A Comprehensive Analysis of Rust Disease Resistance in the Bioenergy Plant Switchgrass (*Panicum virgatum* L.)

Taylor Price Frazier

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Horticulture

Bingyu Zhao, Chair
Mohammad A. Saghai-Maroof
David G. Schmale
Richard E. Veilleux

December 1, 2015
Blacksburg, VA

Keywords:
Switchgrass, *Panicum virgatum*, Biofuel, Switchgrass Rust, Disease Resistance, NB-LRR, Gene Expression, SNPs, RNA-sequencing, Molecular Marker, SHN3, Lignin, Cellulose

A Comprehensive Analysis of Rust Disease Resistance in the Bioenergy Plant Switchgrass (*Panicum virgatum* L.)

Taylor Price Frazier

ABSTRACT

Switchgrass is a C_4 perennial grass that is currently being developed for use as a second generation lignocellulosic biofuel crop. For switchgrass to be fully utilized as a bioenergy crop, large-scale plantings of elite switchgrass germplasm, possibly in monoculture, are likely to occur. This practice may increase the selection pressure on plant pathogens, such as switchgrass rust, which could result in devastating disease epidemics. The identification and deployment of quantitative trait loci (QTLs) and major plant disease resistance genes (*R*) in switchgrass breeding programs could offer broad spectrum and durable disease resistance in commercial switchgrass cultivars. 'Alamo', a lowland cultivar, is generally resistant to switchgrass rust whereas 'Dacotah', an upland cultivar, is highly susceptible. I hypothesized that major *R* genes and/or QTLs were contributing to the differences in disease phenotypes of these two cultivars. In this dissertation, bioinformatics and molecular biology approaches were employed to dissect the genetic mechanisms underlying switchgrass rust disease resistance. Novel pseudo-F_2 mapping populations were created from a cross derived from 'Alamo' and 'Dacotah'. RNA-sequencing of the pseudo-F_2 progenies of 'Alamo' x 'Dacotah' was used to construct a genetic linkage map and to identify potential QTLs correlating with disease resistance. In addition, a homology-based computational method was used to identify

1,011 potential NB-LRR *R* genes in the switchgrass genome (v 1.1). These potential *R* genes were characterized for polymorphism and expression differences between 'Alamo' and 'Dacotah'. Moreover, I found that some NB-LRR genes are developmentally regulated in switchgrass. One of the major objectives of switchgrass breeding programs is to develop cultivars with improved feedstock quality; however, changes in the components of the plant cell wall may affect disease resistance. I hypothesized that genetically modified switchgrass plants with altered cell wall components will respond differently than the wild-type to switchgrass rust. Transgenic switchgrass plants overexpressing *AtSHN3*, a transcription factor with known functions in epicuticular wax accumulation and cell wall deposition, were created. I found that *AtSHN3*-overexpressing transgenic switchgrass lines were more susceptible than wild-type plants in their response to switchgrass rust. Overall, the results of this dissertation provide a platform for elucidating the molecular mechanisms underlying resistance of switchgrass to switchgrass rust. These findings will help breeders create switchgrass cultivars with improved disease resistance, and will ultimately allow switchgrass to be used for sustainable biomass production.

ACKNOWLEDGEMENTS

I would like to thank my adviser, Bingyu Zhao, for his support and guidance. I would also like to thank my committee members, Saghai Maroof, David Schmale, and Richard Veilleux, for their helpful suggestions during my dissertation research. Additionally, I would like to acknowledge and thank the members of the Zhao lab, the Department of Horticulture, and the members of the Translational Plant Sciences program at Virginia Tech. Without their encouragement, motivation, and friendships, this dissertation would not have been possible. Finally, I would also like to acknowledge and thank my parents, Edward and Myra Frazier, as well as my fiancé, Chandler Douglas. Your constant love and support enabled me to pursue my doctoral degree and for that I am forever grateful.

FUNDING

I would like to acknowledge the Department of Horticulture and the Translational Plant Sciences Program for financial support for my graduate education.

I would like to acknowledge the following sources of support for this research:

Chapter 2: This work was supported by a DOE-USDA grant (DE-FG02-08ER64629) awarded to Jan Leach, Bingyu Zhao, John McKay, Daniel Bush, Hei Leung, and Andrew Kern. This work was also supported by a DOE-USDA grant (DE-SC0008338) awarded to Kevin Childs, C. Robin Buell, Bingyu Zhao, and Xunzhong Zhang.

Chapter 3: This project was supported by USDA-NIFA Grant Number 2011-67009-30133 to Bingyu Zhao, Brett Tyler, Stephen Marek, Carla Garzon, and Bing Yang. This work was also supported in part by U. S. Department of Energy Grant (DE-AI02-09ER64829), USDA-NIFA Grant (2011-67009-30096), and by the USDA-ARS CRIS project (3042-21000-030-00D), all awarded to Gautam Sarath and Christian Tobias. In addition, this work was supported by a DOE-USDA grant (DE-FG02-08ER64629) awarded to Jan Leach, Bingyu Zhao, John McKay, Daniel Bush, Hei Leung, and Andrew Kern as well as a DOE-USDA grant (DE-SC0008338) awarded to Kevin Childs, C. Robin Buell, Bingyu Zhao, and Xunzhong Zhang. Finally, a portion of this work was supported by the Office of Science of the U.S. Department of Energy under contract number DE-AC02-05CH11231.

Chapter 4: This work was supported by a DOE-USDA grant (DE-FG02-08ER64629) awarded to Jan Leach, Bingyu Zhao, John McKay, Daniel Bush, Hei Leung, and Andrew Kern. This project was also supported by USDA-NIFA Grant Number 2011-67009-30133 to Bingyu Zhao, Brett Tyler, Stephen Marek, Carla Garzon, and Bing Yang, and by a Virginia Tech CALS integrative grant to Bingyu Zhao.

Chapter 5: This work was supported by an integrated, internal competitive grant from the College of Agriculture and Life Sciences at Virginia Tech. This work was also supported by the Institute for Critical Technology and Applied Science at Virginia Tech, the Biodesign and Bioprocessing Research Center of the College of Agriculture and Life Sciences at Virginia Tech, and the Virginia Agricultural Experimental Station (VA135872).

ATTRIBUTIONS

Several colleagues contributed to the writing and research in several manuscripts presented as chapters in this dissertation.

Chapter 2: "Exploring Differences in the Transcriptomes of the Switchgrass Cultivars 'Dacotah' and 'Alamo'"

Daniel Hupalo, PhD (Kern lab, Department of Genetics, Rutgers University) is currently a postdoctoral researcher at New York University. Dr. Hupalo is a co-author on the paper, created the gene models of 'Alamo' and 'Dacotah', performed the variant detection, and helped write the manuscript.

Yiming Liu, PhD (Zhao lab, Department of Horticulture) is currently a postdoctoral researcher at Virginia Tech. Dr. Liu is a co-author on the paper and contributed to the molecular techniques used in this chapter.

Jiamin Miao, PhD (Zhao lab, Department of Horticulture) is currently a visiting scientist at Virginia Tech. Dr. Miao is a co-author on the paper and contributed to the molecular techniques used in this chapter.

Xunzhong Zhang, PhD (Zhang lab, Department of Crop and Soil Environmental Sciences) is currently an advanced Research Scientist at Virginia Tech. Dr. Zhang is a co-author on the paper, contributed to the molecular techniques used in this chapter, and is a co-PI for the grant supporting this work.

Bingyu Zhao, PhD (Zhao lab, Department of Horticulture) is currently an associate professor at Virginia Tech. Dr. Zhao is a co-author on the paper and is a co-PI for two of the grants supporting this work.

Jeongwoon Kim, PhD (Childs lab, Department of Plant Biology, Michigan State University) is currently a visiting research associate at Michigan State University. Dr. Kim is a co-author on this paper and contributed to the molecular techniques used in this chapter.

Kevin Childs, PhD (Childs lab, Department of Plant Biology, Michigan State University) is currently an assistant professor at Michigan State University. Dr. Childs is a co-author on this paper and a principle investigator of a grant supporting this work.

Daniel Bush, PhD (Department of Biology, Colorado State University) is currently a professor and Vice Provost for Faculty Affairs at Colorado State University. Dr. Bush is a co-author on this paper and a co-PI of a grant supporting this work.

John McKay, PhD (McKay lab, Department of Bioagricultural Sciences and Pest Management, Colorado State University) is currently an associate professor at Colorado State University. Dr. McKay is a co-author on this paper and a co-PI of a grant supporting this work.

Jan Leach, PhD (Leach lab, Department of Bioagricultural Sciences and Pest Management, Colorado State University) is currently the Associate Dean for Research in the College of Agriculture at Colorado State and an adjunct scientist at the International Rice Research Institute (Philippines). Dr. Leach is a co-author on this paper and the principle investigator of a grant supporting this work.

Andrew D. Kern, PhD (Kern lab, Department of Genetics, Rutgers University) is currently an assistant professor at Rutgers University. Dr. Kern is a co-author on this paper and a co-PI of a grant supporting this work.

Chapter 3: "Identification, characterization, and gene expression analysis of Nucleotide Binding site (NB)-type resistance gene homologues in switchgrass"

Nathan Palmer, PhD (Sarath lab, Grain, Forage and Bioenergy Research Unit, USDA-ARS) is currently a biological sciences lab technician with the USDA-ARS. Dr. Palmer is a co-author on this paper, contributed to the molecular techniques used in this chapter, and co-wrote the manuscript.

Fuliang Xie, PhD (Zhang lab, Department of Biology, East Carolina University) is currently a senior data scientist at OmicSoft Corporation. Dr. Xie is a co-author on this paper and contributed to the molecular techniques used in this chapter.

Christian Tobias, PhD (Tobias lab, Crop Improvement and Genetics Research Unit, USDA-ARS) is currently a research molecular biologist with the USDA-ARS. Dr. Tobias is a co-author on this paper, contributed to the molecular techniques used in this chapter, and is co-PI of a grant that supported this work.

Teresa Donze-Reiner, PhD (Donze-Reiner lab, Department of Biology, West Chester University of Pennsylvania) is currently an assistant professor at West Chester University of Pennsylvania. Dr. Donze-Reiner is a co-author on this paper and contributed to the molecular techniques used in this chapter.

Aureliano Bombarely, PhD (Bombarely lab, Department of Horticulture) is currently an assistant professor at Virginia Tech. Dr. Bombarely is a co-author on this paper and contributed to the molecular techniques used in this chapter.

Kevin Childs, PhD (Childs lab, Department of Plant Biology, Michigan State University) is currently an assistant professor at Michigan State University. Dr. Childs is a co-author on this paper and a principle investigator of a grant supporting this work.

Shengqiang Shu, PhD (DOE-JGI) is currently a software engineer at DOE-JGI. Dr. Shu is a co-author on this paper and contributed to the sequencing and assembly of the switchgrass reference genome used in this chapter.

Jerry Jenkins, PhD (Genome Sequencing Center, HudsonAlpha Institute for Biotechnology) is currently a genome analysis group leader at the HudsonAlpha Institute for Biotechnology. Dr. Jenkins is a co-author on this paper and contributed to the sequencing and assembly of the switchgrass reference genome used in this chapter.

Jeremy Schmutz (DOE-JGI, HudsonAlpha Institute for Biotechnology) is currently the Plant Genome Lead at the DOE-JGI and HudsonAlpha Institute for Biotechnology. Mr. Schmutz is a co-author on this paper and contributed to the sequencing and assembly of the switchgrass reference genome used in this chapter.

Baohong Zhang, PhD (Zhang lab, Department of Biology, East Carolina University) is currently an associate professor at East Carolina University. Dr Zhang is a co-author on this paper and contributed to the experimental design of this project.

Gautam Sarath, PhD (Sarath lab, Grain, Forage and Bioenergy Research Unit, USDA-ARS) is currently a research molecular biologist with the USDA-ARS. Dr. Sarath is a co-author on this paper, a co-PI of a grant supporting this work, and co-wrote the manuscript.

Bingyu Zhao, PhD (Zhao lab, Department of Horticulture) is currently an associate professor at Virginia Tech. Dr. Zhao is a co-author on the paper and is a principle investigator of a grant supporting this work.

Chapter 4: "SNP-based linkage map construction from RNA-sequencing and identification of a QTL associated with disease resistance in the bioenergy crop switchgrass"

Gabriela Orquera Tornakian, MS (Marek lab, Department of Entomology and Plant Pathology, Oklahoma State University) is currently a graduate student at Oklahoma State University. Ms. Tornakian is a co-author on this paper and isolated the VT2-1 rust strain used in this chapter.

Aureliano Bombarely, PhD (Bombarely lab, Department of Horticulture) is currently an assistant professor at Virginia Tech. Dr. Bombarely is a co-author on this paper and contributed to the molecular techniques used in this chapter.

Kerri Mills, MS (Zhao lab, Department of Horticulture) is currently a senior laboratory specialist at Virginia Tech. Ms. Mills is a co-author on this paper and contributed to creation of the mapping populations used in this chapter.

Jiamin Miao, PhD (Zhao lab, Department of Horticulture) is currently a visiting scientist at Virginia Tech. Dr. Miao is a co-author on the paper and contributed to creation of the mapping populations used in this chapter.

Caitlin Burklew, MS (Zhao lab, Department of Horticulture) is currently a senior research associate at Novozymes. Ms. Burklew is a co-author on this paper and contributed to the molecular techniques used in this chapter.

Carlos Toros (Zhao lab, Department of Horticulture) is currently an undergraduate research assistant at Virginia Tech. Mr. Toros is a co-author on this paper and counted the number of pustules on infected switchgrass leaves.

Jan Leach, PhD (Leach lab, Department of Bioagricultural Sciences and Pest Management, Colorado State University) is currently the Associate Dean for Research in the College of Agriculture at Colorado State and an adjunct scientist at the International Rice Research Institute (Philippines). Dr. Leach is a co-author on this paper and the principle investigator of a grant supporting this work.

Bingyu Zhao, PhD (Zhao lab, Department of Horticulture) is currently an associate professor at Virginia Tech. Dr. Zhao is a co-author on the paper and the principle investigator of a grant supporting this work.

Chapter 5: "Overexpression of the Arabidopsis SHN3 transcription factor affects the cell wall composition and disease resistance of transgenic switchgrass plants"

Guigui Wan, MS (Frazier lab, Department of Sustainable Biomaterials) is currently a graduate student at Virginia Tech. Ms. Wan is a co-author of this paper and contributed to the solvent-submersion tensile-torsion analysis in this chapter

Ann Norris, PhD (Frazier lab, Department of Sustainable Biomaterials) is currently a lab technician at Virginia Tech. Dr. Norris is a co-author of this paper and contributed to the lignin measurement and solvent-submersion tensile-torsion analysis in this chapter

Fan Lin, (Bartley lab, Department of Microbiology and Plant Biology, University of Oklahoma) is currently a graduate student at the University of Oklahoma. Mr. Lin is a co-author on this paper and contributed to the cell wall component analysis in this chapter

Carlos Toros (Zhao lab, Department of Horticulture) is currently an undergraduate research assistant at Virginia Tech. Mr. Toros is a co-author on this paper and contributed to the agronomic trait measurements collected in this chapter.

Laura Bartley, PhD (Bartley lab, Department of Microbiology and Plant Biology, University of Oklahoma) is currently an assistant professor at the University of Oklahoma. Dr. Bartley is a co-author on this paper and contributed to the cell wall component analysis in this chapter.

Chip Frazier, PhD (Frazier lab, Department of Sustainable Biomaterials) is currently a professor at Virginia Tech and the director of the Wood-Based Composites Center. Dr. Frazier is a co-author on this paper and contributed to the experimental design of this project.

