upon receptor activity. Several lectins, lectin-kinases, kinases, and LRR receptors were mainly expressed in Race 2-inoculated samples. These proteins likely work together to transmit the signal that the fungal pathogen is present. The expression of several transcription factors commonly associated with biotic stress resistance was triggered by inoculation by *Ptr*.

Resistance Components Exist Within Identified TS Resistance QTLs

Additional research targeting tan spot has associated *Ptr* resistance with various QTLs throughout the wheat genome. Kariyawasam et al. (2016) identified QTLs involved in tan spot resistance [52]. While cultivar Penawawa is resistant to several races of *Ptr*, application of ToxA causes necrosis, the same reaction found in the susceptible cultivar Louise. One major QTL associated with non-race-specific resistance exists on the long arm of chromosome 3B, called *QTs.zhl-3B*, which may be the same as *QTs.fcu-3BL* [53], similar in location to other genes associated with tan spot resistance (*tsr2* and *tsr5*). Additional QTLs were found on chromosomes 1A, 2D, and 5A, the latter conferring resistance to all tested races of *Ptr* [52]. Many differentially expressed genes have locations within these QTLs (**Figure 7.8**). Several of these genes possess possible roles in

resistance, either as receptors (NLRs, other LRR-containing receptors), signal transduction factors (kinases, transcription factors), or response proteins (peroxidases, hydrolases). The genes found within these QTLs represent the same types of genes discussed in the previous sections, those that differ in expression between cultivars and inoculation treatments.

Ptr isolate 86-124 penetrates wheat tissue within 3 hours after inoculation and continues infection in the hours following [54], indicating that the 8 hour time point provides a view of expression after *Ptr* has had a chance to initiate the expression of defense genes. The humidity chamber provided the moisture required for *Ptr* infection. *Ptr* uses an appressorium and penetration peg to access wheat epidermal cells [55]. Both resistant and susceptible host plants experience penetration [55], explaining why many similarities may exist between the two cultivars in expression. If *Ptr* was not able to penetrate the resistant cultivar, more diverse expression may have resulted. Since 8 hours after inoculation shows wheat responding to *Ptr* through the production of defense proteins and chemicals, it is likely that some of the genes expressed by *Ptr* at this point are also in response to wheat. Gene expression in a pathogen changes in response to the host just as the host changes gene expression after the pathogen has been detected. Some of the stress-response genes expressed by Ptr, such as those listed in **Figure 7.2**, are likely involved in responding to wheat defense mechanisms.

Ptr Pathogenicity Likely Involves Multiple Factors with Unique Evolutionary Histories

Wheat lines with the specific susceptibility genes *Tsn1*, *Tsc2*, and *Tsc1* are vulnerable to Ptr ToxA, Ptr ToxB, and Ptr ToxC, respectively [6]. Wheat lines with the mutant alleles *tsn1*, *tsc2*, and *tsc1* are not sensitive to the respective toxins [25,56,57], indicating that susceptibility genes play a major role in tan spot infection. Liu et al. (2017) used recombinant inbred lines (RILs) with various combinations of homozygous dominant, heterozygous, and homozygous recessive genotypes of *Tsn1* and *Tsc1*, to show that the plants that lack the *tsn1* and *tsc1* alleles experienced toxin-induced necrosis and chlorosis, respectively [12]. It is not completely understood how these susceptibility genes cause wheat lines to be sensitive to particular toxins or effectors. Similar to the *Cochliobolus victoriae* susceptibility gene *LOV1* in *Arabidopsis thaliana* [58], *Tsn1* encodes a cytoplasmic NBS-LRR protein, belonging to a large family of R-proteins that generally function as cytoplasmic receptors for the detection of pathogenic effectors. The recently available nucleotide and protein sequences of *Tsn1* and Tsn1, respectively, suggest that *Tsn1* is the result of a gene-fusion that took place in the progenitor of wheat's B-genome, a close relative of *Aegilops speltoides* [11]. Early analysis of tan spot described the *Ptr* toxins as necessary for the development of disease symptoms, however, a recent study has confirmed that even non-toxin-producing *Ptr* races still cause necrosis and chlorosis [59]. Wheat cultivars may even show necrosis when inoculated with *Ptr* races that only produce Ptr ToxB [10]. Taken together, these details show that both *Ptr* virulence and wheat resistance are multifaceted.

CONCLUSIONS

Pyrenophora tritici-repentis (*Ptr*) provides a complex example of a pathogen that has evolved the ability to hijack resistance components to cause infection. The results showed that, upon *Ptr* contact with wheat tissue, *Ptr* started expressing ToxA. Wheat, in response to *Ptr* inoculum, expresses many genes associated with plant resistance responses, including chitinases, transporters, pathogen-detecting receptors, and signaling factors. Resistant and susceptible wheat cultivars differed in expression between several groups of genes and plants exposed to *Ptr* inoculum expressed several groups of genes not expressed as highly in control samples, such as transcription factors, kinases, receptors, and peroxidases. Several of these genes can be found in tan spot resistance QTLs on chromosomes 1A, 2D, 3B, and 5A. These results show that resistance to *Ptr* is likely the result of many individual genes being expressed. Elucidating genes involved in resistance to *Ptr* will provide researchers and breeders specific target factors for developing cultivars that possess these components, with the goal of establishing durable *Ptr* resistance. Gaining a more complete molecular picture of *Ptr* infection may also elucidate why *Ptr* has become a major disease of wheat in recent years.

SUPPLEMENTARY DATA

Supplementary Table 7.1 FASTQ files generated from RNA sequencing of samples are available from the Sequence Read Archive (SRA) data repository of NCBI (Submission: SUB5368694, BioProject: PRJNA529906, Data Archive: BioSample). Metadata for this information can be accessed from the following link: https://dataview.ncbi.nlm.nih.gov/object/PRJNA529906?reviewer=3qsuo5i095ic5v1k7fupmmmsbj

Accession	Sample Description	File Name	SRA
SAMN11289175	Glenlea_Control1_0hr_R1_S1	GC10-1_S1_L007_R1_001.fastq.gz	SRR8816284
SAMN11289176	Glenlea_Control1_0hr_R2_S2	GC10-2_S2_L007_R1_001.fastq.gz	SRR8816283
SAMN11289177	Glenlea_Control1_16hr_R1_S3	GC116-1_S3_L007_R1_001.fastq.gz	SRR8816286
SAMN11289178	Glenlea_Control1_16hr_R2_S4	GC116-2_S4_L007_R1_001.fastq.gz	SRR8816285
SAMN11289179	Glenlea_Control1_8hr_R1_S5	GC18-1_S5_L007_R1_001.fastq.gz	SRR8816288
SAMN11289180	Glenlea_Control1_8hr_R2_S6	GC18-2_S6_L007_R1_001.fastq.gz	SRR8816287
SAMN11289181	Glenlea_Control2_0hr_R1_S7	GC20-1_S7_L007_R1_001.fastq.gz	SRR8816290
SAMN11289182	Glenlea_Control2_0hr_R2_S8	GC20-2_S8_L007_R1_001.fastq.gz	SRR8816289
SAMN11289183	Glenlea_Control2_16hr_R1_S9	GC216-1_S9_L007_R1_001.fastq.gz	SRR8816282
SAMN11289184	Glenlea_Control2_16hr_R2_S10	GC216-2_S10_L007_R1_001.fastq.gz	SRR8816281
SAMN11289185	Glenlea_Control2_8hr_R1_S11	GC28-1_S11_L007_R1_001.fastq.gz	SRR8816274
SAMN11289186	Glenlea_Control2_8hr_R2_S12	GC28-2_S12_L007_R1_001.fastq.gz	SRR8816273
SAMN11289187	Glenlea_Pathogen_0hr_R1_S13	GP0-1_S13_L007_R1_001.fastq.gz	SRR8816272
SAMN11289188	Glenlea_Pathogen_0hr_R2_S14	GP0-2_S14_L007_R1_001.fastq.gz	SRR8816271
SAMN11289189	Glenlea_Pathogen_16hr_R1_S15	GP16-1_S15_L007_R1_001.fastq.gz	SRR8816278
SAMN11289190	Glenlea_Pathogen_16hr_R2_S16	GP16-2_S16_L007_R1_001.fastq.gz	SRR8816277
SAMN11289191	Glenlea_Pathogen_8hr_R1_S17	GP8-1_S17_L007_R1_001.fastq.gz	SRR8816276
SAMN11289192	Glenlea_Pathogen_8hr_R2_S18	GP8-2_S18_L007_R1_001.fastq.gz	SRR8816275
SAMN11289193	Glenlea_Toxin_0hr_R1_S19	GT0-1_S19_L007_R1_001.fastq.gz	SRR8816280
SAMN11289194	Glenlea_Toxin_0hr_R2_S20	GT0-2_S20_L007_R1_001.fastq.gz	SRR8816279
SAMN11289195	Glenlea_Toxin_16hr_R1_S21	GT16-1_S21_L007_R1_001.fastq.gz	SRR8816295
SAMN11289196	Glenlea_Toxin_16hr_R2_S22	GT16-2_S22_L007_R1_001.fastq.gz	SRR8816296
SAMN11289197	Glenlea_Toxin_8hr_R1_S23	GT8-1_S23_L007_R1_001.fastq.gz	SRR8816293
SAMN11289198	Glenlea_Toxin_8hr_R2_S24	GT8-2_S24_L007_R1_001.fastq.gz	SRR8816294
SAMN11289199	Salamouni_Control1_0hr_R1_S25	SC10-1_S25_L008_R1_001.fastq.gz	SRR8816299
SAMN11289200	Salamouni_Control1_0hr_R2_S26	SC10-2_S26_L008_R1_001.fastq.gz	SRR8816300