Bingyu Zhao, PhD (Zhao lab, Department of Horticulture) is currently an associate professor at Virginia Tech. Dr. Zhao is a co-author on the paper and the principle investigator of a grant supporting this work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLE OF ABBREVIATIONS

BLAST	Basic Local Alignment Search Tool
bp	Base Pair
CC	Coiled-coil
cDNA	Complementary Deoxyribose Nucleic Aid
cM	Centimorgan
cm	Centimeter
DNA	Deoxyribose Nucleic Acid
DOE-JGI	Department of Energy- Joint Genome Institute
dpi	Days post inoculation
ELF1α	Elongation Factor 1 alpha
EST	Expressed Sequence Tag
EST-SSR	Expressed Sequence Tag- Simple Sequence Repeat
ETI	Effector Triggered Immunity
GO	Gene Ontology
GST	Glutathione-S-Transferase
ha	Hectare
HRGP	Hydroxyproline-rich Glycoprotein
LRR	Leucine Rich Repeat
Mbp	Mega base pairs
Mg	Megagram
NB	Nucleotide Binding Site
NB-LRR	Nucleotide Binding Site- Leucine Rich Repeat
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
pg	Picogram
PAMP	Pathogen Associated Molecular Pattern
PCR	Polymerase Chain Reaction
PTI	Pathogen Associated Molecular Pattern-Triggered Immunity
QTL	Quantitative Trait Locus
RAPD	Random Amplified Polymorphic DNA
R genes	Plant Disease Resistance Genes
RGH	Resistance Gene Homolog
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribose Nucleic Acid
RNAi	Ribose Nucleic Acid Interference
RNA-Seq	Ribose Nucleic Acid- Sequencing
RPKM	Reads Per Kilobase Per Million of Reads Mapped
RT-PCR	Real Time- Polymerase Chain Reaction
SAM	S-adenosylmethionine
SNP	Single Nucleotide Polymorphism
SRAP	Sequence Related Amplified Polymorphism
SSR	Simple Sequence Repeat

TCA Cycle	Tricarboxylic Acid Cycle
TIR	*Drosophila* Toll and Mammalian Interleukin-1 Receptor
U.S.	United States
USDA	United States Department of Agriculture
US-DOE	United States Department of Energy

CHAPTER 1: Literature Review

In the early 1990s, the United States Department of Energy (US-DOE) designated switchgrass as a potential bioenergy feedstock (McLaughlin & Kszos, 2005). Since then, the US-DOE has initiated research programs that have focused on developing this plant species for use in biofuel production (Sanderson *et al.*, 1996). Switchgrass (*Panicum virgatum*) is a warm-season, perennial grass that is native to the prairies of North America. There are several physical and physiological characteristics of switchgrass that make it an ideal candidate for a second generation biofuel feedstock. As a C4 photosynthetic plant, switchgrass is highly efficient at converting light energy and concentrating carbon dioxide into biomass production (Lewandowski *et al.*, 2003). Some switchgrass cultivars can grow upwards of 3 meters in height and can produce root stocks that extend the same length below ground (Weaver, 1968). Additionally, switchgrass requires minimal fertilizer input and can endure multiple harvests per year (Thomason *et al.*, 2005). The amount of biomass that can be produced from established switchgrass stands depends on numerous factors including switchgrass genotype, environmental conditions, nutrient input, and harvest frequency (Lemus *et al.*, 2002; McLaughlin & Kszos, 2005; Thomason *et al.*, 2005). When grown for use as a biofuel, switchgrass can produce on average between 10 and 20 Mg ha^{-1} of biomass per year (McLaughlin & Kszos, 2005).

As a readily out-crossing prairie grass, switchgrass has two major ecotypes that have originated from different habitats. The lowland ecotypes are generally adapted to southern, warm, moist habitats, while the upland ecotypes are better adapted to the northern U.S. and southern Canada. Lowland ecotypes have been shown to have wider stems, coarser leaves, and

taller tillers than their upland counterparts (Porter, 1966). Due to these characteristics, lowland varieties have been found to produce on average three times more biomass than the upland varieties (Cassida *et al.*, 2005). Since upland ecotypes typically grow in the northern regions of North America, they are generally considered to be more tolerant to drought and cold abiotic stresses; however, a recent study found that lowland ecotypes may actually be more tolerant to drought (Liu *et al.*, 2015). Lowland ecotypes have also been shown to be more resistant to diseases whereas the upland ecotypes are more susceptible (Uppalapati *et al.*, 2013).

After switchgrass was identified as a potential bioenergy crop, breeding programs were established to develop elite cultivars with enhanced agronomic traits such as increased biomass production (Taliaferro & Das, 2002a) and improved forage quality (Hopkins *et al.*, 1993). In order to concentrate breeding efforts and quickly identify desirable germplasm, several studies analyzed physiological parameters that could be positive indicators of switchgrass fitness. These included photosynthetic rates (Wullschleger *et al.*, 1996), transpiration efficiency (Byrd & May, 2000), and cold hardiness (Hope & McElroy, 1990).

Cytogenetic studies in switchgrass have aimed to estimate the chromosome number and nuclear size of existing switchgrass populations. Initial studies identified that there are nine chromosomes in this species (Lu *et al.*, 1998). As early as 1944, Nielson reported observing switchgrass accessions with ploidy numbers ranging from diploid (2n=2x=18) to duodecaploid (2n=12x=108) (Nielson, 1944). While all lowland ecotypes seem to be tetraploids (2n=4x=36), upland ecotypes can vary in their ploidy level with aneuploidy being a common phenomenon (Brunken & Estes, 1975; Riley & Vogel, 1982). Tetraploid, hexaploid (2n=6x=54), and octaploid (2n=8x=72) varieties of upland ecotypes have consistently been reported (Riley & Vogel, 1982). Using flow cytometry, the DNA content of lowland ecotypes was estimated to be 3.1 pg/nuclei

(Lu *et al.*, 1998). Correspondingly, the DNA content of octaploid upland ecotypes was estimated to be between 5.2 and 6.1 pg/nuclei (Hopkins *et al.*, 1996; Lu *et al.*, 1998).

The self-incompatibility system that is present in switchgrass is believed to resemble the S-Z system that has previously been described in grasses (Langridge & Baumann, 2008). As such, the outcrossing rate of switchgrass has been reported to be as high as 99.96% (Liu *et al.*, 2014). Thus, only a small percentage of selfed seeds (0.35%-1.39%) have been observed in naturally occurring switchgrass populations (Martinez-Reyna & Vogel, 2002). Recently, however, a report identified a lowland genotype of switchgrass with high self-compatibility (Liu & Wu, 2012). This particular genotype could be useful for developing inbred lines. Regardless of the ecotype, biological crosses of switchgrass are possible if the plants contain the same ploidy level (Martínez-Reyna *et al.*, 2001; Martinez-Reyna & Vogel, 2002; Taliaferro & Das, 2002b). In addition, synthetic polyploids of switchgrass with doubled chromosomes have been shown to breed with germplasm of their new-found ploidy (Yang *et al.*, 2014).

Since switchgrass is highly self-incompatible, much heterozygosity exists in the genetic backgrounds of native switchgrass germplasm. Early genomic studies in switchgrass focused on identifying molecular markers that could be used to construct genetic linkage maps, as well as assess the variation within and among existing switchgrass populations. Gunter *et al* (1996) used five RAPD markers to assess the genetic diversity among 14 upland and lowland switchgrass populations. In this study, the authors found that these switchgrass populations were on average 65% similar to each other while members within each of the populations were approximately 81% similar (Gunter *et al.*, 1996). Another study used RFLP markers to analyze genetic variation within and between three switchgrass cultivars (‘Alamo’, ‘Kanlow’, and ‘Summer’) that are commonly used in breeding programs (Missaoui *et al.*, 2006). Interestingly, more polymorphism

was found within the 'Summer' genotypes (64%), which is an upland ecotype, in comparison to 'Kanlow' (52%) and 'Alamo' (60%), which are lowland ecotypes (Missaoui *et al.*, 2006). In addition to genomic DNA, RFLP markers for switchgrass chloroplast DNA have also been used to determine the genetic relationship between the different ecotypes (Hultquist *et al.*, 1996). Indeed, the results of that study showed that the chloroplast DNA varied between lowland and upland switchgrass (Hultquist *et al.*, 1996). Several other types of molecular markers have also been investigated for use in switchgrass molecular breeding including EST-SSRs (Tobias *et al.*, 2006; Narasimhamoorthy *et al.*, 2008), SSRs (Wang *et al.*, 2011), and SRAPs (Huang *et al.*, 2011).

Focus has shifted from the phenotypic evaluation and characterization of existing switchgrass germplasms to molecular-based studies aimed at elucidating the genetic mechanisms underlying important agronomic traits. The establishment of Next Generation Sequencing (NGS) techniques, along with a reduction in the cost of various NGS sequencing platforms, has produced valuable genomic resources for switchgrass. Recently, a genome sequence for switchgrass (v 1.1) has become available (Goodstein *et al.*, 2012). This sequence is based on the 'AP13' variety of the cultivar 'Alamo', which is a tetraploid (Goodstein *et al.*, 2012). The genome is comprised of 1,230 Mbp of sequencing data located in 319,670 contigs. In an attempt to construct the two linkage groups of tetraploid switchgrass, 99,024 of these contigs were assigned to one of the 18 main chromosomal pseudomolecules, which represent the A and B subgenomes. Additional genomic resources have also become available including a BAC library with 147,456 clones (Saski *et al.*, 2011), an Affymetrix cDNA microarray chip (Zhang *et al.*, 2013), and numerous RNA-sequencing datasets (Wang *et al.*, 2012b; Zhang *et al.*, 2013; Serba *et al.*, 2015).

In addition to traditional breeding, molecular biology and biotechnology techniques have been used to genetically modify switchgrass with desirable agronomic traits. Plant tissue culture methods have been established that allow for the production of embryogenic callus from seeds (Denchev & Conger, 1994; Li & Qu, 2011) and leaf tissue (Denchev & Conger, 1994). A switchgrass line of the 'Alamo' variety, termed HR8, was created by recurrent tissue selection and has been shown to produce more somatic embryogenic callus in comparison to other unselected 'Alamo' types (Xu *et al.*, 2011b). Additional reports have shown that embryogenic switchgrass tissue can be genetically transformed using particle bombardment (Richards *et al.*, 2001) and *Agrobacterium tumefaciens* (Li & Qu, 2011; Xu *et al.*, 2011b).

Initial research at creating transgenic switchgrass inserted a construct carrying the Basta tolerance *bar* gene into embryogenic callus (Richards *et al.*, 2001). Since then, numerous reports have been published about the creation of transgenic switchgrass. RNAi-based strategies have been used to knock down key components of the lignin pathway in an effort to reduce lignin content (Fu *et al.*, 2011a; Xu *et al.*, 2011a). Overexpression of the Arabidopsis NAC transcription factor *LONGER VEGETATIVE PHASE1* led to a change in leaf architecture, increased lignin content, and improved drought tolerance in transgenic switchgrass (Xu *et al.*, 2012). Switchgrass has also been used as a biofactory for the production of value-added traits such as plastics (Somleva *et al.*, 2008).

In order to optimize switchgrass biomass output for biofuel production, desired switchgrass germplasms will likely be planted and grown in monoculture. This has the potential to significantly increase the disease pressure on switchgrass fields and could lead to devastating epidemic outbreaks. To date, several diseases have been reported on switchgrass. A novel member of the *Colletotrichum* genus (*Colletotrichum navitas*) was recently identified as the

cause of Anthracnose disease on switchgrass (Crouch *et al.*, 2009). Additionally, switchgrass panicle seedheads have been shown to be infected by *Tilletia maclaganii* (Berk.) G.P. Clinton and *T. pulcherrima* Ellis & Galloway, which produces smut and bunt, respectively (Gravert *et al.*, 2000; Carris *et al.*, 2008). Thomsen *et al* (2008) found that even a 26% incidence of naturally occurring smut infections in switchgrass fields in Iowa had the ability to reduce switchgrass yield by approximately 17%. Viruses have also been identified to cause disease on switchgrass. These include *Panicum mosaic virus* (Sill & Pickett, 1957), *Barley yellow dwarf virus* (Garrett *et al.*, 2004), and *Switchgrass mosaic virus* (Agindotan *et al.*, 2013). In 2004, Krupinsky *et al* identified *Bipolaris oryzae* as the causal agent of leaf spot on switchgrass (Krupinsky *et al.*, 2004).

One of the most wide spread and devastating disease problems of switchgrass, and the focus of this dissertation, is switchgrass rust. Rusts are foliar fungal pathogens that mainly colonize on the adaxial surface of leaves. Presently, three rust pathogens have been described on switchgrass: *Puccinia graminis*, *Puccinia emaculata*, and *Uromyces graminicola* Burrill (Uppalapati *et al.*, 2013). Of these rust pathogens, *P. emaculata* is the most common and has been reported in Tennessee (Zale *et al.*, 2008), Arkansas (Hirsch *et al.*, 2010), and Virginia (Frazier *et al.*, 2012). *P. emaculata* is a basidiomycete and an obligate plant pathogen. The lifecycle of *P. emaculata* is presumed to follow that of other rust pathogens. Presently, there has been no identification of an alternate host. Infection of switchgrass by switchgrass rust has been shown to result in substantial biomass yield losses (Sykes *et al.*, 2015). *Brachypodium distachyon*, a model organism for the grasses, is currently being investigated as a non-host model for elucidating resistance to *P. emaculata* (Gill *et al.*, 2015).

Plants are stationary organisms that are constantly exposed to microbial and fungal plant pathogens. As such, plants have evolved defense mechanisms to recognize and ward off potential

invaders. The plant innate immune system is comprised of two levels of resistance. The first level involves the recognition of pathogen-associated molecular patterns (PAMPs) by cell surface pattern-recognition receptors and subsequent initiation of PAMP-triggered immunity, or PTI (Jones & Dangl, 2006). Defense responses in PTI include callose deposition at the plant cell wall, initiation of MAP kinase signaling cascades, induction of pathogen-responsive gene expression, and rapid production of reactive oxygen species (Nürnberger *et al.*, 2004).

The second level depends on either the direct or indirect recognition of secreted pathogen effector molecules by plant proteins and downstream activation of defense responses in what has been termed Effector-triggered immunity, or ETI (Jones & Dangl, 2006). In ETI, invading pathogens secrete effector molecules that enter into host plant cells and work to manipulate or alter plant defense responses (Boller & He, 2009). Consequently, plants have evolved Resistance (*R*) genes that encode for proteins that can directly or indirectly recognize pathogen effectors and initiate downstream signaling pathways (Hammond-Kosack & Jones, 1997). The culmination of these pathways is usually a hypersensitive response (HR), which is a type of programmed cell death that prevents pathogen invasion and proliferation (Heath, 2000).

The identification and validation of *R* genes is a main focus in molecular plant-pathogen interactions as researchers aim to create plant cultivars with improved disease resistance. *R* genes that function in ETI belong to a highly conserved family of nucleotide binding site- leucine rich repeat (NBS-LRR) genes (Meyers *et al.*, 1999). These genes encode for proteins that contain a C-terminal LRR domain, a central NBS domain, and a variable N-terminal domain (Meyers *et al.*, 1999). Within the large NBS-LRR gene family, there are two main classes that are categorized by the type of domain encoded in the N-terminal region. One class has an N-terminal domain that is homologous to the intracellular signaling domain of Drosophila toll and

mammalian interleukin-1 receptors. This class is known as TIR-NBS-LRR genes, or TNL genes for short (Meyers *et al.*, 1999). The second class of NBS-LRR genes encodes an N-terminal coiled-coil domain and is subsequently referred to as CC-NBS-LRR genes, or CNL genes for short (Meyers *et al.*, 1999). TNL genes are predominantly found in dicots but are presumably absent in monocot genomes.

Presently, 16,884 putative *R* genes have been identified by experimental and computational methods and have been deposited in the Plant Resistance Gene Database (PRGDB). Of these potential *R* genes, only 113 have been manually curated and their functions are currently being investigated in numerous plant species. Despite recent advances in identifying potential *R* genes, only one study has been performed in switchgrass. Zhu *et al* (2013) used a combination method of PCR and EST database mining to identify 295 potential NB-LRR genes in switchgrass (Zhu *et al.*, 2013). This is only 0.3% of the protein coding genes in the switchgrass genome (v 1.1) (Goodstein *et al.*, 2012). It is estimated that the percentage of *R* genes in plant genomes ranges from 0.6% to 1.76% (Porter *et al.*, 2009). Thus, in order to fully determine the defense gene arsenal present in switchgrass, a more comprehensive analysis of *R* genes is needed.

Despite its importance as an emerging bioenergy crop, a comprehensive analysis of rust disease resistance has not yet been performed in switchgrass and thus, the genetic mechanisms underlying tolerance to disease remains poorly understood in this plant species. This dissertation will focus on the following specific objectives:

Objective 1. Explore differences in the transcriptomes of the switchgrass cultivars 'Dacotah' and 'Alamo' (Chapter 2).

‘Alamo’ is a lowland tetraploid cultivar of switchgrass that produces copious amounts of biomass and is relatively resistant to switchgrass rust. ‘Dacotah’ is an upland tetraploid cultivar of switchgrass that is diminutive in stature and is highly susceptible to switchgrass rust***. I hypothesized that the observed phenotypic variation between these two cultivars is due in part to differences in gene expression.*** RNA-sequencing reads, obtained from leaf tissue cDNA from ‘Alamo’ and ‘Dacotah’, were mapped to the switchgrass reference genome (v 0.0) Gene models were created and Gene Set Enrichment Analysis (GSEA) was performed to investigate the genetic mechanisms in both cultivars that may be contributing to their vastly different phenotypes. In addition, SNPs were identified and analyzed as a function of read depth to determine the extent of polymorphism captured by this sequencing effort.

Objective 2. Identify and characterize switchgrass NB-LRR containing plant disease resistance (*R*) genes (Chapter 3).

The largest class of *R* genes belongs to those of the NB-LRR variety; however, no study has identified NB-LRR genes in switchgrass on a whole genome level. NB-LRR genes have been shown to make up anywhere from 0.5% to 1.8% of the total protein coding genes in plants. Since the switchgrass genome contains approximately 97,000 protein coding genes, ***I hypothesized that approximately 1,000 NB-LRR genes are present in switchgrass and that several of these genes contribute to disease resistance.*** In this chapter, a homology-based computational approach was used to identify and characterize potential NB-LRR disease resistance genes in switchgrass. RNA-sequencing data from ‘Alamo’ and ‘Dacotah’ were used to identify SNPs in the NB-LRR genes, as well as determine gene expression differences in uninoculated leaf tissue, between the two cultivars. Finally, RNA-sequencing datasets from the

flag leaves of field-grown cv ‘Summer’ plants found that some of the NB-LRR genes identified in this study are developmentally regulated.

Objective 3. Evaluate a collection of switchgrass germplasm in response to rust isolate VT2-1 (Chapter 4).

‘Alamo’ and ‘Dacotah’ vary significantly in their response to switchgrass rust with ‘Alamo’ being resistant to rust and ‘Dacotah’ being highly susceptible. ***I hypothesized that one or more QTLs may be contributing to the different disease phenotypes of ‘Alamo’ and ‘Dacotah’.*** In this objective, novel pseudo-F_2 mapping populations were created that were derived from an initial cross of ‘Alamo’ and ‘Dacotah’. RNA-sequencing data from un-inoculated leaf tissue of 154 F_2 progenies were used to analyze the segregation patterns of SNPs between the two cultivars and to construct a genetic linkage map. Disease assays for resistance to switchgrass rust were performed on three biological replicates of each plant in the F_2 population. The disease phenotype data, along with the constructed genetic linkage map, were used to identify potential QTLs that could be contributing to disease resistance in switchgrass.

Objective 4. Transform switchgrass with *AtSHN3* and determine the effects of this transgene on the cell wall components (Chapter 5).

One of the main goals of switchgrass breeding programs is to release cultivars with improved feedstock quality. Switchgrass lines with plant cell walls that have reduced levels of lignin and higher levels of cellulose are ideal for the production of bioethanol. The plant cell wall, however, is the first physical barrier to foliar plant pathogens, such as switchgrass rust. Therefore, ***I hypothesized that alterations in the lignin and cellulose content of switchgrass cell***

walls will affect disease response. In this chapter, *Agrobacterium tumefaciens* was employed to transform embryogenic switchgrass callus with the *AtSHN3* cDNA sequence under the control of the constitutive maize ubiquitin promoter. Four transgenic switchgrass lines were compared to wild-type plants for changes in important agronomic traits (tiller height, biomass, etc.) and differences in lignin and cellulose content. The transgenic plants were also analyzed for altered stem stiffness and mechanical strength, as well as disease response to switchgrass rust.

REFERENCES

Agindotan B, Okanu N, Oladeinde A, Voigt T, Long S, Gray M, Bradley C. 2013. Detection of Switchgrass mosaic virus in Miscanthus and other grasses. *Can. J. Plant Pathol.* **35**: 81-86.

Boller T, He SY. 2009. Innate immunity in plants: an arms race between pattern recognition receptors in plants and effectors in microbial pathogens. *Science* **324**: 742-744.

Brunken JN, Estes JR. 1975. Cytological and morphological variation in *Panicum virgatum* L. *Southwest. Nat.* **19**: 379-385.

Byrd GT, May PA. 2000. Physiological comparisons of switchgrass cultivars differing in transpiration efficiency. *Crop Sci.* **40**: 1271-1277.

Carris L, Castlebury L, Zale J. 2008. First report of *Tilletia pulcherrima* bunt on switchgrass (*Panicum virgatum*) in Texas. *Plant Dis.* **92**: 1707-1707.

Cassida KA, Muir JP, Hussey MA, Read JC, Venuto BC, Ocumpaugh WR. 2005. Biomass yield and stand characteristics of switchgrass in south central U.S. environments. *Crop Sci.* **45**: 673-681.

Crouch JA, Beirn LA, Cortese LM, Bonos SA, Clarke BB. 2009. Anthracnose disease of switchgrass caused by the novel fungal species *Colletotrichum navitas*. *Mycol. Res.* **113**: 1411-1421.

Denchev PD, Conger BV. 1994. Plant regeneraton from callus cultures of switchgrass. *Crop Sci.* **34**: 1623-1627.

Frazier T, Shen Z, Zhao B, Bush E. 2012. First report of *Puccinia emaculata* infection on switchgrass in Virginia. *Plant Dis.* **97**: 424-424.

Fu C, Mielenz JR, Xiao X, Ge Y, Hamilton CY, Rodriguez M, Chen F, Foston M, Ragauskas A, Bouton J, Dixon RA, Wang Z-Y. 2011. Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proc. Natl. Acad. Sci. USA* **108**: 3803-3808.

Garrett K, Dendy S, Power A, Blaisdell G, Alexander HM, McCarron J. 2004. Barley yellow dwarf disease in natural populations of dominant tallgrass prairie species in Kansas. *Plant Dis.* **88**: 574-574.

Gill US, Uppalapati SR, Nakashima J, Mysore KS. 2015. Characterization of *Brachypodium distachyon* as a nonhost model against switchgrass rust pathogen *Puccinia emaculata*. *BMC Plant Biol.* **15**: 113.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**: D1178-D1186.

Gravert CE, Tiffany LH, Munkvold GP. 2000. Outbreak of smut caused by *Tilletia maclaganii* on cultivated switchgrass in Iowa. *Plant Dis.* **84**: 596-596.

Gunter LE, Tuskan GA, Wullschleger SD. 1996. Diversity among populations of switchgrass based on RAPD markers. *Crop Sci.* **36**: 1017-1022.

Hammond-Kosack KE, Jones JDG. 1997. Plant disease resistance genes. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**: 575-607.

Heath MC. 2000. Hypersensitive response-related death. *Plant Mol. Biol.* **44**: 321-334.

Hirsch RL, TeBeest DO, Bluhm BH, West CP. 2010. First report of rust caused by *Puccinia emaculata* on switchgrass in Arkansas. *Plant Dis.* **94**: 381-381.

Hope HJ, McElroy A. 1990. Low-temperature tolerance of switchgrass (*Panicum virgatum* L.). *Can. J. Plant Sci.* **70**: 1091-1096.

Hopkins AA, Taliaferro CM, Murphy CD, Christian D. 1996. Chromosome number and nuclear DNA content of several switchgrass populations. *Crop Sci.* **36**: 1192-1195.

Hopkins AA, Vogel KP, Moore KJ. 1993. Predicted and realized gains from selection for in vitro dry matter digestibility and forage yield in switchgrass. *Crop Sci.* **33**: 253-258.

Huang L-K, Bughrara SS, Zhang X-Q, Bales-Arcelo CJ, Bin X. 2011. Genetic diversity of switchgrass and its relative species in *Panicum* genus using molecular markers. *Biochem. Syst. Ecol.* **39**: 685-693.

Hultquist SJ, Vogel K, Lee D, Arumuganathan K, Kaeppler S. 1996. Chloroplast DNA and nuclear DNA content variations among cultivars of switchgrass, *Panicum virgatum* L. *Crop Sci.* **36**: 1049-1052.

Jones JDG, Dangl JL. 2006. The plant immune system. *Nature* **444**: 323-329.

Krupinsky J, Berdahl J, Schoch C, Rossman A. 2004. Leaf spot on switch grass (*Panicum virgatum*), symptoms of a new disease caused by *Bipolaris oryzae*. *Can. J. Plant Pathol.* **26**: 371-378.

Langridge P, Baumann U 2008. Self-incompatibility in the grasses. *Self-Incompatibility in Flowering Plants*: Springer, 275-287.

Lemus R, Brummer EC, Moore KJ, Molstad NE, Burras CL, Barker MF. 2002. Biomass yield and quality of 20 switchgrass populations in southern Iowa, USA. *Biomass Bioenergy* **23**: 433-442.

Lewandowski I, Scurlock JMO, Lindvall E, Christou M. 2003. The development and current status of perennial rhizomatous grasses as energy crops in the US and Europe. *Biomass Bioenergy* **25**: 335-361.

Li R, Qu R. 2011. High throughput *Agrobacterium*-mediated switchgrass transformation. *Biomass Bioenergy* **35**: 1046-1054.

Liu L, Thames S, Wu Y. 2014. Lowland switchgrass plants in populations set completely outcrossed seeds under field conditions as assessed with SSR markers. *BioEnergy Res.* **7**: 253-259.

Liu L, Wu Y. 2012. Identification of a selfing compatible genotype and mode of inheritance in switchgrass. *BioEnergy Res.* **5**: 662-668.

Liu Y, Zhang X, Tran H, Shan L, Kim J, Childs K, Ervin E, Frazier T, Zhao B. 2015. Assessment of drought tolerance of 49 switchgrass (*Panicum virgatum*) genotypes using physiological and morphological parameters. *Biotechnol. Biofuels* **8**: 1-18.

Lu K, Kaeppler SM, Vogel KP, Arumuganathan K, Lee DJ. 1998. Nuclear DNA content and chromosome numbers in switchgrass. *Great Plains Res.* **8**: 269-280.

Martinez-Reyna J, Vogel K. 2002. Incompatibility systems in switchgrass. *Crop Sci.* **42**: 1800-1805.

Martínez-Reyna JM, Vogel KP, Caha C, Lee DJ. 2001. Meiotic stability, chloroplast DNA polymorphisms, and morphological traits of upland × lowland switchgrass reciprocal hybrids. *Crop Sci.* **41**: 1579-1583.

McLaughlin SB, Kszos LA. 2005. Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. *Biomass Bioenergy* **28**: 515-535.

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

Missaoui A, Paterson A, Bouton J. 2006. Molecular markers for the classification of switchgrass (*Panicum virgatum* L.) germplasm and to assess genetic diversity in three synthetic switchgrass populations. *Genet. Resour. Crop Evol.* **53**: 1291-1302.

Narasimhamoorthy B, Saha MC, Swaller T, Bouton JH. 2008. Genetic diversity in switchgrass collections assessed by EST-SSR markers. *BioEnergy Res.* **1**: 136-146.

Nielson EL. 1944. Analysis of variation in *Panicum virgatum*. *J. Agric. Res.* **69**: 327-353.

Nürnberger T, Brunner F, Kemmerling B, Piater L. 2004. Innate immunity in plants and animals: striking similarities and obvious differences. *Immunol. Rev.* **198**: 249-266.

Porter B, Paidi M, Ming R, Alam M, Nishijima W, Zhu Y. 2009. Genome-wide analysis of *Carica papaya* reveals a small *NBS* resistance gene family. *Mol. Genet. Genomics* **281**: 609-626.

Porter CL, Jr. 1966. An analysis of variation between upland and lowland switchgrass, *Panicum Virgatum* L., in central Oklahoma. *Ecology* **47**: 980-992.

Richards HA, Rudas VA, Sun H, McDaniel JK, Tomaszewski Z, Conger BV. 2001. Construction of a GFP-BAR plasmid and its use for switchgrass transformation. *Plant Cell Rep.* **20**: 48-54.

Riley RD, Vogel KP. 1982. Chromosome numbers of released cultivars of switchgrass, indiangrass, big bluestem, and sand bluestem. *Crop Sci.* **22**: 1082-1083.

Sanderson MA, Reed RL, McLaughlin SB, Wullschleger SD, Conger BV, Parrish DJ, Wolf DD, Taliaferro C, Hopkins AA, Ocumpaugh WR, Hussey MA, Read JC, Tischler CR. 1996. Switchgrass as a sustainable bioenergy crop. *Bioresour. Technol.* **56**: 83-93.

Saski C, Li Z, Feltus F, Luo H. 2011. New genomic resources for switchgrass: a BAC library and comparative analysis of homoeologous genomic regions harboring bioenergy traits. *BMC Genomics* **12**: 1-12.

Serba DD, Uppalapati SR, Mukherjee S, Krom N, Tang Y, Mysore KS, Saha MC. 2015. Transcriptome profiling of rust resistance in switchgrass using RNA-Seq analysis. *Plant Genome* **8**.

Sill W, Pickett R. 1957. A new virus disease of switchgrass, *Panicum virgatum* L. *L. Plant Dis. Rep.* **41**: 241-249.

Somleva MN, Snell KD, Beaulieu JJ, Peoples OP, Garrison BR, Patterson NA. 2008. Production of polyhydroxybutyrate in switchgrass, a value-added co-product in an important lignocellulosic biomass crop. *Plant Biotechnol. J.* **6**: 663-678.

Sykes V, Allen F, Mielenz J, Stewart CN, Jr., Windham M, Hamilton C, Rodriguez M, Jr., Yee K. 2015. Reduction of ethanol yield from switchgrass infected with rust caused by *Puccinia emaculata*. *BioEnergy Res.*: 1-9.

Taliaferro CM, Das M. 2002a. Breeding and selection of new switchgrass varieties for increased biomass production. Oak Ridge, Tennessee: Oak Ridge National Laboratory.

Taliaferro CM, Das M. 2002b. Breeding and selection of new switchgrass varieties for increased biomass production. *ORNL/SUB-02-19XSY162C/01. Oak Ridge National Laboratory, Oak Ridge, Tennessee.*

Thomason WE, Raun WR, Johnson GV, Taliaferro CM, Freeman KW, Wynn KJ, Mullen RW. 2005. Switchgrass response to harvest frequency and time and rate of applied nitrogen. *J. Plant Nutr.* **27**: 1199-1226.

Thomsen PM, Brummer EC, Shriver J, Munkvold GP. 2008. Biomass yield reductions in switchgrass due to smut caused by *Tilletia maclaganii*. *Plant Health Prog. doi* **10**: 1094.

Tobias CM, Hayden DM, Twigg P, Sarath G. 2006. Genic microsatellite markers derived from EST sequences of switchgrass (*Panicum virgatum* L.). *Mol. Ecol. Notes* **6**: 185-187.

Uppalapati S, Serba D, Ishiga Y, Szabo L, Mittal S, Bhandari H, Bouton J, Mysore K, Saha M. 2013. Characterization of the rust fungus, *Puccinia emaculata*, and evaluation of genetic variability for rust resistance in switchgrass populations. *BioEnergy Res.* **6**: 458-468.

Wang Y, Zeng X, Iyer NJ, Bryant DW, Mockler TC, Mahalingam R. 2012. Exploring the switchgrass transcriptome using second-generation sequencing technology. *PloS ONE* **7**: e34225.

Wang YW, Samuels TD, Wu YQ. 2011. Development of 1,030 genomic SSR markers in switchgrass. *Theor. Appl. Genet.* **122**: 677-686.

Weaver JE. 1968. *Prairie plants and their environment. A fifty-year study in the midwest*. Lincoln, NE University of Nebraska Press

Wullschleger SD, Sanderson MA, McLaughlin SB, Biradar DP, Rayburn AL. 1996. Photosynthetic rates and ploidy levels among populations of switchgrass. *Crop Sci.* **36**: 306-312.

Xu B, Escamilla-Trevino LL, Sathitsuksanoh N, Shen Z, Shen H, Zhang YH, Dixon RA, Zhao B. 2011a. Silencing of 4-coumarate:coenzyme A ligase in switchgrass leads to reduced lignin content and improved fermentable sugar yields for biofuel production. *New Phytol.* **192**: 611-625.

Xu B, Huang L, Shen Z, Welbaum GE, Zhang X, Zhao B. 2011b. Selection and characterization of a new switchgrass (*Panicum virgatum* L.) line with high somatic embryogenic capacity for genetic transformation. *Sci. Hortic.* **129**: 854-861.

Xu B, Sathitsuksanoh N, Tang Y, Udvardi MK, Zhang J-Y, Shen Z, Balota M, Harich K, Zhang P, Zhao B. 2012. Overexpression of *AtLOV1* in switchgrass alters plant architecture, lignin content, and flowering time. *PLoS ONE* **7**: e47399.

Yang Z, Shen Z, Tetreault H, Johnson L, Friebe B, Frazier T, Huang L-k, Burklew C, Zhang X-Q, Zhao B. 2014. Production of autopolyploid lowland switchgrass lines through in vitro chromosome doubling. *BioEnergy Res.* **7**: 232-242.

Zale J, Freshour L, Agarwal S, Sorochan J, Ownley BH, Gwinn KD, Castlebury LA. 2008. First report of rust on switchgrass (*Panicum virgatum*) caused by *Puccinia emaculata* in Tennessee. *Plant Dis.* **92**: 1710-1710.

Zhang JY, Lee YC, Torres-Jerez I, Wang M, Yin Y, Chou WC, He J, Shen H, Srivastava AC, Pennacchio C, Lindquist E, Grimwood J, Schmutz J, Xu Y, Sharma M, Sharma R, Bartley LE, Ronald PC, Saha MC, Dixon RA, Tang Y, Udvardi MK. 2013. Development of an integrated transcript sequence database and a gene expression atlas for gene discovery and analysis in switchgrass (*Panicum virgatum* L.). *Plant J.* **74**: 160-173.

Zhu Q, Bennetzen JL, Smith SM. 2013. Isolation and diversity analysis of resistance gene homologues from switchgrass. *G3 Genes Genom. Genet.* **3**: 1031-1042.