SAMN11289201	Salamouni_Control1_16hr_R1_S27	SC116-1_S27_L008_R1_001.fastq.gz	SRR8816297
SAMN11289202	Salamouni_Control1_16hr_R2_S28	SC116-2_S28_L008_R1_001.fastq.gz	SRR8816298
SAMN11289203	Salamouni_Control1_8hr_R1_S29	SC18-1_S29_L008_R1_001.fastq.gz	SRR8816291
SAMN11289204	Salamouni_Control1_8hr_R2_S30	SC18-2_S30_L008_R1_001.fastq.gz	SRR8816292
SAMN11289205	Salamouni_Control2_0hr_R1_S31	SC20-1_S31_L008_R1_001.fastq.gz	SRR8816264
SAMN11289206	Salamouni_Control2_0hr_R2_S32	SC20-2_S32_L008_R1_001.fastq.gz	SRR8816263
SAMN11289207	Salamouni_Control2_16hr_R1_S33	SC216-1_S33_L008_R1_001.fastq.gz	SRR8816266
SAMN11289208	Salamouni_Control2_16hr_R2_S34	SC216-2_S34_L008_R1_001.fastq.gz	SRR8816265
SAMN11289209	Salamouni_Control2_8hr_R1_S35	SC28-1_S35_L008_R1_001.fastq.gz	SRR8816268
SAMN11289210	Salamouni_Control2_8hr_R2_S36	SC28-2_S36_L008_R1_001.fastq.gz	SRR8816267
SAMN11289211	Salamouni_Pathogen_0hr_R1_S37	SP0-1_S37_L008_R1_001.fastq.gz	SRR8816270
SAMN11289212	Salamouni_Pathogen_0hr_R2_S38	SP0-2_S38_L008_R1_001.fastq.gz	SRR8816269
SAMN11289213	Salamouni_Pathogen_16hr_R1_S39	SP16-1_S39_L008_R1_001.fastq.gz	SRR8816262
SAMN11289214	Salamouni_Pathogen_16hr_R2_S40	SP16-2_S40_L008_R1_001.fastq.gz	SRR8816261
SAMN11289215	Salamouni_Pathogen_8hr_R1_S41	SP8-1_S41_L008_R1_001.fastq.gz	SRR8816301
SAMN11289216	Salamouni_Pathogen_8hr_R2_S42	SP8-2_S42_L008_R1_001.fastq.gz	SRR8816302
SAMN11289217	Salamouni_Toxin_0hr_R1_S43	ST0-1_S43_L008_R1_001.fastq.gz	SRR8816303
SAMN11289218	Salamouni_Toxin_0hr_R2_S44	ST0-2_S44_L008_R1_001.fastq.gz	SRR8816304
SAMN11289219	Salamouni_Toxin_16hr_R1_S45	ST16-1_S45_L008_R1_001.fastq.gz	SRR8816305
SAMN11289220	Salamouni_Toxin_16hr_R2_S46	ST16-2_S46_L008_R1_001.fastq.gz	SRR8816306
SAMN11289221	Salamouni_Toxin_8hr_R1_S47	ST8-1_S47_L008_R1_001.fastq.gz	SRR8816307
SAMN11289222	Salamouni_Toxin_8hr_R2_S48	ST8-2_S48_L008_R1_001.fastq.gz	SRR8816308

ACKNOWLEDGEMENTS

Support for this dissertation work came from the USDA-NIFA hatch Projects to M. Nepal (SD00H469-13 and SD00H659-18), South Dakota Agriculture Experiment Station, and Department of Biology & Microbiology at South Dakota State University.

REFERENCES

1. Singh RP, Singh PK, Rutkoski J, et al. Disease Impact on Wheat Yield Potential and Prospects of Genetic Control. *Annual review of phytopathology.* 2016;54:303-322.

2. FAOSTAT. FAO Database. 2013; http://www.fao.org/faostat/en/ - data. Accessed August 3, 2017.

3. Murray TD, Bockus, W.W., Bowden, R.L., Hunger, R.M., and Smiley, R.W., collators. Diseases of Wheat (Triticum spp. L.). *Common Names of Plant Diseases* 2015; http://www.apsnet.org/publications/commonnames/Pages/Wheat.aspx. Accessed July 28, 2017.

4. Rees R, Platz G. Effects of yellow spot on wheat: comparison of epidemics at different stages of crop development. *Crop and Pasture Science.* 1983;34(1):39-46.

5. Shi G, Zhang Z, Friesen TL, et al. The hijacking of a receptor kinase–driven pathway by a wheat fungal pathogen leads to disease. *Science advances.* 2016;2(10):e1600822.

6. Faris JD, Liu Z, Xu SS. Genetics of tan spot resistance in wheat. *Theoretical and Applied Genetics.* 2013;126(9):2197-2217.

7. Figueroa M, Manning VA, Pandelova I, Ciuffetti LM. Persistence of the host-selective toxin Ptr ToxB in the apoplast. *Molecular Plant-Microbe Interactions.* 2015;28(10):1082-1090.

8. Ciuffetti LM, Tuori RP. Advances in the characterization of the Pyrenophora tritici-repentis—wheat interaction. *Phytopathology.* 1999;89(6):444-449.

9. Ali S, Francl LJ. Population race structure of Pyrenophora tritici-repentis prevalent on wheat and noncereal grasses in the Great Plains. *Plant Disease.* 2003;87(4):418-422.

10. Lamari L, Gilbert J, Tekauz A. Race differentiation in Pyrenophora tritici-repentis and survey of physiologic variation in western Canada. *Canadian journal of plant pathology.* 1998;20(4):396-400.

11. Faris JD, Zhang Z, Lu H, et al. A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. *Proceedings of the National Academy of Sciences.* 2010;107(30):13544-13549.

12. Liu Z, Zurn JD, Kariyawasam G, et al. Inverse gene-for-gene interactions contribute additively to tan spot susceptibility in wheat. *Theoretical and Applied Genetics.* 2017;130(6):1267-1276.

13. Manning VA, Chu AL, Steeves JE, Wolpert TJ, Ciuffetti LM. A host-selective toxin of Pyrenophora tritici-repentis, Ptr ToxA, induces photosystem changes and reactive oxygen species accumulation in sensitive wheat. *Molecular Plant-Microbe Interactions.* 2009;22(6):665-676.

14. Ciuffetti LM, Manning VA, Pandelova I, Betts MF, Martinez JP. Host‐selective toxins, Ptr ToxA and Ptr ToxB, as necrotrophic effectors in the Pyrenophora tritici‐repentis‐wheat interaction. *New Phytologist.* 2010;187(4):911-919.

15. Lamari L, Bernier C. Virulence of isolates of Pyrenophora tritici-repentis on 11 wheat cultivars and cytology of the differential host reactions. *Canadian Journal of Plant Pathology.* 1989;11(3):284-290.

16. Lamari L, Sayoud R, Boulif M, Bernier C. Identification of a new race in Pyrenophora tritici-repentis: implications for the current pathotype classification system. *Canadian Journal of Plant Pathology.* 1995;17(4):312-318.

17. Strelkov SE, Lamari L, Sayoud R, Smith RB. Comparative virulence of chlorosis-inducing races of Pyrenophora tritici-repentis. *Canadian Journal of Plant Pathology.* 2002;24(1):29-35.

18. Lamari L, Strelkov S, Yahyaoui A, Orabi J, Smith R. The identification of two new races of Pyrenophora tritici-repentis from the host center of diversity confirms a one-to-one relationship in tan spot of wheat. *Phytopathology.* 2003;93(4):391-396.

19. Petrov V, Qureshi MK, Hille J, Gechev T. Occurrence, biochemistry and biological effects of host-selective plant mycotoxins. *Food and Chemical Toxicology.* 2017.

20. Friesen TL, Stukenbrock EH, Liu Z, et al. Emergence of a new disease as a result of interspecific virulence gene transfer. *Nature genetics.* 2006;38(8):953-956.

21. Manning VA, Hardison LK, Ciuffetti LM. Ptr ToxA interacts with a chloroplast-localized protein. *Molecular plant-microbe interactions.* 2007;20(2):168-177.

22. Tai YS, Bragg J, Lu H, et al. Functional Characterization of Ptr ToxA and Molecular Identification of its Intracellular Targeting Protein in Wheat. Paper presented at: Plant and Animal Genome VX Conference Abstracts2007.