CHAPTER 2: Exploring Differences in the Transcriptomes of the Switchgrass Cultivars 'Dacotah' and 'Alamo'

ABSTRACT

Panicum virgatum, commonly known as switchgrass, has risen to prominence as a biofuel feedstock because it is a C_4 perennial grass with high biomass productivity. As a readily outcrossing species, naturally occurring switchgrass germplasms have a highly heterozygous genetic background. The large degree of genetic diversity within this species contributes to a wide range of phenotypic variations in important agronomic traits such as disease resistance, abiotic stress tolerance, and biomass production. In this study, we performed Illumina RNA-sequencing on the leaf transcriptomes of two switchgrass cultivars that exhibit significantly different phenotypes, 'Dacotah' and 'Alamo'. In total, 125,257,298 and 221,647,550 raw sequencing reads were collected for 'Dacotah' and 'Alamo', respectively. Two methods were used to create a full transcriptome for each cultivar: reference-guided assembly and *de novo* assembly. After mapping to the early release switchgrass genome, a total number of 77,421 and 83,094 gene models was assembled for 'Dacotah' and 'Alamo', respectively. Approximately 97,000 single copy *de novo* transcripts were assembled for each cultivar. To estimate the variation captured within our sequencing effort, we evaluated SNP discovery based on an increasing scale of included reads. Even with the deep transcriptome sequencing (18X-35X) performed in this study, the full breadth of polymorphism in these samples of switchgrass has not been entirely described. Gene expression analysis identified 3,165 genes that were significantly differentially expressed between the two cultivars. A total of 1,137 genes was found to be up-

regulated in ‘Dacotah’ and 1,798 genes was found to be up-regulated in ‘Alamo’. In addition, we identified 81 and 161 genes that were expressed only in ‘Dacotah’ and ‘Alamo’, respectively. The results of this study provide a foundation for identifying genes that may contribute to desirable agronomic traits in switchgrass. Furthermore, our results may guide the design of future RNA-sequencing experiments of switchgrass so that the extent of polymorphism within the species is fully captured.

INTRODUCTION

Panicum virgatum, commonly known as switchgrass, is a perennial prairie grass that has been the recent focus of several initiatives to develop it as a bioenergy feedstock (McLaughlin & Kszos, 2005). Once used solely for hay production and originally bred for fodder, switchgrass has been identified as a dedicated feedstock for cellulosic biomass production. There are many attributes of switchgrass that make it appealing as a bioenergy crop. For instance, switchgrass is a C_4 plant that is native to North America. When grown for use as a biofuel, it has the potential to improve soil conservation while thriving in a low-input cultivation scheme (McKendry, 2002). Furthermore, there are many varieties of switchgrass that are well-adapted to grow in different environmental conditions across the vast prairie regions of North America (Casler *et al.*, 2004; Casler *et al.*, 2007).

As a readily outcrossing species, native switchgrass germplasms have highly heterozygous genetic backgrounds and grow over a range of habitats. Two main ecotypes of switchgrass, upland and lowland, have emerged that are suited to different growth environments

(Das *et al.*, 2004). In addition to their wide geographic growth range, switchgrass cultivars display substantial diversity in important agronomic traits such as abiotic stress tolerance, disease resistance, and biomass production (Lemus *et al.*, 2002). Recently, interest in switchgrass production has shifted to maximizing bioethanol output from its cellulosic biomass. This effort has focused upon a select few varieties including the lowland varieties 'Alamo' and 'Kanlow' and the upland variety 'Cave-In-Rock' (Lewandowski *et al.*, 2003). Although interest in switchgrass as a feedstock has increased, few major breeding efforts have been employed over the last decade to improve its agronomic traits (Wu *et al.*, 2006; Bouton, 2007).

Several studies have used cutting-edge genetic and genomic tools to dissect the underlying molecular mechanisms controlling agronomically important traits in widely grown switchgrass cultivars. For instance, a previous investigation into gene expression used microarrays and RT-PCR to determine how regulatory RNAs impact cellulosic biomass accumulation (Fu *et al.*, 2012). Affymetrix Gene Chips have been used to identify genes in switchgrass that are influenced by heat stress (Li *et al.*, 2013). In addition to these methods, RNA-sequencing technologies have recently been employed in switchgrass in an attempt to construct a full transcriptome for this species (Wang *et al.*, 2012b) and to identify genes that are differentially regulated during infection by switchgrass rust (Serba *et al.*, 2015). The results of Wang *et al.* (2012) set a precedent for gene discovery and use of a closely-related species, *Setaria italica*, as a reference when investigating switchgrass. Despite recent advances in technology, few studies have looked at transcriptome differences between lowland and upland switchgrass cultivars. One study by Childs *et al.* (2014) examined transcript and SNP differences between the two ecotypes; however, it focused on tetraploid lowland cultivars and octoploid upland cultivars (Childs *et al.*, 2014b). Aside from the study by Serba *et al.* (2015) that focused

on disease resistance and the report by Childs *et al* (2014), no reports have analyzed transcriptome differences between lowland and upland switchgrass ecotypes that are of the same ploidy.

The upland and lowland varieties of switchgrass differ in ploidy number, and are usually reproductively isolated; however, crosses between the two ecotypes are possible if the plants have the same ploidy (Martínez-Reyna *et al.*, 2001). The switchgrass cultivar 'Alamo' is a lowland tetraploid that is popular for its high biomass yield. Like other lowland cultivars, it is adapted to humid climates and can grow upwards in height of 61 to 305 cm (Cortese *et al.*, 2010). Its coarse stem and general robustness have made it one of the highest-yielding varieties of switchgrass that has been identified, and as a result, it has become a major target as a bioenergy crop (McLaughlin & Kszos, 2005). 'Dacotah' is an upland tetraploid that was formally identified two decades ago (Barker *et al.*, 1990). It is a diminutive but thick-stemmed grass, collected from the north-central plains of North America and is known for low biomass yield, even compared to other upland varieties such as 'Cave-In-Rock' (Barker *et al.*, 1990).

Understanding the genetic mechanisms underlying the phenotypic differences in lowland and upland switchgrass will lead to the development of new switchgrass cultivars with improved agronomic traits. In this study, we used RNA-sequencing to sequence, assemble, and analyze the leaf transcriptomes of two tetraploid switchgrass cultivars of different ecotypes, 'Alamo' and 'Dacotah'. We examined nucleotide variation and gene expression differences that naturally exist between these two switchgrass cultivars. The results of this study provide a foundation for further genetic studies of lowland and upland switchgrass. In addition, these results will aid in the design of future switchgrass RNA-sequencing projects.

MATERIALS AND METHODS

Seed material and plant growth conditions

Seeds for switchgrass cultivars Alamo and Dacotah were obtained from the USDA Plant Genetic Resources Conservation Unit (Griffin, GA). The seeds were planted separately in pots (15 cm diameter x 15 cm) containing Sunshine Mix #1 series soil and grown in the greenhouse facility at Virginia Tech. The plants were maintained in the greenhouse at 28/26 °C (day/night) under a 14-hour photoperiod. Ten weeks following germination, one 'Alamo' plant and one 'Dacotah' plant were selected for further analysis. Tillers from each of these plants were clonally propagated into four biological replicates and after transplanting, the plants resumed growth in the greenhouse.

RNA isolation, cDNA library generation, and sequencing

For total RNA isolation, approximately 100-200 mg of leaf tissue was collected from the first fully expanded leaf of E3 stage switchgrass tillers. Tissue was sampled from each of the four 'Alamo' and 'Dacotah' plants, and placed in separate 2 mL epi-tubes that contained a metal BB. The tubes were immediately frozen in liquid nitrogen and stored at -80°C. Using a Geno/Grinder 2000 (SPEX SamplePrep, Metuchen, NJ, USA), the tissue was ground into a fine powder and total RNA was extracted using a modified TRIzol combined with columns method. This method included an initial RNA extraction phase using the TRIzol reagent (Life technologies, Grand Island, NY, USA) followed by RNA purification using columns and reagents contained in the QIAGEN RNeasy Plant Mini Kit (QIAGEN, Germantown, MD, USA).

The total RNA was re-suspended in 50 µL of RNase-free ddH_2O and the overall quality and quantity was assessed using a Nanodrop ND-1000 (Wilmington, DE, USA) and a BioAnalyzer (Virginia Bioinformatics Institute, Blacksburg, VA, USA). The samples were stored at -80°C until sequencing.

One single sample of 'Dacotah' and one single sample of 'Alamo' were sent to The Genomics Resources Core Facility at Weill Cornell Medical College for deep RNA-sequencing. cDNA library preparation and Illumina HiSeq2000 100 bp paired-end high-throughput sequencing were performed at the core facility according to the manufacturer's protocol. The 100 bp RNA-seq reads generated for 'Alamo' and 'Dacotah' were used for construction of gene models and variant detection. The remaining three total RNA samples (three biological replicates) were sent to Michigan State University for Illumina HiSeq2000 50 bp paired-end high-throughput sequencing. The 50 bp RNA-seq reads generated for three samples of 'Dacotah' and three samples of 'Alamo' were used for differential gene expression analysis.

Transcriptome mapping, assembly, and annotation software

The early release switchgrass reference genome (v 0.0), which is based on the AP13 variety of 'Alamo', and the foxtail millet genome (*S. italica,* v 1.0) were both obtained from Phytozome (Goodstein *et al.*, 2012). *De novo* assembly of the Alamo and Dacotah transcriptomes was performed using the Trinity package (release 2013-02-25, (Grabherr *et al.*, 2011)) with default settings for filtering and assembly. To assess putative transcript quality for the *de novo* assembly, we used NCBI's BLASTx (v 2.2.28 implementation) and a query database

of known proteins sourced from UniProt. A lower cutoff for e-values was set at 0.005 and the lowest e-value hit was accepted as the putative annotation for the switchgrass *de novo* transcript.

Reference-guided assembly of the switchgrass reads was achieved using the programs that are readily available in the Tuxedo Suite pipeline, including Bowtie (v 0.12.7, (Langmead *et al.*, 2009)), TopHat (v 2.0.1, (Trapnell *et al.*, 2009)) and Cufflinks (v 2.1.1, (Trapnell *et al.*, 2010)). All of the programs were executed using the default parameters for filtering, alignment, and creation of gene models.

Transcriptome profiling, variant detection, and SNP depth analyses

The method for variant prediction in this study follows the pipeline for variant prediction established by the 1000 Genomes Project (Lozano *et al.*, 2012). To compare sequence variations between the two cultivars, we used the early release switchgrass reference genome (v 0.0), the foxtail millet reference genome (v 1.0), and the de novo assembled transcriptomes of the two cultivars that we created. The clean 100 bp Illumina sequence reads from each cultivar were then aligned to each of the references of choice using the Burrows-Wheeler Aligner (v 0.7.4, (Li & Durbin, 2009)). In the case of the de novo transcriptomes, the Alamo reads were aligned to the assembled Dacotah transcriptome and vise-versa. After read mapping, the samtools program was used to process the alignment (v 0.1.18, (Li *et al.*, 2009)). Additionally, Picard software (v 1.66, http://broadinstitute.github.io/picard) was used to alter the read group tags across each alignment. The sorted and formatted alignments of the 'Dacotah' and 'Alamo' reference-assembled transcriptomes, as well as the 'Dacotah' and 'Alamo' *de novo* assembled transcriptomes, were then used for polymorphism discovery. Polymorphisms were identified using the Genome

Analysis Toolkit (v 1.6-5, (McKenna et al., 2010)) with default parameters. For each set, singleton reads were removed from the analysis when calling polymorphisms.

To investigate how many reads are necessary to identify the majority of polymorphisms in switchgrass, we created several pseudo-read sets of diminishing size and called polymorphisms for each set. All scripting was done using Python. Using the transcriptomes sequenced for 'Alamo' and 'Dacotah', we down-sampled the full set of clean 100 bp reads randomly in order to create data sets of smaller size. This sampling algorithm set a probability and based on whether a random integer between 0 and 100 fell below that set probability, the algorithm randomly pulled a read and its mate pair. This allowed for the creation of sampled sets of reads representing 75%, 50%, 25%, 12%, and 6.5% of the total sequenced reads. The sets of down-sampled reads were then used as inputs to predict polymorphisms across both *de novo* transcriptomes and the genome reference-guided assemblies.

Differential expression and gene set enrichment analyses

The 50 bp paired-end Illumina sequencing reads for the three biological replicates of 'Alamo' and 'Dacotah were imported into CLC Genomics Workbench (v 7.5.1, http://www.clcbio.com/) and were assessed for quality. Next, the reads were trimmed 10 bp from the 5' end and 1 bp from the 3' end, and reads with an overall quality score of less than Q20 were discarded. After trimming, the reads were mapped to the switchgrass reference genome (v.0.0) using the default parameters. Following read mapping, an experiment was created in CLC Genomics Workbench that compared the RPKM (reads per kilobase per million of reads mapped) gene expression values between the three 'Alamo' and three 'Dacotah' samples.

Statistical analyses were performed on the data using the Empirical analysis of Differential Gene Expression (EDGE) tool that is embedded in the software. Gene Set Enrichment Analysis (GSEA) was performed using the topGO package in R. All statistical analyses were performed using Fisher's exact test and a p-value cut off of <0.01.

RESULTS AND DISCUSSION

Assembly and mapping of switchgrass transcriptomes

In this study, we used RNA-sequencing to analyze the leaf transcriptomes of 'Dacotah', a low biomass yielding cultivar, and 'Alamo', a high biomass yielding cultivar (Figure 2-1). Switchgrass is a readily out-crossing species and as such, switchgrass plants have highly heterozygous genetic backgrounds. Even among members of the same cultivar, a large degree of genetic variation can exist (Cortese *et al.*, 2010). Therefore, only one 'Dacotah' plant and one 'Alamo' plant were used for this sequencing effort. Total RNA was extracted from leaf tissue of 'Dacotah' and 'Alamo' and sent for Illumina 100 bp paired-end RNA-sequencing. Overall, 125,257,298 (18x coverage) and 221,647,550 (35x coverage) raw sequencing reads were collected for 'Dacotah' and 'Alamo', respectively. After filtering for low-quality reads, 101,884,203 clean reads were obtained for 'Dacotah' and 178,687,444 clean reads were obtained for 'Alamo'. The clean reads were used for downstream analyses.

Due to the complex nature of switchgrass genetics, two independent approaches were used to create a full leaf transcriptome for each cultivar. The first approach used a reference-guided method to map the clean sequencing reads to the 410,030 contigs of the first release

switchgrass genome (v 0.0). The switchgrass reference genome is based on the DNA sequence from the AP13 variety of the cv. Alamo. The use of a reference sequence improves the quality of transcriptome assembly because it presents genomic regions, along with their intrinsic polymorphism, to which the aligner can unambiguously map transcript reads. Using the switchgrass reference genome to guide the mapping, we assembled 77,421 gene models in 'Dacotah' and 83,094 in 'Alamo' (Table 2-1).

Since the draft switchgrass genome is comprised of thousands of contigs, which may break genes and confound alignments, we also exploited the known synteny in the grasses by mapping the switchgrass reads to the foxtail millet genome (v 1.0). Foxtail millet has a less complex and more refined reference genome that is phylogenetically close to switchgrass (Vicentini *et al.*, 2008; Doust *et al.*, 2009). When the foxtail millet genome was used as a reference to guide the read mappings, 44,144 gene models were predicted within 'Dacotah' and 43,786 gene models were predicted within 'Alamo' (Table 2-1). This is roughly half the number of transcripts detected when mapping the reads to the switchgrass genome. The creation of far fewer gene models, when aligned to the foxtail millet genome, could be attributed to the simplified architecture (diploid) and relatively smaller size (~490 Mb) of this species' genome.

The second approach used a *de novo* assembly to construct and identify transcripts within the leaf transcriptome of each cultivar. Using Trinity, a total of 233,284 transcripts were constructed for 'Dacotah', and 260,460 transcripts were constructed for 'Alamo'. After filtering this data to remove alternatively spliced transcripts and leave only single-copy gene annotations, 97,298 'Dacotah' and 97,446 'Alamo' transcripts remained (Table 2-1). A potential pitfall of creating a *de novo* transcript assembly from a polyploid genome is that it may merge transcripts and create composite genes. Therefore, to evaluate the quality of transcripts produced, we used

BLASTx to compare the collection of assembled transcripts to the UniProt database of known proteins. Based on their homology to switchgrass, the proteins from this multi-species database helped assess the quality of the *de novo* assembled transcripts. Interestingly, we found limited homology to non-switchgrass proteins, with only 13,270 'Dacotah' and 13,530 'Alamo' transcripts having significant hits to the UniProt database. This is compared to 23,249 unique UniProt gene IDs detected when using the switchgrass reference-guided gene models as a query. Between the 'Dacotah' and 'Alamo' *de novo* assemblies, only 8,887 UniProt gene IDs were detected within both cultivars.