23. Manning VA, Hamilton SM, Karplus PA, Ciuffetti LM. The Arg-Gly-Asp–containing, solvent-exposed loop of Ptr ToxA is required for internalization. *Molecular plant-microbe interactions.* 2008;21(3):315-325.

24. Ciuffetti L, Francl L, Ballance G, et al. Standardization of toxin nomenclature in the Pyrenophora tritici-repentis/wheat interaction. *Canadian Journal of Plant Pathology.* 1998;20(4):421-424.

25. Effertz R, Meinhardt S, Anderson J, Jordahl J, Francl L. Identification of a chlorosis-inducing toxin from Pyrenophora tritici-repentis and the chromosomal location of an insensitivity locus in wheat. *Phytopathology.* 2002;92(5):527-533.

26. Abdullah S, Sehgal SK, Ali S, Liatukas Z, Ittu M, Kaur N. Characterization of Pyrenophora tritici-repentis (Tan Spot of Wheat) Races in Baltic States and Romania. *The plant pathology journal.* 2017;33(2):133.

27. Lamari L, Bernier C. Evaluation of wheat lines and cultivars to tan spot [Pyrenophora tritici-repentis] based on lesion type. *Canadian Journal of Plant Pathology.* 1989;11(1):49-56.

28. Ali S, Francl L. Recovery of Pyrenophora tritici-repentis from barley and reaction of 12 cultivars to five races and two host-specific toxins. *Plant disease.* 2001;85(6):580-584.

29. Jordahl J, Francl L. Increase and storage of cultures of Pyrenophora tritici-repentis. 1992.

30. Faris J, Anderson J, Francl L, Jordahl J. Chromosomal location of a gene conditioning insensitivity in wheat to a necrosis-inducing culture filtrate from Pyrenophora tritici-repentis. *Phytopathology.* 1996;86(5):459-463.

31. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010.

32. Kong Y. Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics.* 2011;98(2):152-153.

33. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nature methods.* 2015;12(4):357.

34. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015;31(2):166-169.

35. Kersey PJ, Allen JE, Christensen M, et al. Ensembl Genomes 2013: scaling up access to genome-wide data. *Nucleic acids research.* 2014;42(D1):D546-D552.

36. Team RC. R: A language and environment for statistical computing. 2013.

37. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology.* 2014;15(12):550.

38. Jones P, Binns D, Chang H-Y, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014:btu031.

39. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of molecular biology.* 1990;215(3):403-410.

40. Ge SX. iDEP: An integrated web application for differential expression and pathway analysis. *bioRxiv.* 2017:148411.

41. Jones JD, Dangl JL. The plant immune system. *Nature.* 2006;444(7117):323-329.

42. Bacete L, Mélida H, Miedes E, Molina A. Plant cell wall‐mediated immunity: cell wall changes trigger disease resistance responses. *The Plant Journal.* 2018;93(4):614-636.

43. Pandelova I, Betts MF, Manning VA, Wilhelm LJ, Mockler TC, Ciuffetti LM. Analysis of transcriptome changes induced by Ptr ToxA in wheat provides insights into the mechanisms of plant susceptibility. *Molecular plant.* 2009;2(5):1067-1083.

44. Frey M, Schullehner K, Dick R, Fiesselmann A, Gierl A. Benzoxazinoid biosynthesis, a model for evolution of secondary metabolic pathways in plants. *Phytochemistry.* 2009;70(15-16):1645-1651.

45. Du Fall LA, Solomon PS. The necrotrophic effector S n T ox A induces the synthesis of a novel phytoalexin in wheat. *New Phytologist.* 2013;200(1):185-200.

46. Ordonio R, Ito Y, Morinaka Y, Sazuka T, Matsuoka M. Molecular breeding of Sorghum bicolor, a novel energy crop. *International review of cell and molecular biology.* Vol 321: Elsevier; 2016:221-257.

47. Bak S, Paquette SM, Morant M, et al. Cyanogenic glycosides: a case study for evolution and application of cytochromes P450. *Phytochemistry Reviews.* 2006;5(2-3):309-329.

48. Wang Q, Chen D, Wu M, et al. MFS transporters and GABA metabolism are involved in the self-defense against DON in Fusarium graminearum. *Frontiers in plant science.* 2018;9.

49. Ma Q-H, Liu Y-C. TaDIR13, a dirigent protein from wheat, promotes lignan biosynthesis and enhances pathogen resistance. *Plant molecular biology reporter.* 2015;33(1):143-152.

50. Yoo TH, Park C-J, Ham B-K, Kim K-J, Paek K-H. Ornithine decarboxylase gene (CaODC1) is specifically induced during TMV-mediated but salicylate-independent resistant response in hot pepper. *Plant and cell physiology.* 2004;45(10):1537-1542.

51. Woldemariam MG, Ahern K, Jander G, Tzin V. A role for 9-lipoxygenases in maize defense against insect herbivory. *Plant signaling & behavior.* 2018;13(1):4709-4723.

52. Kariyawasam GK, Carter AH, Rasmussen JB, et al. Genetic relationships between race-nonspecific and race-specific interactions in the wheat–Pyrenophora tritici-repentis pathosystem. *Theoretical and Applied Genetics.* 2016;129(5):897-908.

53. Faris J, Friesen T. Identification of quantitative trait loci for race-nonspecific resistance to tan spot in wheat. *Theoretical and applied genetics.* 2005;111(2):386-392.

54. Dushnicky L, Ballance G, Sumner M, MacGregor A. Penetration and infection of susceptible and resistant wheat cultivars by a necrosis toxin-producing isolate of Pyrenophora tritici-repentis. *Canadian Journal of Plant Pathology.* 1996;18(4):392-402.

55. Larez C, Hosford Jr R, Freeman T. Infection of wheat and oats by Pyrenophora tritici-repentis and initial characterization of resistance. *Phytopathology.* 1986;76(9):931-938.

56. Cheong J, Wallwork H, Williams K. Identification of a major QTL for yellow leaf spot resistance in the wheat varieties Brookton and Cranbrook. *Crop and Pasture Science.* 2004;55(3):315-319.

57. Friesen T, Faris J. Molecular mapping of resistance to Pyrenophora tritici-repentis race 5 and sensitivity to Ptr ToxB in wheat. *Theoretical and applied genetics.* 2004;109(3):464-471.

58. Lorang JM, Sweat TA, Wolpert TJ. Plant disease susceptibility conferred by a “resistance” gene. *Proceedings of the National Academy of Sciences.* 2007;104(37):14861-14866.

59. Ali S, Gurung S, Adhikari TB. Identification and characterization of novel isolates of Pyrenophora tritici-repentis from Arkansas. *Plant disease.* 2010;94(2):229-235.

APPENDIX I: EVOLUTIONARY CONSERVATION OF DISEASE RESISTANCE COMPONENTS ACROSS MULTIPLE GRASS SPECIES

Plant pathogens utilize effectors to interfere with host resistance responses as a way to facilitate infection. Many targets of these effectors regulate the expression of genes associated with defense system. Plants initiate responses when pathogen effectors cleave critical target proteins. *Pseudomonas syringae* cysteine protease effector *Pseudomonas phaseolicola* B (AvrPphB) cleaves the *Arabidopsis thaliana* cytoplasmic kinase protein AVRPPHB SUSCEPTIBLE1 (PBS1) and other PBS1-like (PBL) kinases to interfere with immune signaling. The resistance protein RESISTANCE TO PSEUDOMONAS SYRINGAE5 (RPS5) recognizes the cleavage of PBS1 and initiates a response to the bacterial pathogen. While elucidated in *Arabidopsis thaliana*, recent literature suggests that grasses, such as wheat, also possess PBS1 homologs that share the cleavage domain utilized by AvrPphB. The objectives of this study were to sequence and compare the region directly downstream of the cleavage site in PBS1 homologs of both cultivated and wild grass family members. DNA was extracted from leaf tissue and PCR was used to amplify a conserved region of PBS1 homologs. These results show that PBS1 has been evolutionarily conserved throughout Poaceae lineages, specifically the region coding for the cleavage site of the AvrPphB effectors. While the role of PBS1 as a regulator of resistance response may not have been retained among these species, a selective advantage was likely present to preserve the function of PBS1 as a decoy for bacterial effectors.

Plant pathogens use effectors to inhibit plant defense responses, thereby facilitating infection [1]. Effectors are diverse molecules that are often proteins, although non-proteinaceous effectors also can be produced. Method of effector delivery also varies, some exogenously applied to the host tissue and others directly injected into host cytoplasm through type III secretion systems [2]. Plants have evolved to detect these effectors by producing receptor proteins that initiate defense responses upon activation [1]. These receptors can directly bind to effector proteins or become activated upon interaction with a modified effector target. In the latter case, the pathogen effector typically cleaves a target protein that is then recognized by the receptor. Effector targets generally act as regulators of resistance responses, cleavage then leading to inhibited response [3,4]. The *Arabidopsis thaliana* protein AVRPPHB SUSCEPTIBLE1 (PBS1) acts as a target protein for the *Pseudomonas syringae* cysteine protease effector Pseudomonas phaseolicola B (AvrPphB) [5]. This cleavage of PBS1 is recognized by RESISTANCE TO PSEUDOMONAS SYRINGAE5 (RPS5), initiating a resistance response. Recently discovered in 2017, grasses, such as wheat, also possess *PBS1* homologs that share cleavage domains utilized by AvrPphB [6]. The objectives of this project are to sequence *PBS1* homologs in multiple crop and wild grass species and assess *PBS1* homologs for AvrPphB cleavage sites and other conserved features.

PBS1 homolog sequences were accessed from Sun et al. (2017) and the National Center for Biotechnology Information (NCBI) [6]. Sequences were aligned and used to construct a maximum-likelihood phylogenetic tree using the program MEGA 7 with 100 bootstrap replicates. Based on aligned PBS1 homolog sequences, primers were designed to amplify an approximately 700 base pair long region of *PBS1*. Plants were grown until

enough biomass was available, tissue was harvested, and DNA was extracted. Polymerase Chain Reaction (PCR) was used to amplify the *PBS1* region after optimization for annealing temperature and magnesium chloride concentration. Amplification was confirmed through gel electrophoresis, amplicons were purified, and samples were sent to Iowa State University for sequencing. Sequences were assessed for quality and contigs were assembled using the program Sequencer. Sequences were aligned to assess structural conservation

PBS1 homologs were successfully sequenced in 11 of the species in which successful amplification was observed (see **Figures I.1** and **I.2**). Variation between species was not only observed in the nucleotide sequences of *PBS1* (**Figure I.1**), but also in the overall length of amplicons (**Figure I.2**). Based on more complete phylogeny using sequences from NCBI (**Figure I.3**), *PBS1* homologs show a similar phylogeny to the expected nesting of taxa based on current classification. Data on wild grasses (**Table I.1**) was not available in NCBI, thus prompting this study to assess the evolutionary conservation of *PBS1*. Presence of clear bands after gene amplification indicates that each of the species in Table 1 possess homologs of PBS1. The region amplified appears to have variable lengths due to the difference in size of the amplicons. This indicates that *PBS1* will show variation in its sequence between the species, which was observed in the those that yielded sequences (**Figure I.1**). Alignment of the *PBS1* homologs delivered an explanation for the diversity in amplicon lengths: unique insertions exist in many of the species, especially in barley, rice, and maize. The first ~200 bases of the amplicons correspond to one of the *PBS1* exons, followed by a ~400 base intron, and ending in ~50 bases of the next exon. This region was chosen due to the presence of the cleavage site

for AvrPphB and sequence conservation allowing for multi-species amplification. AvrPphB cleaves PBS1 immediately following the GDK motif (amino acids: GDK, nucleotides: GGTGACAAG, downstream: TCTCATGTCTCTACACGTGTGATGGG). As shown in the alignment, this region is conserved among wild grass species. This indicates that *P. syringae* effector AvrPphB can cleave PBS1 homologs in wild grasses, as well as crops. Future research should focus on adding sequence data by re-sequencing unavailable taxa and sequencing this region in additional species. Primers should be designed for additional components of disease resistance in grass species. With such data, comparison in evolutionary rate of various component genes could be assessed to determine which groups evolve more quickly in response to pathogen pressures.

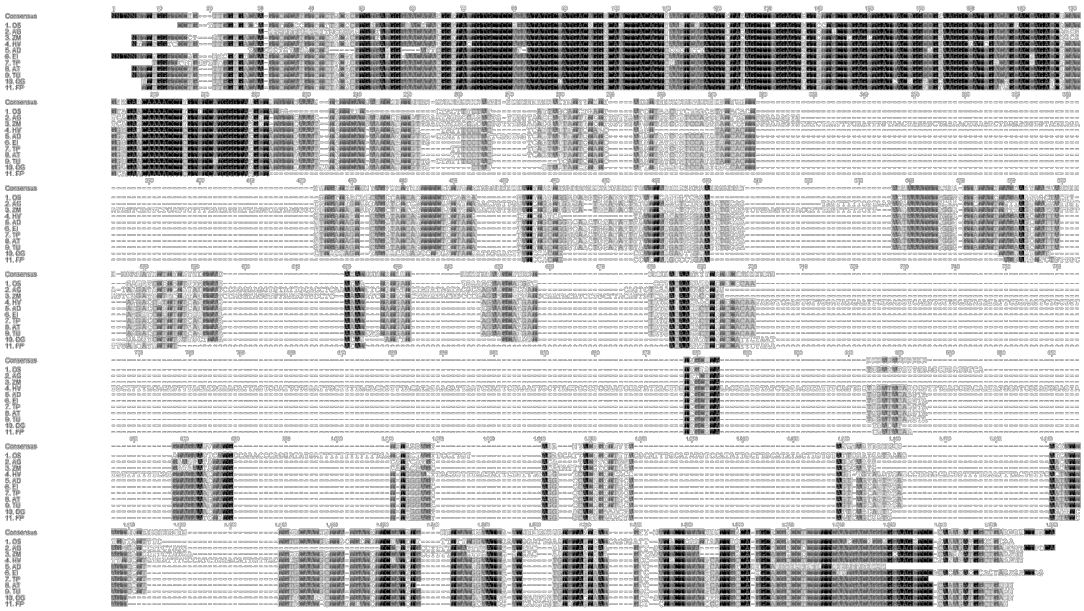


Figure I.1 Alignment of *PBS1* homologs in *Oryza sativa* (OS), *Andropogon gerardii* (AG), *Zea mays* (ZM), *Hordeum vulgare* (HV), *Agropyron desertorum* (AD), *Elytrigia intermedium* (EI), *Thinopyrum ponticum* (TP), *Aegilops tauschii* (AT), *Triticum uruartu* (TU), *Dactylis glomerata* (DG), and *Festuca pratensis* (FP).

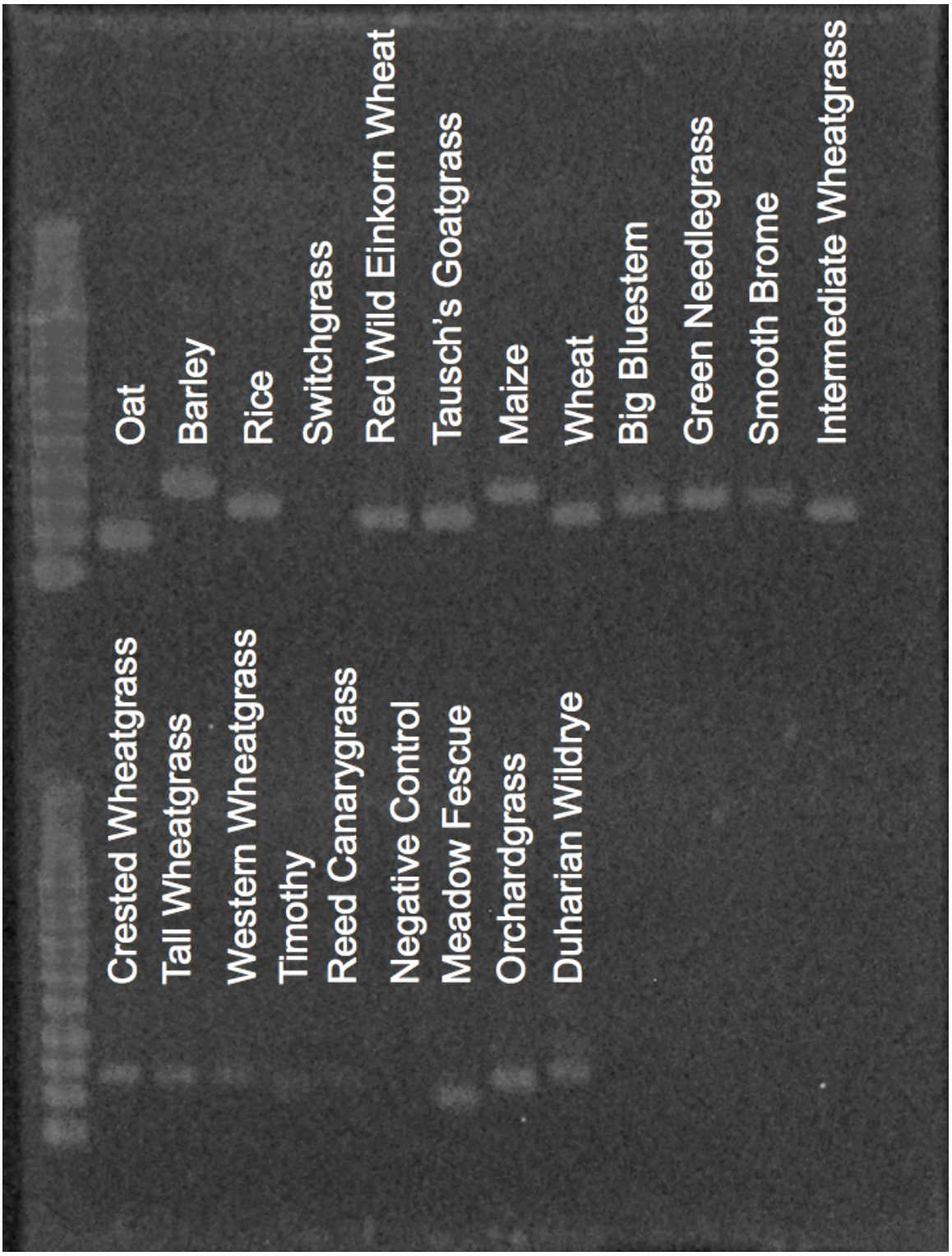


Figure I.2 Electrophoresis gel (agarose, 1.5%) used to confirm the presence of *PBS1* amplicons in various grass species.