Variation within the transcriptomes of 'Dacotah' and 'Alamo'

Sequencing the transcriptomes of these two cultivars provided a rich dataset to exploit for new discovery of variation. In this study, variants were detected using the gene models created from the two reference-guided alignments, as well as the *de novo*-assembled transcriptomes. The number of variants discovered between these two cultivars of switchgrass is listed in Table 2-2. For the switchgrass genome reference-guided assemblies, we found that the total number of SNPs between 'Dacotah' and the reference (1,260,677) was almost twice the total number of SNPs between 'Alamo' and the reference (652,817). However, the number of insertions (176,361) and deletions (244,635) found between the switchgrass reference genome and the 'Dacotah' was less than the number of insertions (227,179) and deletions (392,124) between 'Alamo' and the reference. Far fewer insertions and deletions were detected between the two cultivars and the foxtail millet genome (Table 2-2). This could be attributed either to the smaller size of the foxtail millet genome, or to the ~13.1 million year phylogenetic distance between the

two species (Bennetzen *et al.*, 2012). To help facilitate the use of these polymorphisms for either molecular breeding or marker development, a switchgrass genome browser was created based on the first release draft of the switchgrass genome (v 0.0). The individualized variant detection data, as well as the gene models assembled in this study, can be accessed here: http://genome.genetics.rutgers.edu/. An example of the browser interface is shown in Figure 2-2 and it displays the polymorphisms identified in Pavirv0007384, a gene that encodes a putative hydroxyproline-rich glycoprotein.

Although numerous polymorphisms were detected in 'Alamo' and 'Dacotah', we wanted to investigate whether or not the sequencing depth of this study was sufficient to capture the majority of variation within the switchgrass transcriptome. To accomplish this, we measured SNP discovery as a function of included reads. Several datasets of decreasing read number were randomly generated from the full set of transcriptome reads. Using these sets of down-sampled reads, we identified polymorphisms and observed at what level SNPs were no longer captured (Figure 2-3). If the extent of variation within switchgrass was observed in this experiment, an upward trend of polymorphism would be expected to plateau as it approached the largest set of sampled reads. As seen in Figure 3, there is no plateau in the number of SNPs identified, even with 100% of the reads included. Thus, the results of this analysis show that the breadth of variation within the switchgrass transcriptome was not fully captured by this sequencing effort, and a greater sequencing depth would be necessary to maximize SNP discovery.

The least number of SNPs detected in the down-sampled read sets was between the reference switchgrass genome (v 0.0), which was created based on the AP13 plant of cv. Alamo, and the sequences from the 'Alamo' plant used in this study. The large number of SNPs detected between these two plants is still relatively high considering that the two plants are from the same

cultivar. Interestingly, when comparing the reference genome to the 'Dacotah' reads, the polymorphism observed matched the results from the 'Alamo' reads in the smaller down-sampled sets. As the included read count increased, a greater number of polymorphism was observed between 'Dacotah' and the reference genome than was observed between 'Alamo' and the reference genome. 'Dacotah' is an upland switchgrass cultivar that may be genetically distant from the lowland 'Alamo'. Therefore, it was not surprising to find that a greater number of polymorphism was found between 'Dacotah' and the switchgrass reference genome than was found between 'Alamo' and the reference genome.

Differential gene expression between 'Dacotah' and 'Alamo'

'Dacotah' and 'Alamo' exhibit vastly different phenotypes, which may be due, at least in part, to alterations in gene expression. In this study, total RNA was extracted from three biological replicates of 'Dacotah' and 'Alamo' and sent for Illumina 50 bp paired-end RNA-sequencing. Overall, 171,147,538 raw reads were sequenced for 'Dacotah' and 160,280,986 raw reads were sequenced for 'Alamo'. The reads were trimmed to a length of 39 bp and filtered to discard reads with quality scores less than Q20. After quality trimming and filtering, 164,101,166 and 153,727,808 clean reads were obtained for 'Dacotah' and 'Alamo', respectively. For each biological replicate, the clean reads were mapped to the switchgrass reference genome (v 0.0) using CLC Genomics Workbench (v 7.5). Once the reads were mapped to the switchgrass reference genome, readily available tools in CLC Genomics Workbench were used to analyze differences in gene expression between the two cultivars. After filtering for genes that had a False Discovery Rate corrected p-value of less than 0.01 and a fold change

greater than two, 3,165 genes were found to be differentially expressed between the ‘Alamo’ and ‘Dacotah’ samples.

Of the genes differentially expressed between the two cultivars, 1,786 were up-regulated in ‘Alamo’ in comparison to ‘Dacotah’. Four of these genes were found to be expressed greater than 1,000 fold (Table 2-3). The first gene is Pavirv00007384, a putative hydroxyproline-rich glycoprotein (HRGP). There are several classes of HRGPs and all of these classes are recognized as fundamental components of the plant cell wall (Sommer-Knudsen *et al.*, 1998). Cross-linking of HRGPs has also been shown to strengthen the cell wall. Differences in the HRGPs of the cell walls of lowland and upland switchgrass cultivars could help explain several biomass-related phenotypes. For instance, lowland switchgrass cultivars are generally taller than their upland counterparts (Porter, 1966). In addition, lowland cultivars also tend to have long, broad, thick leaves while the upland cultivars have short, succulent, thinner leaves (Cortese *et al.*, 2010). The plant cell wall is also the first physical barrier encountered by plant pathogens (Deepak *et al.*, 2010). In addition to structure, HRGPs have been shown to be up-regulated during pathogen invasion (Deepak *et al.*, 2010). Most lowland ecotypes, including the cultivar ‘Alamo’, are more resistant to disease than their upland counterparts. Therefore, the up-regulation of Pavirv0007384 under normal conditions may also contribute to disease resistance in lowland switchgrass cultivars. The second gene is Pavir00039173, which codes for a potential glutathione-S-transferase (GST) family protein. GST proteins are involved in numerous biological and physiological processes in plants including response to abiotic stresses such as salt and drought, detoxification, and transport of phytochemicals (Edwards *et al.*, 2000). Our previous study on salinity tolerance in switchgrass has suggested that ‘Alamo’ is more tolerant to salt stress than ‘Dacotah’ (Xie *et al.*, 2014). It will be interesting to further investigate if this putative GST gene

indeed has a key role in abiotic stress response in ‘Alamo’ and other lowland switchgrass cultivars. The other two genes, Pavirv00024836 and Pavirv00037271, are not annotated in the switchgrass genome (v 0.0). BLAST searches of these protein sequences against the non-redundant protein database show that they have limited homology to hypothetical and uncharacterized proteins of other plant species. Therefore, further studies are needed in order to determine what role these genes have in switchgrass.

In addition to the ‘Alamo’ up-regulated genes, we identified 1,137 genes that were up-regulated in ‘Dacotah’. Two of these genes, Pavirv00039194 and Pavirv00057008, showed greater than 1,000 fold change difference in expression over ‘Alamo’ (Table 2-3). Pavirv00039194 is annotated as a putative thioredoxin superfamily protein. Thioredoxin proteins have been shown to play an important role in plant response to oxidative stresses and the induction of reactive oxygen species (Vieira Dos Santos & Rey, 2006). These reactive oxygen species, in particular hydrogen peroxide and superoxide radicals, ultimately damage important cellular components such as DNA, protein, and lipids (Reddy *et al.*, 2004). Drought stress is one of the major inducers of reactive oxygen species in plants. In the case of switchgrass, upland switchgrass cultivars are adapted to the northern environment with low soil moisture conditions (Porter, 1966). Therefore, the significantly higher expression of Pavirv00039194 in ‘Dacotah’, in comparison to ‘Alamo’, might have evolved as a mechanism to tolerate arid conditions. The second gene that was up-regulated, Pavirv00057008, was not annotated in the first release switchgrass genome. A BLAST search of this gene’s protein sequence reveals limited homology to other proteins with no significant hits. The current protein sequence for this gene is comprised of 77 amino acids. Since the switchgrass protein sequences are derived from the genome sequence, this suggests that the nucleotide sequence for this gene is incomplete, or that this is a

pseudogene that is expressed. More studies will be necessary to determine what role this protein has in ‘Dacotah’ growth and development.

Alternative to changes in gene expression, the difference in phenotypes of ‘Alamo’ and ‘Dacotah’ could be attributed to a subset of genes that are specifically expressed in each cultivar. A total of 161 genes were found to be expressed only in ‘Alamo’. In the ‘Alamo’-only gene dataset, Pavirv00057975, Pavirv00056926, and Pavirv00066321 are the three genes that exhibited the greatest levels of expression (Table 2-3). The first gene, Pavirv00057975, is predicted to encode a protein that is homologous to the cysteine-rich receptor-like kinase (CRK) 22 protein of Arabidopsis. CRKs have been shown to play a role in numerous processes such as response to pathogen infection (Yeh *et al.*, 2015), oxidative stress response (Burdiak *et al.*, 2015), and cell death regulation (Chen *et al.*, 2004). Since this large subfamily of receptor-like kinases has so many functions in plant growth and development, additional studies are needed to determine the exact role of this gene in ‘Alamo’. Pavirv00056926 is predicted to encode a F-box family protein. Plant F-box proteins are also involved in various pathways including ubiquitination (Gagne *et al.*, 2004), self-compatibility recognition (Ushijima *et al.*, 2003), hormone signaling (Potuschak *et al.*, 2003), and leaf senescence (Woo *et al.*, 2001). Therefore, more studies are needed to determine the precise function of this gene. BLAST analysis of the third gene, Pavirv00066321, shows that it has homology to the EGG APPARATUS-1-LIKE protein of *Brachypodium distachyon* (62% identical, e-value = 2e-18). This protein is expressed in the floral organs and has been shown to regulate pollen tube growth (Márton *et al.*, 2005). Other studies, however, have demonstrated that members of this protein family are expressed in all tissues of the plant (Gray-Mitsumune & Matton, 2006). In addition, most of these proteins contain signal peptides, implicating that they may be extracellularly secreted and can potentially

act as signaling molecules (Gray-Mitsumune & Matton, 2006). It would be interesting to investigate the role that this protein has in ‘Alamo’ growth and development.

In contrast to Alamo, 81 genes were found to be expressed only in ‘Dacotah’. In the ‘Dacotah’-only gene data set, Pavirv00000718, Pavirv00068373, and Pavirv00059379 are the three genes that exhibited the greatest levels of expression (Table 2-3). Pavirv00000718 and Pavirv00068373 are not annotated in the switchgrass genome and BLAST analysis did not reveal any significant homology to any known proteins. Pavirv00059379 is predicted to code for a zinc-binding ribosomal protein. The annotated gene ontology (GO) terms that are associated with this gene suggest that it encodes for a protein that is a structural component of the ribosome and thus, it plays a role in protein translation. Since this protein is involved in an integral part of plant growth and development, it is interesting that this particular gene was found to be expressed only in ‘Dacotah’. It may be that this gene is not functional in ‘Alamo’ in this particular tissue, and that in leaf tissue ‘Alamo’ utilizes a similar but different gene. On the other hand, this gene may be functional in ‘Alamo’ but may be expressed in a spatiotemporal manner that differs from ‘Dacotah’.

Gene set enrichment analysis of differentially expressed genes in ‘Dacotah’ and ‘Alamo’

While the expression levels of certain genes could lend some explanation to the phenotypic traits observed between these two cultivars, variations in the phenotypes of ‘Alamo’ and ‘Dacotah’ could also be due to a collection of genes that regulate these traits. Using the topGO package in R, gene set enrichment analysis was performed on each of the four differentially expressed gene datasets (‘Dacotah’ up-regulated genes, ‘Alamo’ up-regulated

genes, ‘Dacotah’-only expressed genes, and ‘Alamo’-only expressed genes) to determine if these genes played an important role in any biological processes. A gene ontology category was considered significant if the results of all three statistical tests were $p < 0.01$ (Table 2-4).

For the genes up-regulated in ‘Alamo’, those underlying five biological processes were found to be enriched in the dataset including: 1) translation, 2) protein folding, 3) malate metabolic process, 4) DNA topological change, and 5) photosynthesis. The presences of genes responsible for protein synthesis, as well as genes involved in the TCA cycle and photosynthesis, suggests that the genes up-regulated in ‘Alamo’ may directly contribute to growth and development. Interestingly, the only biological processes that were significant for the genes expressed only in ‘Alamo’ were negative regulation of translation and intra-Golgi vesicle-mediated transport. This indicates that regulatory measures are in place to control gene expression in ‘Alamo’ at the post-transcriptional level.

Alternatively, for the genes up-regulated in ‘Dacotah’, six biological processes were found to be enriched in the dataset including: 1) oxidation-reduction process, 2) response to stress, 3) S-adenosylmethionine (SAM) biosynthetic process, 4) steroid biosynthetic process, 5) glutamate biosynthetic process, 6) G-protein coupled receptor signalling pathway, and 7) spermine biosynthetic process. It is interesting to observe that genes involved in SAM biosynthetic process and spermine biosynthesis are enriched in this dataset. Decarboxylated SAM contributes the α-aminopropionyl group to spermine (Slocum, 1991), which has been shown to be involved in salt and drought stress tolerance in Arabidopsis (Kusano *et al.*, 2007). Only one biological process, DNA catabolic activity, was found to be significant in the genes expressed only in ‘Dacotah’. Since one single gene was found in the dataset for this term, this

process may not have any significant application to the phenotypic variation observed between the two cultivars.

Despite the identification of these biological processes, inadequate annotation of the early release switchgrass genome prevents a comprehensive analysis of all of the differentially expressed genes. For instance, in the ‘Alamo’ and ‘Dacotah’ up-regulated datasets, 55% and 49% of the genes, respectively, are not annotated with any GO term and were not included in gene ontology analysis. Similarly, 75% of ‘Alamo’-only expressed genes and 81% of ‘Dacotah’-only expressed genes did not have a GO annotation and were excluded from this analysis. Therefore, improvement in the annotation of the switchgrass genome will lead to the identification of additional biological processes that may explain the variation in phenotypes between these two switchgrass cultivars.

CONCLUSION

The two main ecotypes of switchgrass, lowland and upland, exhibit vastly different phenotypic traits and therefore, they may be exploited as new resources for genetically improving the agronomic traits of this important biofuel crop. In this study, we performed RNA-sequencing and transcriptome analysis on the leaf tissue of two phenotypically different switchgrass cultivars: one lowland cultivar, ‘Alamo’, and one upland cultivar, ‘Dacotah’. The prevalence of polymorphism within these two cultivars of switchgrass exceeded our expectations, and we failed to capture the full extent of it in our experiment. Here we have set a lower bound for what to expect from polymorphism within switchgrass, but the upper bound still

must be identified. In addition, we identified a handful of genes that are significantly differentially expressed between the two cultivars. Further studies are needed to validate the function of these genes and to determine what role these genes play in switchgrass growth and development.

PROGRAMMING DATA AND ACKNOWLEDGEMENTS

All scripting done was using Python. All software used in the development and analyses are available upon request. Whole genome data generated by this analysis, such as gene models, are available in full on genome browser located at genome.genetics.rutgers.edu (Hupalo and Kern, 2013). Reads generated by the sequencing effort have been made available through the Sequence Read Archive. This project was supported by grants from the Joint feed stock genome program of U.S. Department of Energy and U.S. Department of Agriculture. The study was also partially supported by Virginia Tech's Virginia Agricultural Experiment Station (VA135872 to BZ).

ACKNOWLEDGEMENTS

This work was supported by a DOE-USDA grant (DE-FG02-08ER64629) awarded to JL, BZ, JM, DB, and AK. This work was also supported by a DOE-USDA grant (DE-SC0008338) awarded to KC, BZ, and XZ.

AUTHOR CONTRIBUTIONS

BZ, DB, JM, AK, and JL designed the sequencing experiments. TF, YL, JM, XZ, and JK carried out germination, tissue collection, and RNA extraction. Sequencing was performed at the Cornell Core Facility as well as at Michigan State University. DH, AK, TF, and KC designed the analysis of transcriptome data. DH and TF performed analysis and were the primary authors, with comments made by all other contributors.

References

Barker R, Haas R, Berdahl J, Jacobson E. 1990. Registration of 'Dacotah' switchgrass. *Crop Sci.* **30**: 1158.

Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Estep M, Feng L, Vaughn JN, Grimwood J, Jenkins J, Barry K, Lindquist E, Hellsten U, Deshpande S, Wang X, Wu X, Mitros T, Triplett J, Yang X, Ye C-Y, Mauro-Herrera M, Wang L, Li P, Sharma M, Sharma R, Ronald PC, Panaud O, Kellogg EA, Brutnell TP, Doust AN, Tuskan GA, Rokhsar D, Devos KM. 2012. Reference genome sequence of the model plant Setaria. *Nat. Biotech.* **30**: 555-561.

Bouton JH. 2007. Molecular breeding of switchgrass for use as a biofuel crop. *Curr. Opin. Genet. Dev.* **17**: 553-558.

Burdiak P, Rusaczonek A, Witoń D, Głów D, Karpiński S. 2015. Cysteine-rich receptor-like kinase CRK5 as a regulator of growth, development, and ultraviolet radiation responses in *Arabidopsis thaliana*. *J. Exp. Bot.*

Casler M, Vogel K, Taliaferro C, Ehlke N, Berdahl J, Brummer E, Kallenbach R, West C, Mitchell R. 2007. Latitudinal and longitudinal adaptation of switchgrass populations. *Crop Sci.* **47**: 2249-2260.

Casler M, Vogel K, Taliaferro C, Wynia R. 2004. Latitudinal adaptation of switchgrass populations. *Crop Sci.* **44**: 293-303.