Table I.1 Species selected for analysis based on their status as crop species or wild grasses used for forage in the Great Plains.

Species	Common Name	Type	Native
Aegilops tauschii	Tausch's Goatgrass	Crop Progenitor	
Agropyron desertorum	Crested Wheatgrass	Wild	
Andropogon gerardii	Big Bluestem	Wild	Yes
Avena sativa	Oat	Crop	
Bromus inermis	Smooth Bromegrass	Wild	
Dactylis glomerata	Orchardgrass	Wild	
Elymus dahuricus	Duharian Wildrye	Wild	
Elytrigia intermedium	Intermediate Wheatgrass	Wild	
Festuca pratensis	Meadow Fescue	Wild	
Hordeum vulgare	Barley	Crop	
Nassella viridula	Green Needlegrass	Wild	Yes
Oryza sativa	Rice	Crop	
Panicum virgatum	Switchgrass	Wild	Yes
Pascopyrum smithii	Western Wheatgrass	Wild	Yes
Phalaris arundinacea	Reed canarygrass	Wild	Yes
Phleum pretense	Timothy	Wild	Yes
Thinopyrum ponticum	Tall Wheatgrass	Wild	
Triticum aestivum	Common Wheat	Crop	
Triticum urartu	Red Wild Einkorn Wheat	Crop Progenitor	
Zea mays	Maize/Corn	Crop	

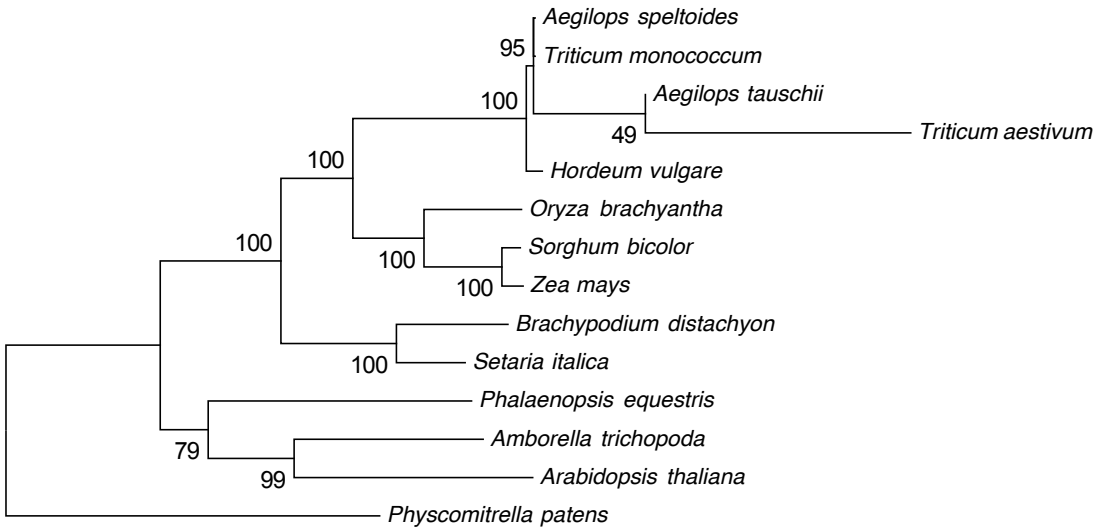


Figure I.3 Maximum-likelihood phylogenetic tree (MEGA 7, 100 bootstraps) using *PBS1* homologs accessed from NCBI. Sequence alignment was used to design primers for amplification shown in Figures 1 and 2.

REFERENCES

1. Jones JD, Dangl JL. The plant immune system. *Nature.* 2006;444(7117):323-329.

2. Boch J, Bonas U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annual review of phytopathology.* 2010;48.

3. Meyers BC, Kaushik S, Nandety RS. Evolving disease resistance genes. *Current opinion in plant biology.* 2005;8(2):129-134.

4. van der Hoorn RA, Kamoun S. From guard to decoy: a new model for perception of plant pathogen effectors. *The Plant Cell.* 2008;20(8):2009-2017.

5. Shao F, Golstein C, Ade J, Stoutemyer M, Dixon JE, Innes RW. Cleavage of Arabidopsis PBS1 by a bacterial type III effector. *Science.* 2003;301(5637):1230-1233.

6. Sun J, Huang G, Fan F, et al. Comparative study of Arabidopsis PBS1 and a wheat PBS1 homolog helps understand the mechanism of PBS1 functioning in innate immunity. *Scientific reports.* 2017;7(1):5487.

APPENDIX II: EVOLUTION OF THE NB-ARC PROTEIN DOMAIN AS A MAJOR SIGNALING COMPONENT OF THE PLANT DEFENSE RESPONSE

Conserved cellular signaling mechanisms in diverse species point to the utility of these complexes throughout evolutionary history. Plants, animals, fungi, and bacteria all share proteins containing a highly conserved domain that causes conformational shifts after protein activation and exchange ADP for ATP [1]. This domain is named after its constituent motifs and the identified proteins that it is found in: Nucleotide-Binding, Apoptotic protease activating factor 1, plant R-protein, *Caenorhabditis elegans* Death protein-4 (NB-ARC) [2]. This domain is generally associated with planned cell death in the form of apoptosis in animal systems and hypersensitive response in plants [2]. In plants, these proteins are immune system receptors that detect pathogenic effectors in order to initiate a response. In order for plants to resist devastating diseases, they make use of the NB-ARC by having it flanked by a C-terminal Leucine-Rich Repeat (LRR) and either an N-terminal Toll/Interleukin Receptor (TIR) or Coiled-Coil (CC), TNL and CNL, respectively [1]. TNL and CNL genes have been identified in dicot plants, but TNL genes are rare in monocots [3]. These genes have become a major focus of research due to the economic importance of cereal crop production in the maintenance of the global food supply [4,5]. The exponentially increasing availability of genomic and other biological data provides insight into how gene families evolve, through the identification and comparison of conserved sequences [3]. The number and variety of sequences containing NB-ARCs in plants have expanded voluminously in comparison to similar sequences in other organisms [1]. Understanding why the NB-ARC became such a large player in the plant immune system may provide information as to the advantages organisms gain through the

use of NB-ARC signaling mechanisms. The main objectives of this project are to identify NB-ARC protein sequences in diverse genomes that represent many phylogenetic groups, elucidate the evolutionary relationships between NB-ARCs in early and higher plants, and predict alternative functions for NB-ARC-containing sequences that may not be involved in pathogen resistance responses.

Genomic data from Phytozome and Ensembl Genomes was accessed and all proteomes were downloaded for species of interest. R-gene identification was run with two Hidden Markov Models using *Arabidopsis thaliana* NB-ARC reference sequences. MEME software was then used to isolate genes that contained P-loop, Kinase-2, and GLPL motifs. Sequences were then analyzed using BLAST to identify homologs and predict gene function. MEGA 7, Biomart, InterProScan 5, and DnaSP 5.10.1 were used to generate phylogenetic trees, sequence annotations, and selection pressure information. NB-ARCs were identified through InterProScan annotation using the Pfam Family PF00931. Since the focus of this study is the identification of NB-ARCs in a wide range of species, this was a more inclusive method than only identifying sequences with all conserved motifs. Biomart was also used to search candidate genomes for any sequences that contained the PF00931 annotation.

The expansion of genes encoding proteins with NB-ARCs can be seen in **Figure II.1**. Earlier plant species and non-plants generally have only a few to a few dozen NB-ARC proteins. This massive expansion is likely caused by the important role that these receptor proteins play. Acting as a necessary component of pathogenic effector receptors makes these proteins highly involved in the survival of plants when pathogen pressure is present. Also, with some plants, such as wheat and soybean, hybridization and whole

genome duplications, respectively, cause a large increase in the number of genes encoding these proteins. Number of CNL resistance genes identified based upon conserved motifs within the NB-ARC can be seen in **Table II.1**. The single celled green algae *Chylamydomonas reinhardtii* contained no NB-ARC sequences at all, at least none that InterProScan could annotate as containing an NB-ARC. The same follows for other members of the Chlorophyte group: *Dunaliella salina*, *Volvox carteri*, *Coccomyxa subellipsoidea*, *Micromonas pusilla*, and *Ostreococcus lucinmarinus*. Only looking at the plant lineage, it would be assumed that NB-ARCs arose in early plants. However, since NB-ARC sequences have been identified in animals, fungi, and bacteria [6], it would appear that many Chlorophytes lost NB-ARC containing sequences, possibly through pseudogenization or deletion. Availability of the *Chromochloris zofingiensis* genome [7] allows for the identification of NB-ARCs that are more closely related to the Chlorophyte group, used to search for similar sequences in the Chlorophyte genomes. When *C. zofingiensis* NB-ARC-containing sequences are used to BLAST *C. reinhardtii* proteins, it can be seen that some *C. reinhardtii* proteins possess amino acids similar to NB-ARC-containing proteins after the GLPL motif, including the Leucine-rich repeats (**Figure II.2**).

The red algae *Chondrus crispus* has 58 NB-ARC-containing sequences in its genome. This indicates that red algae may have experienced an expansion in these genes independent of the expansion seen in green plants. Also, in contrast to some green algae, cyanobacteria possess several NB-ARC containing sequences, although the number varies greatly with the species of cyanobacteria. Several fungi protein sequences possess NB-ARCs, for instance, *Fusarium oxysporum* has four proteins with NB-ARCs. Several

other fungal species for which genomes were available also showed similar numbers of NB-ARC-containing genes. While these genes have not diversified nearly to the extent that they have in plants, they may serve in important roles, like APAF1 in humans. Ten genes within the *Streptomyces coelicolor* genome contain NB-ARCs. Submitting these sequences to NCBI BLAST results in ATP-binding protein similarity, which does not give much information as to their function. Accessions CAB88433 and CAB38588 are most similar to a regulatory protein that is involved in antibiotic production [8]. This could be expected since the NB-ARC in plants does not directly carry out defense responses, but rather is involved in the signaling that leads to such responses. Because of this, the signaling function of the NB-ARC could have many applications in different cellular systems. It can be hypothesized that NB-ARC receptors, initially used for recognition of pathogen effectors could evolve to become receptors for other proteins to initiate entirely different cellular responses, or vice versa.

Overall, the NB-ARC sequences of even highly divergent species showed multiple conserved regions (**Figure II.3**). The NB-ARC proteins in *C. zofingiensis* all possessed TIR domains, possibly indicating that TNL genes may have been ancestral to CNL genes. This would mean that TNL genes were lost from the genomes of grasses, instead of TNL genes independently arising in dicot genomes. Future studies should focus on characterizing the exact functions of early NB-ARC-containing proteins to determine if they are involved in pathogen resistance responses or if these organisms have evolved to utilize them for some other purpose. Knowledge about the NB-ARC and associated domains will become more important as more studies work to elucidate plant

resistance pathways and as genome editing becomes sophisticated enough to engineer plants with novel resistance mechanisms for devastating pathogens.

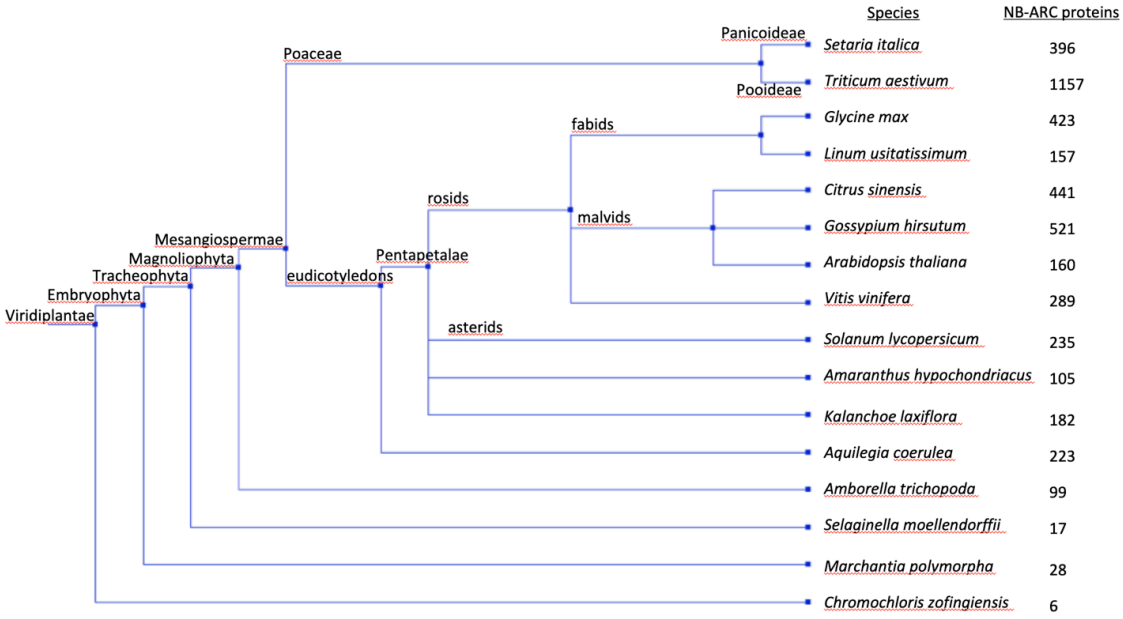


Figure II.1 Unique proteins containing NB-ARC amino acid domain sequences per species, arranged based upon phylogeny with selected groups labeled. These numbers do not include the various alternative transcripts available for many of the proteins found, with alternative splicing a possible mechanism for further variation in the proteins with these domains. The phylogeny shown here is based on the NCBI Taxonomy Browser.

Table II.1 CNL genes for selected species. These numbers are lower than the overall amount of NB-ARC-containing sequences since not all NB-ARCs contain the conserved domains used to categorize CNL sequences.

Species	CNL	Species	CNL
Aegilops tauschii	392	*Physcomitrella patens*	9
Amborella trichopoda	46	*Selaginella moellendorffii*	7
Ananas cosmosus	126	*Setaria italica*	278
Arabidopsis thaliana	52	*Sorghum bicolor*	177
Brachypodium distachyon	223	*Spirodela polyrhiza*	60
Hordeum vulgare	183	*Triticum aestivum*	609
Musa acuminata	64	*Triticum urartu*	318
Oropetium thomaeum	4	*Zea mays*	75
Oryza sativa	301	*Zostera marina*	26

Figure II.2 Example InterProScan annotation comparison of *C. zofingiensis* and *C. reinhardtii* homolog sequences. The *C. zofingiensis* NB-ARC-containing gene above contains TIR, NB-ARC, and LRR domains, appearing very similar to a dicot TNL gene in species such as *Glycine max*. The *C. reinhardtii* sequence contained mostly LRR annotations, with strong similarity to the LRR of the above gene. Dashed lines represent 100 amino acids.

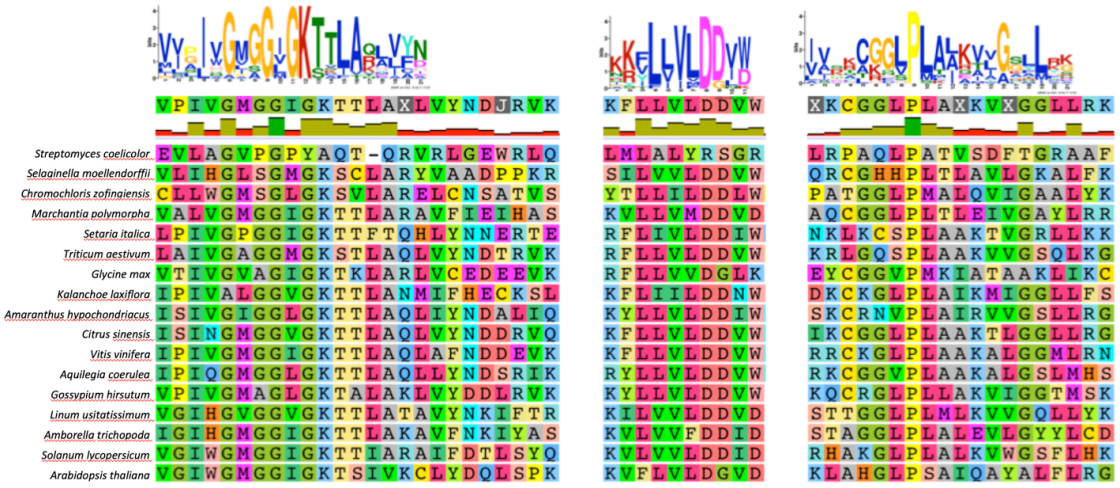


Figure II.3 Alignment of selected NB-ARC regions: P-loop, Kinase-2, GLPL motifs. Above the selected alignment sections are the Multiple Expectation Maximization for Motif Elicitation (MEME) diagrams representing the prevalence of amino acids at each given location.

REFERENCES

1. Takken FL, Goverse A. How to build a pathogen detector: structural basis of NB-LRR function. *Current opinion in plant biology.* 2012;15(4):375-384.

2. van der Biezen EA, Jones JD. The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. *Current Biology.* 1998;8(7):R226-R228.

3. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of NBS-LRR–encoding genes in Arabidopsis. *The Plant Cell Online.* 2003;15(4):809-834.