Chen K, Fan B, Du L, Chen Z. 2004. Activation of hypersensitive cell death by pathogen-induced receptor-like protein kinases from Arabidopsis. *Plant Mol. Biol.* **56**: 271-283.

Childs KL, Nandety A, Hirsch CN, Góngora-Castillo E, Schmutz J, Kaeppler SM, Casler MD, Buell CR. 2014. Generation of transcript assemblies and identification of single nucleotide polymorphisms from seven lowland and upland cultivars of switchgrass. *Plant Genome* **7**.

Cortese L, Honig J, Miller C, Bonos S. 2010. Genetic Diversity of Twelve Switchgrass Populations Using Molecular and Morphological Markers. *Bioenerg. Res.* **3**: 262-271.

Das MK, Fuentes RG, Taliaferro CM. 2004. Genetic variability and trait relationships in switchgrass. *Crop Sci.* **44**: 443-448.

Deepak S, Shailasree S, Kini RK, Muck A, Mithöfer A, Shetty SH. 2010. Hydroxyproline-rich glycoproteins and plant defence. *J. Phytopathol.* **158**: 585-593.

Dos Santos CV, Rey P. 2006. Plant thioredoxins are key actors in the oxidative stress response. *Trends Plant Sci.* **11**: 329-334.

Doust AN, Kellogg EA, Devos KM, Bennetzen JL. 2009. Foxtail millet: a sequence-driven grass model system. *Plant Physiol.* **149**: 137-141.

Edwards R, Dixon DP, Walbot V. 2000. Plant glutathione S-transferases: enzymes with multiple functions in sickness and in health. *Trends Plant Sci.* **5**: 193-198.

Fu C, Sunkar R, Zhou C, Shen H, Zhang JY, Matts J, Wolf J, Mann DG, Stewart CN, Jr., Tang Y, Wang ZY. 2012. Overexpression of miR156 in switchgrass (Panicum virgatum L.) results in various morphological alterations and leads to improved biomass production. *Plant Biotechnol. J.* **10**: 443-452.

Gagne JM, Smalle J, Gingerich DJ, Walker JM, Yoo S-D, Yanagisawa S, Vierstra RD. 2004. *Arabidopsis* EIN3-binding F-box 1 and 2 form ubiquitin-protein ligases that repress

ethylene action and promote growth by directing EIN3 degradation. *Proc. Natl. Acad. Sci. USA* **101**: 6803-6808.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**: D1178-D1186.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotech.* **29**: 644-652.

Gray-Mitsumune M, Matton DP. 2006. The Egg apparatus 1 gene from maize is a member of a large gene family found in both monocots and dicots. *Planta* **223**: 618-625.

Hupalo D, Kern AD. 2013. Conservation and Functional Element Discovery in 20 Angiosperm Plant Genomes. *Mol. Biol. Evol.* **30**: 1729-1744.

Kusano T, Yamaguchi K, Berberich T, Takahashi Y. 2007. The polyamine spermine rescues Arabidopsis from salinity and drought stresses. *Plant Signal Behav.* **2**: 251-252.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Lemus R, Brummer EC, Moore KJ, Molstad NE, Burras CL, Barker MF. 2002. Biomass yield and quality of 20 switchgrass populations in southern Iowa, USA. *Biomass Bioenergy* **23**: 433-442.

Lewandowski I, Scurlock JMO, Lindvall E, Christou M. 2003. The development and current status of perennial rhizomatous grasses as energy crops in the US and Europe. *Biomass Bioenergy* **25**: 335-361.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754-1760.

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

Li Y-F, Wang Y, Tang Y, Kakani VG, Mahalingam R. 2013. Transcriptome analysis of heat stress response in switchgrass (Panicum virgatum L.). *BMC Plant Biol.* **13**: 153-153.

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

Martínez-Reyna JM, Vogel KP, Caha C, Lee DJ. 2001. Meiotic stability, chloroplast DNA polymorphisms, and morphological traits of upland × lowland switchgrass reciprocal hybrids. *Crop Sci.* **41**: 1579-1583.

Márton ML, Cordts S, Broadhvest J, Dresselhaus T. 2005. Micropylar pollen tube guidance by egg apparatus 1 of maize. *Science* **307**: 573-576.

McKendry P. 2002. Energy production from biomass (part 1): overview of biomass. *Bioresour. Technol.* **83**: 37-46.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**: 1297-1303.

McLaughlin SB, Kszos LA. 2005. Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. *Biomass Bioenergy* **28**: 515-535.

Porter CL, Jr. 1966. An analysis of variation between upland and lowland switchgrass, *Panicum Virgatum* L., in central Oklahoma. *Ecology* **47**: 980-992.

Potuschak T, Lechner E, Parmentier Y, Yanagisawa S, Grava S, Koncz C, Genschik P. 2003. EIN3-dependent regulation of plant ethylene hormone signaling by two Arabidopsis F box proteins: EBF1 and EBF2. *Cell* **115**: 679-689.

Reddy AR, Chaitanya KV, Vivekanandan M. 2004. Drought-induced responses of photosynthesis and antioxidant metabolism in higher plants. *J. Plant Physiol.* **161**: 1189-1202.

Serba DD, Uppalapati SR, Mukherjee S, Krom N, Tang Y, Mysore KS, Saha MC. 2015. Transcriptome profiling of rust resistance in switchgrass using RNA-Seq analysis. *Plant Genome* **8**.

Slocum RD. 1991. Polyamine biosynthesis in plants. *Biochemistry and physiology of polyamines in plants*: 23-40.

Sommer-Knudsen J, Bacic A, Clarke AE. 1998. Hydroxyproline-rich plant glycoproteins. *Phytochemistry* **47**: 483-497.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105-1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotech.* **28**: 511-515.

Ushijima K, Sassa H, Dandekar AM, Gradziel TM, Tao R, Hirano H. 2003. Structural and transcriptional analysis of the self-incompatibility locus of almond: identification of a pollen-expressed F-box gene with haplotype-specific polymorphism. *Plant Cell* **15**: 771-781.

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

Wang Y, Zeng X, Iyer NJ, Bryant DW, Mockler TC, Mahalingam R. 2012. Exploring the switchgrass transcriptome using second-generation sequencing technology. *PloS ONE* **7**: e34225.

Woo HR, Chung KM, Park J-H, Oh SA, Ahn T, Hong SH, Jang SK, Nam HG. 2001. ORE9, an F-box protein that regulates leaf senescence in Arabidopsis. *Plant Cell* **13**: 1779-1790.

Wu M, Wu Y, Wang M. 2006. Energy and emission benefits of alternative transportation liquid fuels derived from switchgrass: a fuel life cycle assessment. *Biotechnol. Prog.* **22**: 1012-1024.

Xie F, Stewart CN, Jr., Taki FA, He Q, Liu H, Zhang B. 2014. High-throughput deep sequencing shows that microRNAs play important roles in switchgrass responses to drought and salinity stress. *Plant Biotechnol. J.* **12**: 354-366.

Yeh Y-H, Chang Y-H, Huang P-Y, Huang J-B, Zimmerli L. 2015. Enhanced Arabidopsis pattern-triggered immunity by overexpression of cysteine-rich receptor-like kinases. *Front. Plant Sci.* **6**: 322.

Table 2-1. Transcriptome information gathered for each cultivar.

Included in the table is the number of clean reads, the number of *de novo* transcripts assembled, the number of single copy *de novo* transcripts assembled, the number of *de novo* transcripts that have homology to known proteins, and the number of single copy gene models created when mapping to two different reference genomes.

	Clean reads	*De novo* transcripts	*De novo* single copy	High Confidence *de novo* transcripts	*P. virgatum* reference single copy	*S. italia* reference single copy
'Alamo'	178,687,444	260460	97446	13530	83094	43786
'Dacotah'	101,884,203	233284	97298	13270	77421	44144

Table 2-2. Variants identified throughout the two switchgrass cultivars.

Shown are the polymorphisms found within the *de novo* assembly as well as the insertions, deletions, and SNPs found in gene models created from mapping reads to a reference genome.

	De novo variants	*P. virgatum* reference insertions	*P. virgatum* reference deletions	*P. virgatum* reference SNPs	*S. italica* reference insertions	*S. italica* reference deletions
'Alamo'	1205305	227179	392124	652817	7227	20977
'Dacotah'	1287078	176361	244635	1260677	4830	14655

Table 2-3. Top 5 expressed genes in the differentially expressed gene datasets.

The datasets include 'Alamo'-upregulated (AU) genes, 'Alamo'-only (AO) expressed genes, 'Dacotah'-upregulated (DU) genes, and 'Dacotah'-only (DO) expressed genes. Fold change and Mean RPKM are indicated where applicable and a "-" indicates not applicable. Putative gene descriptions and rice orthologues are listed as annotated in the switchgrass reference genome (v 0.0) and a "-" indicates missing information from the genome annotation .

Gene	Dataset	Fold Change	Mean RPKM	p-value	Putative Description	Rice orthologues
Pavirv00007384	AU	2,598	-	3.16e-102	hydroxyproline-rich glycoprotein family protein	Os07g48500.1
Pavirv00024836	AU	2,320	-	101.e-140	-	-
Pavirv00037271	AU	2,191	-	1.55e-62	-	Os06g38210.1
Pavirv00039173	AU	1,314	-	1.04e-62	Glutathione S-transferase family protein	Os01g27360.1
Pavirv00060275	AU	898	-	2.06e-72	cysteine proteinase inhibitor precursor protein	Os01g58890.1
Pavirv00057975	AO	-	170	4.61e-48	cysteine-rich RLK 22	Os07g43570.1
Pavirv00056926	AO	-	128	2.35e-60	F-box family protein	Os07g35050.1
Pavirv00066321	AO	-	109	2.05e-85	-	Os07g41410.1
Pavirv00051085	AO	-	94	6.12e-52	cycloartenol synthase 1	Os02g04730.1
Pavirv00060623	AO	-	74	7.37e-43	-	-
Pavirv00039194	DU	1,288	-	9.07e-105	Thioredoxin superfamily protein	Os09g38670.2
Pavirv00057008	DU	1,044	-	3.61e-129	-	-
Pavirv00014487	DU	857	-	2.77e-57	-	Os07g43730.1
Pavirv00036696	DU	525	-	3.41e-16	Leucine-rich repeat (LRR) family protein	Os02g40130.1
Pavirv00034225	DU	495	-	2.04e-33	HSP20-like chaperones superfamily protein	Os07g33350.2
Pavirv00000718	DO	-	88	1.00e-39	-	-
Pavirv00068373	DO	-	74	7.33e-31	-	-
Pavirv00059379	DO	-	58	1.67e-30	Zinc-binding ribosomal protein family protein	Os05g48320.1
Pavirv00001723	DO	-	58	1.21e-30	-	-
Pavirv00045727	DO	-	49	3.78e-25	-	Os12g04010.1

Table 2-4. TopGO gene set enrichment analysis of the GO category "Biological Process" for the four differentially expressed gene datasets.

Datasets include 'Alamo'-upregulated (AU) genes, 'Alamo'-only (AO) genes, 'Dacotah'-upregulated (DU) genes, and 'Dacotah'-only (DO) genes. Only the GO terms that were significant ($p < 0.01$) for all three Fisher tests are listed.

Data set	GO ID	GO Term	Annotated	Significant	Expected	Classic	Elim	Weight
AU	0006412	translation	1014	85	28.35	9.90E-20	1.40E-22	6.30E-25
AU	0006457	protein folding	241	15	6.74	0.00337	0.00131	0.0013
AU	0006108	malate metabolic process	20	4	0.56	0.00205	0.00143	0.0014
AU	0006265	DNA topological change	25	4	0.7	0.0048	0.00337	0.0034
AU	0015979	photosynthesis	138	12	3.86	0.00051	0.00021	0.0038
AO	0017148	negative regulation of translation	59	5	0.06	2.20E-09	1.70E-09	1.70E-09
AO	0006891	intra-Golgi vesicle-mediated transport	3	1	0	0.00281	0.0027	0.0027
DU	0055114	oxidation-reduction process	2866	95	59.82	2.90E-06	1.40E-08	4.80E-08
DU	0006950	response to stress	932	23	19.45	0.23274	0.10812	9.20E-06
DU	0006556	S-adenosylmethionine biosynthetic process	11	3	0.23	0.00132	0.00096	0.00096
DU	0006694	steroid biosynthetic process	194	11	4.05	0.00264	0.00111	0.00111
DU	0006537	glutamate biosynthetic process	12	3	0.25	0.00173	0.00126	0.00126
DU	0007186	G-protein coupled receptor signaling pathway	74	6	1.54	0.00449	0.00262	0.0064
DU	0006597	spermine biosynthetic process	7	2	0.15	0.00852	0.00688	0.00688
DO	0006308	DNA catabolic process	16	1	0.01	0.0075	0.0075	0.0075

Figure 2-1. Phenotypes of A) lowland switchgrass cultivar ‘Alamo’ and B) upland switchgrass cultivar ‘Dacotah’.

Pictured is A) ‘Alamo’, a high-yielding lowland cultivar and B) ‘Dacotah’, a low-yielding upland cultivar.

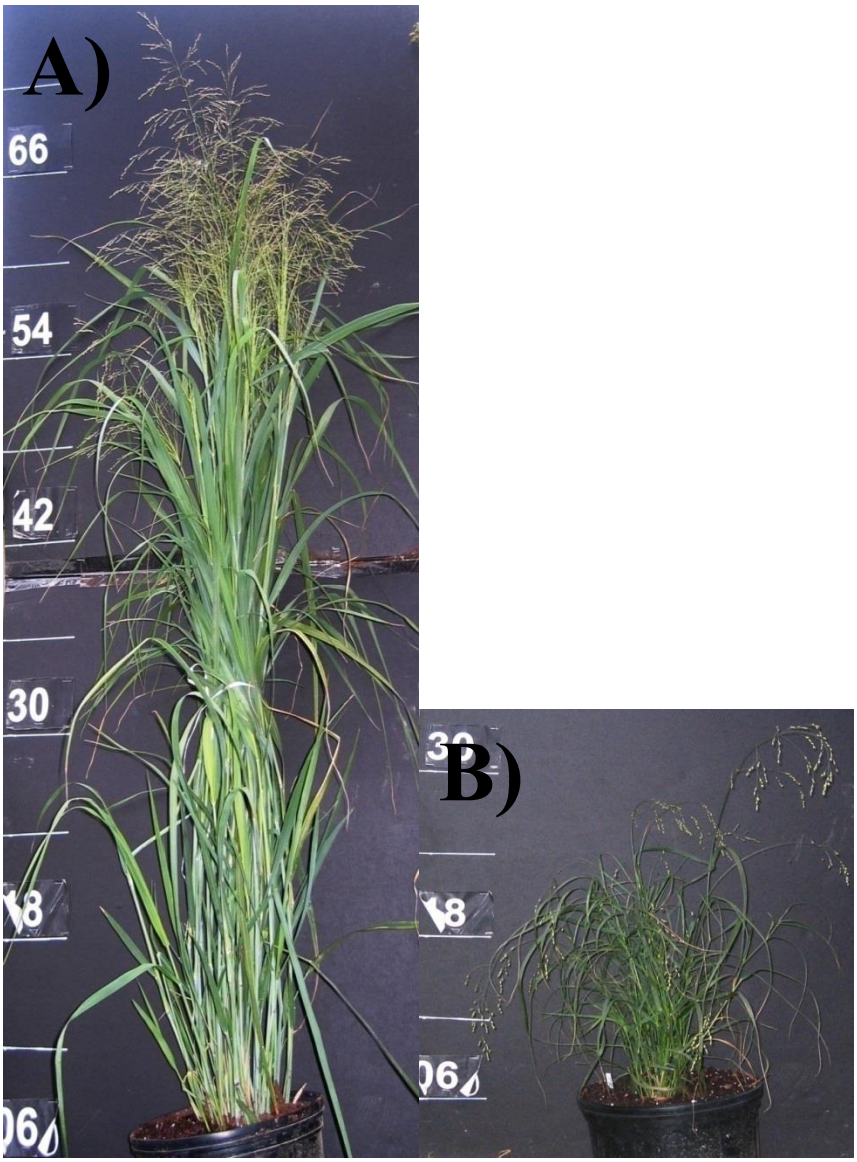


Figure 2-2. A UCSC browser screenshot of the Pavirv0007384 gene, a putative hydroxyproline-rich glycoprotein.

The browser shows the gene is located on sg0.contig45581 within the early release switchgrass genome. Shown in the figure are insertions, deletions, predicted splice junctions, and polymorphisms within the two cultivars. Below that is the gene models predicted for each cultivar from the RNA-sequencing data.

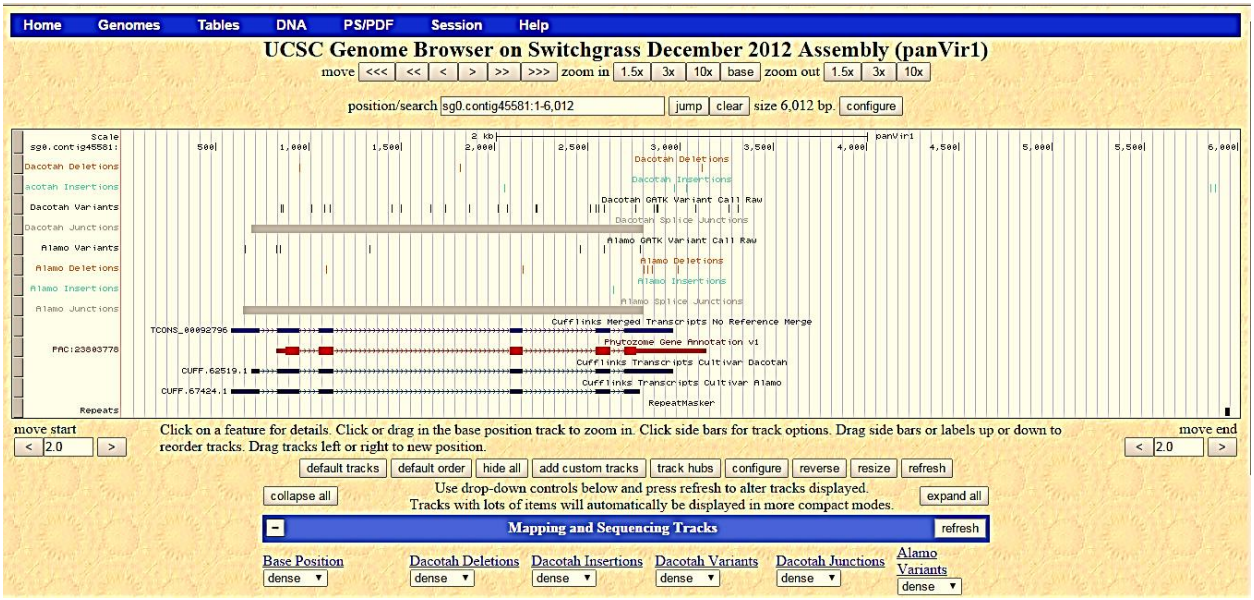
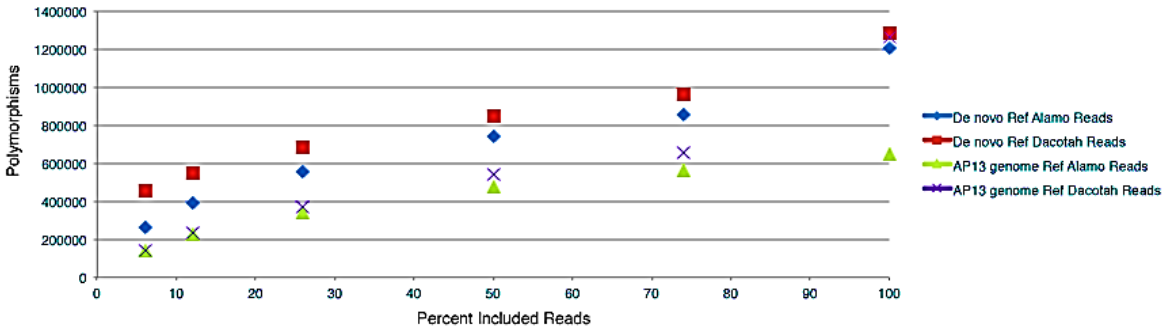


Figure 2-3. Polymorphisms identified based on down-sampled sets of reads.

Shown is the number of polymorphisms identified as a function of included reads for the AP13 switchgrass reference genome and the 'Dacotah' (purple) and 'Alamo' (green) transcriptome reads. Also shown is the number of polymorphisms identified as a function of included reads for the 'Dacotah' *de novo* assembled transcriptome and 'Alamo' reads (blue), and the 'Alamo' *de novo* assembled transcriptome and 'Dacotah' reads (red).



CHAPTER 3: Identification, characterization, and gene expression analysis of Nucleotide Binding site (NB)-type resistance gene homologues in switchgrass

ABSTRACT

Switchgrass is a warm-season perennial grass that can be used as a second generation bioenergy crop. However, foliar fungal pathogens, like switchgrass rust, have the potential to significantly reduce switchgrass biomass yield. Despite its importance as a prominent bioenergy crop, a genome-wide comprehensive analysis of NB-LRR disease resistance genes has yet to be performed in switchgrass. In this study, we used a homology-based computational approach to identify 1,011 potential NB-LRR resistance gene homologs (RGHs) in the switchgrass genome (v 1.1). In addition, we identified 40 RGHs that potentially contain unique domains including major sperm protein domain, jacalin-like binding domain, calmodulin-like binding, and thioredoxin. RNA-sequencing analysis of leaf tissue from 'Alamo', a rust-resistant switchgrass cultivar, and 'Dacotah', a rust-susceptible switchgrass cultivar, identified 2,634 high quality variants in the RGHs between the two cultivars. RNA-sequencing data from field-grown cv. Summer plants indicated that the expression of some of these RGHs was developmentally regulated. Our results provide useful insight into the molecular structure, distribution, and expression patterns of members of the NB-LRR gene family in switchgrass. These results also provide a foundation for future work aimed at elucidating the molecular mechanisms underlying disease resistance in this important bioenergy crop.

INTRODUCTION

Switchgrass is a North American prairie grass that can be used as a second generation bioenergy feedstock. Switchgrass is a readily outcrossing species and as such, native switchgrass germplasms have maintained a high level of genetic diversity over time (Gunter *et al.*, 1996; Narasimhamoorthy *et al.*, 2008). Two ecotypes of switchgrass, lowland and upland, have emerged that are adapted to different growth habitats. Lowland ecotypes are generally tetraploid in nature and grow in warm, moist, southern climates whereas upland ecotypes of switchgrass are generally tetraploid, hexaploid, or octoploid in nature and can be found growing in the northern part of the United States into southern Canada (Porter, 1966). Lowland ecotypes also typically produce more biomass and are more tolerant to diseases than their upland counterparts; however, upland ecotypes are generally more tolerant to drought and cold stresses (Porter, 1966; Stroup *et al.*, 2003).

Currently, industrial scale breeding programs of switchgrass have focused on optimizing biomass yield and improving feedstock quality in order to produce more biofuel (Bouton, 2007). However, this practice is likely to reduce the genetic diversity in switchgrass that promotes disease resistance. Airborne foliar fungal pathogens, like switchgrass rust, have potential to cause nationwide epidemics on switchgrass and have been shown to cause significant biomass yield losses (Sykes *et al.*, 2015). The causal agent of switchgrass rust, *Puccinia emaculata* Schw., is widespread and has been reported in Tennessee (Zale *et al.*, 2008), Arkansas (Hirsch *et al.*, 2010), Virginia (Frazier *et al.*, 2012), and Mississippi (Gilley *et al.*, 2013). The majority of switchgrass cultivars, including lowland and upland ecotypes, have been shown to be moderate to highly susceptible to this rust pathogen (Gustafson *et al.*, 2003; Uppalapati *et al.*, 2013).

The molecular mechanisms underlying tolerance to pathogen infection must be elucidated in order to understand and improve disease resistance in switchgrass. A recent report found that many genes were differentially expressed during switchgrass rust infection and that some of these genes belonged to the NB-LRR gene family, which is the largest family of plant disease resistance (*R*) genes (Serba *et al.*, 2015). Plant NB-LRR genes encode for proteins that contain a C-terminal leucine rich repeat (LRR) domain, a highly conserved central nucleotide binding (NB) domain, and a variable N-terminal region (Hammond-Kosack & Jones, 1997). The NB domain has been well characterized with regards to several preserved motifs including the Walker-A/P-loop, kinase 2, RNBS/kinase 3, and GLPL (Meyers *et al.*, 2003). Recent studies have suggested that the LRRs within the C terminal region may contain either a 'LxxLxxLxxLxLxx' signature (Jones & Jones, 1997) or a 'LxxLxxL' signature (Bryan *et al.*, 2000). Two major classes of NB-LRR genes have emerged based on conserved sequences within the N-terminal. These classes are characterized by the presence of either a coiled-coil (CC) motif or a domain homologous to the *Drosophila* Toll and mammalian Interleukin-1 Receptor (TIR) (Meyers *et al.*, 2003). To date, no TIR-NB-LRR genes have been identified in cereal grass species (Bai *et al.*, 2002).

Putative NB-containing *R* genes have been identified in numerous plant species by experimental methods, such as PCR cloning (Yu *et al.*, 1996) which has been used to identify potential RGHs in species such as Arabidopsis (Meyers *et al.*, 1999), rice (Bai *et al.*, 2002), and cotton (He *et al.*, 2004). Homology-based bioinformatics approaches have also been used to identify thousands of putative NB-containing *R* genes in plants, including several important crop species such as rice (Monosi *et al.*, 2004), potato (Lozano *et al.*, 2012), and soybean (Kang *et al.*, 2012). Alternatively, recent advances in next generation sequencing technologies have allowed

for preferential high throughput sequencing of *R* genes in a process known as Resistance gene enrichment Sequencing (Ren-Seq) (Jupe *et al.*, 2013). This method has been successfully used in potato and tomato to identify NB-LRR resistance genes (Jupe *et al.*, 2013).

Despite recent advances in determining potential *R* genes, few studies have been performed in switchgrass. One recent report by Zhu *et al* (2013) identified RGHs in switchgrass using a combination of PCR-cloning and EST database mining. The results of this study identified approximately 380 RGHs (Zhu *et al.*, 2013), which is roughly 0.39% of the total number of complete and incomplete genes in the switchgrass genome (v 1.1) (Goodstein *et al.*, 2012). Depending on the size of the plant genome, it is estimated that NB-LRR genes comprise between 0.5% - 1.8% of the protein coding genes (Meyers *et al.*, 2003; Mun *et al.*, 2009). Therefore, we anticipated that several hundred more RGHs should be present in the switchgrass genome. In this study, we used a homology-based computational method to identify 1,011 putative NB-containing RGHs in the switchgrass genome (v 1.1). We also identified several switchgrass RGHs that contained unique domains, including a jacalin-like lectin binding domain, a calmodulin-like binding domain, and a major sperm protein domain. Additionally, RNA-sequencing (RNA-seq) data from a rust-resistant cultivar, 'Alamo' (Frazier *et al.*, 2012), and a rust-susceptible cultivar, 'Dacotah' (Frazier *et al.*, 2012), were used to identify variants within the RGHs and to examine basal expression differences of these RGHs in uninfected leaf tissue. An analysis of RNA-seq datasets from the flag leaves of field-grown cultivar 'Summer' plants indicated that the expression levels of some RGHs were developmentally regulated. The results of this study significantly improve our understanding of disease resistance genes in this important biofuel crop.

MATERIALS AND METHODS

Switchgrass genome resources and identification of putative NB-LRR resistance genes

The switchgrass genome (*Panicum virgatum* v1.1, DOE-JGI) and its annotation resources were accessed from the DOE-JGI website (http://www.phytozome.net/panicumvirgatum.php) (Goodstein *et al.*, 2012). HMMER 3.0 (http://hmmer.janelia.org/software) (Finn *et al.*, 2011) and PfamScan (ftp://ftp.sanger.ac.uk/pub/databases/Pfam/Tools/) were employed to search for switchgrass genes that contained Pfam NB (NB-ARC) family (PF00931) domains. The search was conducted using switchgrass protein sequences that are based on the curated Pfam-A dataset (Pfam10.0) (Punta *et al.*, 2012) and the threshold was set to < 1E-10. To predict LRRs, a Perl script was developed that identified and counted the number of 'LxxLxxLxx' signatures in the C-terminal region of the switchgrass RGHs.

Structural analysis of the newly identified RGHs

Coils (v 2.2) (Lupas *et al.*, 1991), a program embedded in the InterProScan software (v 5.11) (Jones *et al.*, 2014), was employed to search the putative switchgrass RGHs for the presence of CC domains. SMART (Simple Modular Architecture Research Tool, http://smart.embl.de/), a widely used online resource that identifies and annotates protein domains and protein domain architectures (Letunic *et al.*, 2012), was used to search the RGHs for TIR domains. The SMART and Pfam databases were also used to detect other highly conserved unique domains that may be present in the RGHs. SignalP (v 4.0) (Petersen *et al.*, 2011), as part of InterProScan, was used to analyze the N-terminal region of the putative switchgrass RGHs for the presence of signal peptides. TargetP 1.1

(http://www.cbs.dtu.dk/services/TargetP/) was then employed to analyze the subcellular localization of the switchgrass RGHs that contained a signal peptide (Emanuelsson *et al.*, 2007). Finally, NLStradamus (Nguyen Ba *et al.*, 2009) was used to determine if any of the RGHs may contain a putative nuclear localization signal.

Phylogenetic analysis of switchgrass RGHs containing full-length NB domains

The amino acid sequences of the switchgrass RGHs were manually searched for the presence of four highly conserved motifs within the NB domain: the P-loop/WalkerA, the Kinase 2, the RNBS, and the GLPL (Meyers *et al.*, 2003). For those RGHs in which all four motifs could be identified, the amino acid sequences were extracted and compiled together. Clustal Omega (Sievers *et al.*, 2011) was used to align the sequences and Skylign (Wheeler *et al.*, 2014) was used to identify and determine the consensus sequences for the 4 conserved motifs mentioned above. To analyze the genetic diversity present in the switchgrass RGHs, the amino acid sequences of the full-length NB domains, starting from the beginning of the P-loop and ending at the GLPL motif, were compared to 116 full-length NB domains from *Brachypodium distachyon* (Tan & Wu, 2012). All of the sequences were aligned using ClustalW, as part of the MEGA software (v 6) program, and the default parameters (Tamura *et al.*, 2013). After alignment, a Maximum-likelihood phylogenetic tree was constructed in MEGA using the default parameters and a bootstrap value of 100.

RNA-seq analysis and variant detection between 'Alamo' and 'Dacotah'

RNA-seq datasets (Hupalo et al., unpublished) from un-inoculated leaf tissue of 'Alamo', a rust-resistant switchgrass cultivar, and 'Dacotah', a rust-susceptible switchgrass cultivar, were imported and analyzed using CLC Genomics Workbench (v 7.5) (Table 3-1). Raw reads that passed default quality scores and were greater than 50 bp in length were mapped to the switchgrass (v 1.1) reference genome using the RNA-seq analysis tool with the following parameters: mapping was also performed to intergenic regions, only one hit was allowed per read, a similarity fraction of 0.9, a length fraction of 1.0, a mismatch cost of 2, an insertion cost of 3, and a deletion cost of 3.

After read mapping, the reads that mapped were locally realigned and variants were called using the Basic Variant Detection tool with the following parameters: broken pairs were not ignored, non-specific matches were ignored, minimum read coverage of 3, minimum variant count of 2, and a minimum variant frequency (%) of 25.0. Following variant detection, the variant tracks for both 'Alamo' and 'Dacotah' were filtered against the RNA-seq mapped reads of the other cultivar, respectively, to identify variants uniquely different between them. High-quality variants, including single nucleotide polymorphisms (SNPs), multiple nucleotide polymorphisms (MNPs), insertions/deletions, and replacements were selected for based on the following criteria: zygosity = homozygous, Frequency = 100%, coverage $\geq$ 3, control count = 0, control coverage $\geq$ 3, and control frequency = 0%. The resulting lists from both variant detections were compiled together for a comprehensive list of high quality variants between the two cultivars.

Gene expression analysis of switchgrass RGHs in 'Alamo', 'Dacotah', and 'Summer'

In order to determine if any of the RGHs identified in this study are expressed basally in switchgrass leaves, the RNA-Seq datasets mentioned above for three biological replicates of 'Alamo' and 'Dacotah' were analyzed using CLC Genomics Workbench. Reads with a quality score (Q-score) >20 were mapped to the switchgrass reference genome (v 1.1). After read-mapping, the gene expression output tracks for all replicates of 'Alamo' and 'Dacotah' were compared using readily available tools within the program. A gene was considered expressed if five or more total reads mapped to it. Differential expression analysis between the two cultivars was carried out using an Empirical analysis of Differential Gene Expression (EDGE) test that is available as part of the CLC Genomics Workbench software. The results were filtered based on a False Discovery Rate (FDR) p-value of < 0.05 and a corrected fold change > 2.

We utilized the raw RNA-sequencing data obtained from flag leaves of field grown cv. 'Summer' plants (Palmer *et al.*, 2015) to analyze developmental gene expression of the 1,011 switchgrass RGHs over the course of a growing season. Only RGHs that exhibited detectable expression levels in all three biological replicates, and also in at least one of the five sampled time points, were kept for further analysis. Weighted Gene Co-expression Network Analysis (WGCNA, version 1.43) in R was then used to identify RGHs that demonstrated similar expression patterns (Zhang & Horvath, 2005; Langfelder & Horvath, 2008; RC, 2015). Signed co-expression networks were identified using WGCNA with a soft threshold value of 16 and a minimum module size of 30. The module eigengene (ME), which corresponds to the first principal component of a specific module, represents the expression pattern of each co-expression module.