4. Periyannan S, Moore J, Ayliffe M, et al. The gene Sr33, an ortholog of barley Mla genes, encodes resistance to wheat stem rust race Ug99. *Science.* 2013;341(6147):786-788.

5. Saintenac C, Zhang W, Salcedo A, et al. Identification of wheat gene Sr35 that confers resistance to Ug99 stem rust race group. *Science.* 2013;341(6147):783-786.

6. Finn RD, Bateman A, Clements J, et al. Pfam: the protein families database. *Nucleic acids research.* 2013:gkt1223.

7. Roth MS, Cokus SJ, Gallaher SD, et al. Chromosome-level genome assembly and transcriptome of the green alga Chromochloris zofingiensis illuminates astaxanthin production. *Proceedings of the National Academy of Sciences.* 2017;114(21):E4296-E4305.

8. Lee PC, Umeyama T, Horinouchi S. afsS is a target of AfsR, a transcriptional factor with ATPase activity that globally controls secondary metabolism in Streptomyces coelicolor A3(2). *Mol Microbiol.* 2002;43(6):1413-1430.

APPENDIX III: SHORT COMMUNICATION: INITIAL INVESTIGATION INTO DISEASE RESISTANCE GENES IN WHEAT (*TRITICUM AESTIVUM* L.; POACEAE)

Crop diseases pose a major threat to farmers around the globe. Given the correct environmental conditions, pathogens can cause the loss of an entire field, severely reducing income from yield as well as impacting prices paid by consumers. Due to the major role wheat (*Triticum aestivum* L.) plays in the global food supply, diseases impacting its production have been the topic of much research. Epidemics of stem rust (*Puccinia graminis*) in the early 20th century promoted the development of resistant cultivars, such as Hope Wheat developed by Edgar McFadden [1]. It was later discovered that the resistance gene (R-gene) Sr2 allowed Hope Wheat to resist infection by stem rust [2]. R-genes encode proteins that detect pathogen effectors and initiate defense responses [3]. A major component of an R-protein is the NB-ARC, a highly conserved amino-acid sequence that is flanked by a Leucine-Rich Repeat (LRR) on the C-terminal end and either a Toll/Interleukin Receptor (TIR) or Coiled-Coil (CC) at the N-terminal end (altogether named TNL and CNL, respectively) [3]. Dicot plants generally possess both TNL and CNL genes, whereas monocots only posses CNL genes [4]. CNL genes have been shown to confer resistance to the devastating Ug99 strain of stem rust that has recently become a major issue for farmers in Africa and parts of Asia [5,6]. The hexaploid bread wheat genome was formed through the hybridization of three different species, typically represented as AABBDD for the A, B, and D genomes [7]. The complete wheat genome sequence was recently made available, along with two of wheat's progenitors, *Triticum urartu* and *Aegilops tauschii*, contributors of the A and D genomes, respectively [7-9]. This enables the study of gene families, such as NB-ARC-containing CNL genes, and the

comparison of the three genomes. The main objectives of this project are to identify CNL R-genes in *T. aestivum*, *T. urartu*, and *A. tauschii* genomes using the most recent genomic data and predict functionality of the R-genes based upon homology and elucidate sequence patterns existing among these three species.

Protein sequences were accessed from Phytozome and two Hidden Markov Models were constructed to identify NB-ARC sequences. MEME software was then used to isolate genes that contained P-loop, Kinase-2, and GLPL motifs. Sequences were then analyzed using BLAST to identify homologs and predict gene function. NB-ARC sequences were aligned using Clustal Omega. MEGA 7 was then used to generate a Neighbor-Joining tree with 100 bootstrap replicates and the JTT+G model. Exon-intron variation, genome location, syntenic maps, selection pressures, and duplications/losses were determined using data from the Phytozome Biomart and the Gene Structure Display Server 2.0, Geneious 8.1, SyMAP 4.2, DnaSP 5.10.1, and Notung programs, respectively.

Synteny between wheat and selected other monocot genomes was visualized, with wheat-barley, wheat-rice, and wheat-maize synteny displayed in **Figure III.1**. In the most current *T. aestivum*, *T. urartu*, and *A. tauschii* genomes, 609, 318, and 392 CNL R-genes were identified. Homologs of the 609 wheat sequences showed a large number of RPM1, RPP13, and RGA homologs (**Table III.1**). Duplication-loss analysis showed many duplications in wheat and its progenitors, but an even larger number of losses (**Figure III.2** and **Figure III.3**). Wheat, barley, and wheat progenitors are all members of the Triticeae tribe, sharing very similar genomic organization. This can be seen in the syntenic map (**Figure III.1**). All seven chromosomes of wheat's A, B, and D genomes shares a majority of synteny to barley's seven chromosomes, respectively. One major

difference is that barley chromosomes are much larger than individual wheat chromosomes, wheat possessing a 17 billion base genome spread amongst 21 chromosome pairs. Barley contains a 5.1 billion base genome spread amongst seven chromosome pairs, likely the result of chromosomal duplications within the barley lineage. The largest contributor of the wheat genome, a relative of *Aegilops speltoides*, has yet to be confirmed in wheat literature [7]. Neither has the genome of *Aegilops speltoides* been sequenced. This elusive progenitor has contributed much in terms of pathogen resistance and the 3B chromosome has been studied for its R-gene content [10]. A neighbor-joining tree (100 bootstrap replicates; model: JTT+G) comparing the 1319 R-protein NB-ARCs between wheat and its progenitors was excluded due to size constraints. Interspecific nesting was consistent throughout the tree. This was expected since the hybridization events that led to the formation of bread wheat happened recently, only in the last few thousand years, and has not given the wheat progenitors enough time to diversify, while wheat undergoes artificial selection through breeding for economically valuable traits. In many cases, the *T. urartu* and *A. tauschii* genes nest with their homolog in wheat. In several cases, wheat contains multiple homologs of single genes in the progenitors, indicating tandem duplications that took place since the two diverged, or the specific progenitor population of *A. tauschii* or *T. urartu* contained the duplication that is not represented in their respective genomes. As with previous analyses of monocot CNL genes, the CNL-C clade, as defined in Meyers et al. (2003), has greatly expanded through duplication early in the monocot lineage. Along with this expansion in CNL-C, there is a reduction in CNL-A, CNL-B, and complete absence of CNL-D, as observed across other grasses.

Implementation of a duplication-loss model allows for the visualization of where duplications and losses took place in the context of a phylogenetic tree. Many losses were found for each of the three genomes studied, with many more losses present in wheat progenitors than wheat itself. Also, several characteristics were consistent with previous monocot CNL gene analyses. These R-genes were composed of multiple exons, indicating a possible mechanism of alternative splicing to regulate resistance responses and diversification of protein function. Purifying selection shown among the coding sequences indicates that the sequences are not undergoing positive selection, but are retaining key sequences necessary for their function. Future research should focus on improved genomic annotation of the three wheat genomes in this study. Genomic data should be compiled for *Aegilops speltoides*, the closest proposed relative of the wheat B genome progenitor. Building on the functional prediction of R-genes, expression data should be gathered from resistant and susceptible varieties for various pests in order to characterize R-genes. Cultivar-specific sequences could provide insight into how resistant and susceptible varieties differ in their defense responses, leading to the development of cultivars with durable resistance.

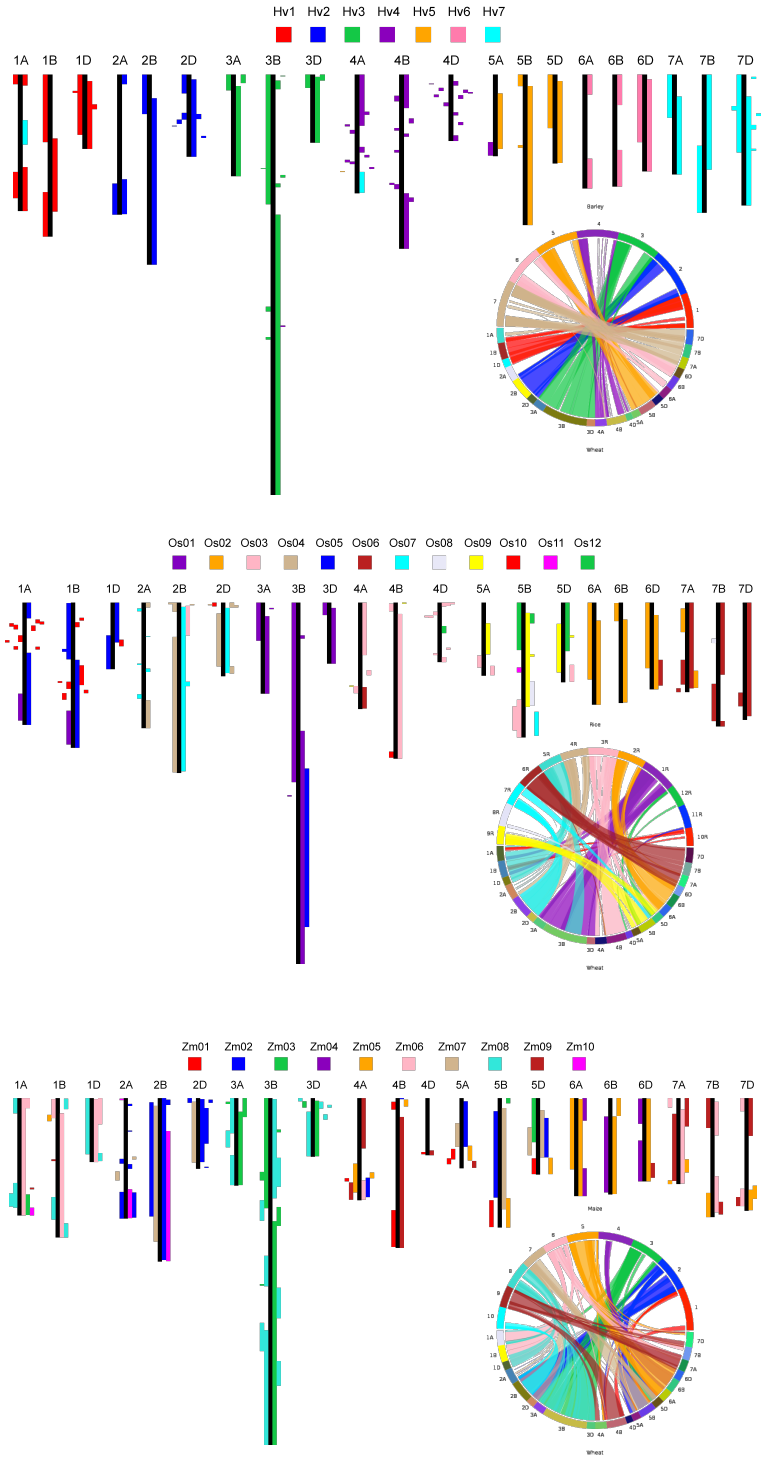


Figure III.1 Genomic synteny between wheat, barley, rice, and maize. Black lines represent the 21 wheat chromosomes with colors used to represent similar regions

corresponding to the labeled chromosomes of barley, rice, and maize. The CIRCOS-style figures show a different view of the same data, with the 21 wheat chromosomes located below the barley, rice, and maize chromosomes, respectively. Colors for the non-wheat chromosomes are not representative of the CIRCOS-style figures.

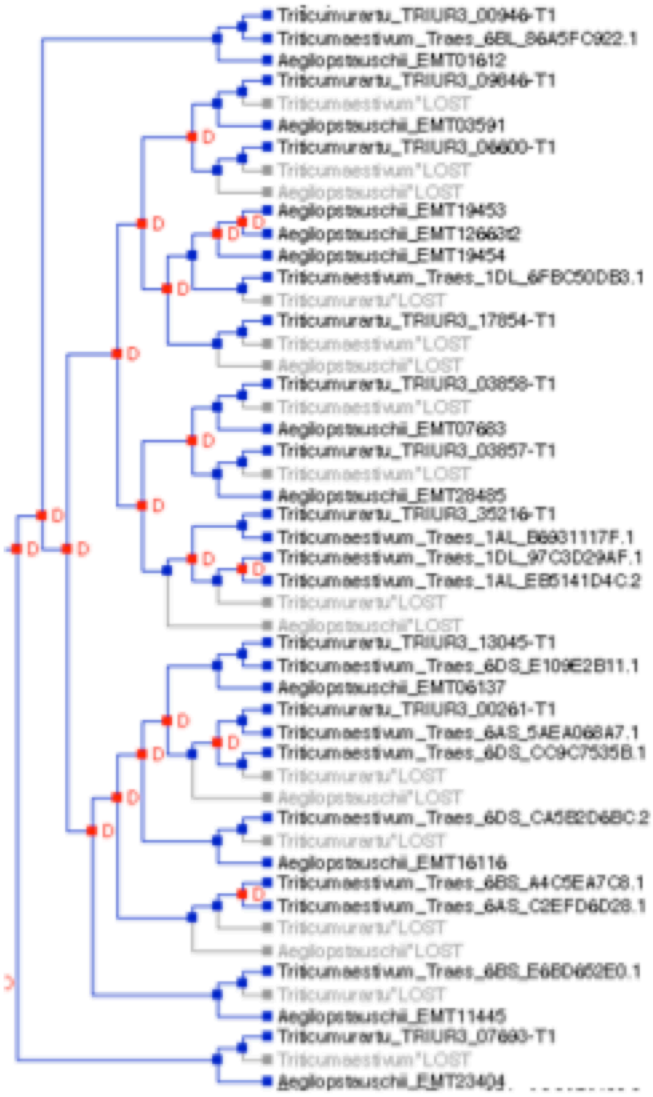


Figure III.2 Selected section of the neighbor-joining tree inferred using a duplication loss model, with red squares representing locations where proposed duplications might have occurred and grey accession names where proposed losses

have taken place. The N-J tree was constructed using the JTT+G model with 100 bootstrap replicates.

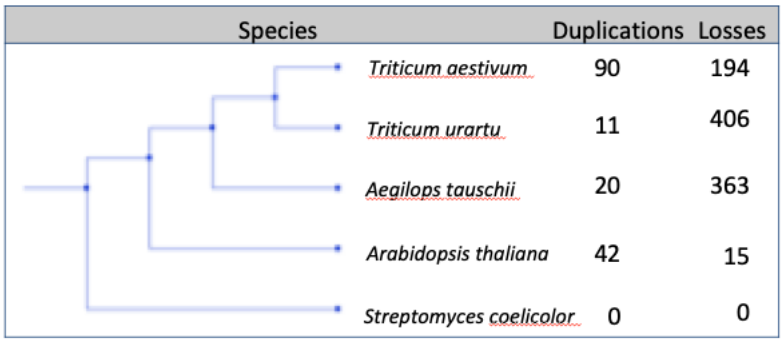


Figure III.3 Predicted CNL gene losses and duplications in the genomes of wheat and its progenitors. Species are arranged in order of their phylogenetic relationships, and gene duplication losses were determined using the program Notung 2.9.

Table III.1 Number of wheat homologs of various characterized R-proteins as identified using BLAST search. Conferral of pathogen resistance was noted in the second column if the pathogen was known and if the R-protein plays a major role in resistance to the pathogen. The number of wheat homologs is out of the total 609 wheat CNL genes identified.

Homolog	Pathogen (if known)	Wheat Homologs
RPM1	*Pseudomonas syringae*	198
RPP13	*Peronospora parasitica*	88
RGA4	*Magnaporthe oryzae*	38
RGA3	--	37
RGA2	*Phytophthora infestans*	33
RPP8	*Peronospora parasitica*	22
RGA1	--	21
NBS2-RDG2A	*Pyrenophora graminea*	7
RPS2	*Pseudomonas syringae*	6
RXW24L	--	5
SKIP11	--	4
APAF1	*NA: apoptosis factor*	3
Pm3b	*Blumeria graminis f. sp. tritici*	3
PM	*Blumeria graminis f. sp. tritici*	3
R1B-14	*Phytophthora infestans*	3
RDL5/RF45	--	2
R1A-6	*Phytophthora infestans*	2
RDG2A	*Pyrenophora graminea*	2
RPS5	*Pseudomonas syringae*	2
MLA	*Blumeria graminis f. sp. tritici*	1
NRC1b	--	1

REFERENCES

1. McFadden E. 2259461. A successful transfer of emmer characters to Vulgare wheat. *Agronomy journal.* 1930;22(12):1020-1034.
2. Spielmeyer W, Sharp P, Lagudah E. Identification and Validation of Markers Linked to Broad-Spectrum Stem Rust Resistance Gene in Wheat (L.). *Crop Science.* 2003;43(1):333-336.
3. Takken FL, Goverse A. How to build a pathogen detector: structural basis of NB-LRR function. *Current opinion in plant biology.* 2012;15(4):375-384.
4. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of NBS-LRR–encoding genes in Arabidopsis. *The Plant Cell Online.* 2003;15(4):809-834.
5. Periyannan S, Moore J, Ayliffe M, et al. The gene Sr33, an ortholog of barley Mla genes, encodes resistance to wheat stem rust race Ug99. *Science.* 2013;341(6147):786-788.
6. Saintenac C, Zhang W, Salcedo A, et al. Identification of wheat gene Sr35 that confers resistance to Ug99 stem rust race group. *Science.* 2013;341(6147):783-786.
7. Brenchley R, Spannagl M, Pfeifer M, et al. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature.* 2012;491(7426):705-710.
8. Jia J, Zhao S, Kong X, et al. Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature.* 2013;496(7443):91-95.
9. Ling H-Q, Zhao S, Liu D, et al. Draft genome of the wheat A-genome progenitor Triticum urartu. *Nature.* 2013;496(7443):87-90.
10. Kariyawasam GK, Carter AH, Rasmussen JB, et al. Genetic relationships between race-nonspecific and race-specific interactions in the wheat–Pyrenophora tritici-repentis pathosystem. *Theoretical and Applied Genetics.* 2016;129(5):897-908.