Isolation of DNA from 'Alamo' and 'Dacotah'

DNA was isolated from young leaf tissue using a modified CTAB method (Porebski *et al.*, 1997). The DNA was re-suspended in 1x TE buffer (10 mM Tris pH 8.0, 10 mM EDTA pH 8.0) and the quantity and quality was measured using a Nanodrop ND-1000 (Wilmington, DE, USA).

Validation of SNPs using allele-specific PCR primers and DNA sequencing

Allele-specific SNP PCR primers for both 'Alamo' and 'Dacotah' were designed as previously described (Hirotsu *et al.*, 2010). Both the forward and reverse primers for each locus accounted for a SNP at the 3' end. The 3rd base pair from the 3' end of each primer was changed according to the recommendations from Hirotsu *et al* (2010) in order to maximize allele-specificity. Four sets of forward and reverse primers that detected 16 SNPs were designed and used for PCR amplification in a total reaction volume of 20 µL (Table 3-2 and Table 3-3). After the reactions were completed, the products were separated by gel electrophoresis on a 1.0% agarose gel and visualized using a GelDoc system (Bio-Rad, Hercules, CA, USA).

Conserved PCR primers were also designed to amplify DNA fragments containing SNPs between 'Alamo' and 'Dacotah' (Table 3-4). These SNPs were validated using traditional DNA sequencing. PCR reactions were performed in a total volume of 30 µL and contained the following components: 15 µL of high fidelity iProof (Bio-Rad, Hercules, CA, USA), 11 uL of ddH_2O, 2 µL of 200 ng/µL DNA, 1 µL of 10 µM forward primer, and 1 µL of 10 µM reverse primer. The conditions for the PCR reactions were: 98°C for 3 min, followed by 30 cycles of denaturation at 98°C for 30 sec, annealing at 60°C for 45 sec, extension at 72°C for 1 min and 30

sec, and a final extension at 72°C for 7 min. Once the PCR reactions finished, 2 µL of each reaction was run on a 0.8% agarose gel in order to verify amplification. Next, the remaining 28 µL of each reaction was purified. Finally, all purified PCR products were sent for DNA sequencing.

RESULTS

Identification of 1,011 putative NB-containing RGHs in switchgrass

In this study, a total of 1,542 switchgrass proteins were identified that contained *R* gene-like NB-ARC domains. After selecting for proteins with an NB-ARC e-value of <1E-10 and removing alternatively spliced transcripts, 1,011 unique switchgrass proteins were collected that contained *R* gene-like NB-ARC domains (Figure 3-1). This accounts for approximately 1.03% of the total protein-coding genes in switchgrass. From here on, these will be referred to as putative switchgrass RGHs. The putative RGH proteins ranged in length from 98 amino acids (Pavir.J31808) to 1,649 amino acids (Pavir.Ba02898) with an average length of 693 amino acids. Of the 1,011 RGH proteins, 695 were annotated as complete protein sequences in the switchgrass genome (v 1.1), meaning they contained a start methionine (ATG) and a stop codon. The remaining 316 RGH proteins lacked one or both of these features.

Since NB-LRR disease resistance genes tend to cluster together in plant genomes (Meyers *et al.*, 2003), the availability of a draft reference genome for switchgrass provided an opportunity to analyze the chromosomal distribution of the 1,011 RGHs. The current version of the switchgrass genome (v 1.1) is comprised of 18 main pseudomolecules that represent the A and B subgenomes (Goodstein *et al.*, 2012). An additional several hundred thousand sequences

are located on unanchored contigs (Goodstein *et al.*, 2012). Of the 1,011 RGHs identified in this study, 511 were assigned to one of the major pseudomolecules while the remaining 500 were dispersed among unanchored contigs (Figure 3-2). Switchgrass chromosome 8 was found to contain the most RGHs with 92 genes located on Chr08a and 93 genes located on Chr08b. In contrast, chromosomes 4 and 7 contained the least total number of RGHs with 20 and 21, respectively (Figure 3-2).

Structural analysis of the 1,011 newly identified RGHs in switchgrass

The two major classes of NB-LRR disease resistance genes contain either a coiled-coil (CC) motif or a domain homologous to the intracellular signaling domain of *Drosophila* Toll and mammalian Interleukin-1 Receptors (TIR) in their N-terminals (Hammond-Kosack & Jones, 1997). A total of 405 genes was discovered that contain putative CC domains (Figure 3-2). In contrast to the CC domains, no TIR domains were detected in any of the 1,011 RGHs.

Using SignalP, we determined that 19 of the 1,011 switchgrass RGHs contained an N-terminal signal peptide. None of these signal peptides were predicted to span a cellular membrane. Based on the amino acid sequence of the signal peptide, TargetP (v 1.1) was used to predict the subcellular localization of these proteins. Of the 19 proteins, 13 were predicted to enter the secretory pathway and four were predicted to go to the mitochondria. The remaining proteins, Pavir.Ea02039.1 and Pavir.Ba00602.1, were predicted to go to other subcellular locations that were not the mitochondria or the chloroplast. Several NB-LRR disease resistance proteins, such as RRS1 (Deslandes *et al.*, 2003), have been shown to localize in the plant nucleus upon pathogen detection. NLStradamus was employed to search the 1,011 RGHs for the

presence of putative nuclear localization signals (NLSs). A total of 104 of the 1,011 RGHs were identified to contain a putative NLS. Of these proteins, the NLS was predicted to be in the N-terminal region in 72 of the RGHs, while 32 of the proteins were predicted to contain an NLS in their C-terminal region.

The majority of NB-containing resistance genes contain highly variable leucine rich repeats (LRRs) in their C-terminals. LRRs are believed to function in protein-protein interactions, as well as in ligand binding (Hammond-Kosack & Jones, 1997). The 1,011 switchgrass RGH proteins were manually screened to identify a consensus sequence (LxxLxxLxx) that was predominant among the RGH proteins. A Perl script determined that a total of 682 of the 1,011 switchgrass RGHs contained one or more 'LxxLxxLxx' signatures downstream of the end of the NB-ARC domain (Figure 3-1). The number of 'LxxLxxLxx' signatures identified ranged from one to 19 per protein with the majority of these proteins containing between five and nine (445 or 65.2%). Therefore, the majority of the switchgrass RGHs do in fact have the typical NB-LRR gene structure.

Identification of unique domains in the 1,011 switchgrass RGHs

Previous reports have identified unique domains in the N- and C-terminals of NB-LRR resistance genes and have suggested that these domains may play a role in *R* gene function (Cesari *et al.*, 2014). As summarized in Table 3-5, 40 switchgrass NB-LRR proteins were also predicted to contain other known functional domains. These unique domains can be classified into seven different functional categories: protein modification, DNA binding/transcription, protein trafficking and vesicle movement, protein-protein interaction, sugar binding, signal

transduction, and transposable element. The majority of switchgrass RGHs with unique domains (24 out of 40, or 60%) fell into the protein modification category. Of these 24 proteins, two RGHs contained a thioredoxin domain (Pavir.Fa01782 and Pavir.Hb00484), one RGH contained a glutaredoxin domain (Pavir.Bb01048), and one RGH contained a phosphatase domain (Pavir.J24356). A putative C-terminal NLS was also detected for Pavir.Fa01782. Interestingly, the remaining 20 RGHs were found to contain a protein kinase domain (Table 3-5). Of these 20 proteins, ten are located on chromosome 8 (five on Chr08a and the other five on Chr08b). Two of these protein kinase-containing RGHs, Pavir.Ha00561 and Pavir.Ha01108, were also found to contain a NLS in their C-terminal and N-terminal, respectively.

The next largest category included RGHs with domains that function in DNA binding and transcription. Two switchgrass RGHs, Pavir.Ba02315 and Pavir.J20878, were found to contain an N-terminal B3 DNA binding domain. Interestingly, Pavir.J20878 is also predicted to contain a C-terminal WRKY domain. Five other RGHs including Pavir.Fa02339, Pavir.Ga00028, Pavir.Gb00931, Pavir.J19380, and Pavir.J40131 were also found to have an N-terminal zinc finger-BED DNA binding domain (Table 3-5).

Finally, the remaining eight switchgrass RGHs fell into the last five categories. Three switchgrass RGHs (Pavir.Ib02384, Pavir.Ib02433, and Pavir.J18369) were predicted to contain an N-terminal domain homologous to major sperm protein (PF00635) (Table 3-5). These proteins are classified in the protein trafficking and vesicle movement category. Two switchgrass RGHs were predicted to contain domains that may play a role in protein-protein interactions (Table 3-5). These include Pavir.J03445, which has a C-terminal WD40 domain, and Pavir.Fb01504, which contained a C-terminal HAT family dimerization region that is common to transposable elements. Pavir.J03445 was also predicted to contain an N-terminal nuclear

localization signal. Another element that is found in some transposons, a gag-polypeptide of LTR copia-type domain, was predicted in the C-terminal of Pavir.Aa01444. The final two RGHs with unique domains, Pavir.Hb01174, which contains a C-terminal jacalin-like lectin binding domain, and Pavir.Hb00190, which has calmodulin binding protein-like domain in the C-terminal, are predicted to function in sugar binding and signal transduction, respectively.

Phylogenetic analysis of switchgrass sequences containing a full-length NB domain

Of the 1,011 switchgrass RGHs, 578 were found to contain a full-length NB domain in which four highly conserved motifs could be found: the Walker-A/P-loop, the Kinase 2, the Kinase 3/RNBS-B, and the GLPL. These 578 sequences were then aligned to determine the consensus sequences for these four motifs (Figure 3-3). The consensus sequences for the four motifs are as follows: 1) Walker-A/P-loop = GxGGxGKT, 2) Kinase 2 = KR(Y/F)L(I/L)VLDD(V/L)W, 3) Kinase 3/RNBS-B = SR(I/V)(I/L)VTTR, and 4) GLPL = GLPLA. These results are consistent with similar sequences identified in NB-LRR genes in other plant species, such as rice (Zhou *et al.*, 2004) and *Brachypodium* (Tan & Wu, 2012).

The amino acid sequences of the NB domains of the 578 RGHs with full-length NB domains, along with amino acid sequences of the NB domains of 116 *Brachypodium* NB-LRR genes (Tan & Wu, 2012), were extracted and subjected to phylogenetic tree analysis. The un-rooted phylogenetic tree can be divided into 35 different groups, with the majority of groups containing both switchgrass and *Brachypodium* NB-LRR genes (Figure 3-4). Switchgrass and *Brachypodium* diverged about 50 million years ago (Opanowicz *et al.*, 2008). However, most NB-LRR genes are still conserved in the two plant species, since no clear separation of

switchgrass or *Brachypodium* genes clusters were observed (Figure 3-4). This is also supported by strong bootstrap values (>50) of many clades that included sequences from both species.

Switchgrass NB-LRR genes are enriched on chromosome 8 (Figure 3-2). Therefore, we examined if chromosome 8 NB-LRR genes (highlighted in red) tended to cluster together in the phylogenetic tree (Figure 3-4). As shown in Figure 3-4, four small clusters of chromosome 8 NB-LRR genes are evident. These clusters were further analyzed to determine if the ten chromosome 8 RGHs that were predicted to contain a protein kinase domain were phylogenetically similar. Seven of the ten protein kinase-containing RGHs (highlighted in green) were located in the same cluster, indicating that these genes are highly homologous. These genes could have rapidly evolved from the same ancestor and duplicated on chromosome 8. Interestingly, no *Brachypodium* genes were located in this cluster. The remaining three protein kinase-containing RGHs were dispersed across the phylogenetic tree with two of the genes clustering together in the same group as the bigger cluster. A similar trend of genes on the same chromosome clustering together was also observed for the *Brachypodium* genes (Figure 3-4).

Identification of variants between RGHs in 'Alamo', a rust-resistant switchgrass cultivar, and 'Dacotah', a rust-susceptible switchgrass cultivar

The RNA-seq mapping statistics obtained from the analysis of these datasets is shown in Table 3-2. Approximately 99,283,683 'Dacotah' reads (82.03%) and 180,996,166 'Alamo' reads (84.35%) were mapped to the reference genome (Table 3-1). The slightly higher number of 'Alamo' reads mapping to the reference genome could be attributed to the fact that the switchgrass reference genome was sequenced from DNA of the cv. 'Alamo'.

After alignment to the switchgrass reference genome (v 1.1), the RNA-seq reads for both 'Alamo' and 'Dacotah' were analyzed in order to identify variants, including SNPs, MNPs, insertions/deletions (indels), and replacements in the RGHs. A total of 23,156 variants was found in 781 RGHs between 'Alamo' and 'Dacotah'. After filtering for high quality homozygous variants, 2,634 variants were found in 344 RGHs. These variants include 2,347 SNPs, 136 MNPs, 145 indels, and 6 replacements.

Allele-specific PCR is a high throughput and cost-effective way of distinguishing SNPs at a particular locus without the need for DNA sequencing (Hirotsu *et al.*, 2010). A total of eight primer pairs, four for 'Alamo' and four for 'Dacotah', was designed to validate 16 SNPs identified by RNA-seq analysis (Table 3-2 and Table 3-3). The results showed that the primers designed to detect the 'Dacotah' alleles were more specific to 'Dacotah' DNA than the primers designed to detect the 'Alamo' alleles (Figure 3-5). The presence of unwanted PCR products using these allele-specific primers suggests that the primers may be amplifying unwanted PCR products. Alternatively, the other allele may actually be present in each DNA sample but is either not expressed or expressed at a level not captured by the sequencing depth of this experiment and thus, was not detected in our RNA-sequencing data.

Traditional PCR amplification and DNA sequencing with conserved primers was also used to validate SNPs identified from RNA-seq analysis. Eight different primer sets were designed to detect 12 SNPs within eight putative RGHs (Table 3-4). DNA sequencing confirmed the presence of 11 out of the 12 SNPs. For the one SNP in Pavir.Ib01513 that could not be validated, the PCR products for the 'Alamo' and 'Dacotah' sequences differed greatly from the expected DNA sequence, indicating that this primer set may have amplified unwanted targets.

Analysis of gene expression of the 1,011 switchgrass RGHs between 'Alamo' and 'Dacotah'

It has been suggested that NB-LRR disease resistance genes are basally expressed in plant tissues (McHale *et al.*, 2006). Upon pathogen detection, the expression of these genes is up-regulated in order to initiate defense responses (Zipfel *et al.*, 2004). RNA-sequencing reads from three biological replicates of 'Alamo' and 'Dacotah' (Table 3-1) were used to determine if any of the 1,011 switchgrass RGHs were basally expressed in un-inoculated leaf tissue. Of the 1,011 RGHs, 338 were expressed in both cultivars. For the individual cultivars, 117 RGHs were found to be expressed only in 'Alamo' and 134 RGHs were found to be expressed only in 'Dacotah'.

Differential expression of NB-LRR resistance genes has been shown to play a role in plant disease response. The results of gene expression analysis were filtered such that the 338 genes expressed in both cultivars were considered significantly differentially expressed if their RPKM values had a fold change greater than 2 and an FDR p-value less than 0.05. Using these criteria, 21 genes were found to be significantly differentially expressed between the two cultivars (Table 3-6). Overall, the results of this analysis suggest that these two different cultivars basally express RGHs.

Expression of specific switchgrass NB-LRR genes are regulated by leaf developmental stages

Recent studies have suggested that NB-LRR resistance genes are also differentially regulated during various developmental stages (Century *et al.*, 1999; Cao *et al.*, 2007). RNA-seq data from the flag leaves of field grown cv. 'Summer' switchgrass plants, collected over five time points from July to September of 2012 (Palmer *et al.*, 2015), were obtained and analyzed

for NB-LRR gene expression. Approximately 755 of the 1,011 RGHs were found to be expressed in all three biological replicates from at least one time point. These 755 RGHs were selected for Weighted Gene Correlation Network Analysis (WGCNA) which classified these genes into eight co-expression modules (Figure 3-6). The first module contained the most RGHs (190 members) which had the highest expression at the end of July, when the flag leaves were fully expanded (Figure 3-6a). The second module consisted of RGHs (146 members) with peak gene expression at the third time point (8/19, Figure 3-6b). The RGHs that are co-expressed in module 3 (99 members) displayed greater variance in gene expression (Figure 3-6c). For the most part, peak gene expression was seen at the fourth time point (8/31) for all three samples; however, some minor expression in one sample collected at time point 1 (7/03) can be observed. Modules 4 (92 members) and 6 (56 members) are comprised of RGHs that demonstrated the highest level of gene expression at the first sampled time point (7/03) (Figure 3-6d and 3-6f, respectively). Modules 4 and 6 differ in that module 4 showed an additional minor expression at time point four (8/31) while module 6 showed an additional expression at time point three (8/16). The fifth and eighth modules were comprised of RGHs (85 and 34 members respectively) that demonstrated gene expression at time point five (9/19; Figure 3-6e and 6h, respectively). The primary difference between these two modules was seen in module 5 showing reduced expression in 9/19 sample three and minor expression in 8/16 sample one while module 8 had high expression in 9/19 sample three. Finally, the members of module 7 (53 members) presented rather consistent expression across two sequential harvest dates (7/27 and 8/16) in all biological replicates (Figure 3-6g).

DISCUSSION

The identification and validation of plant disease resistance genes is a major focus in the molecular investigations of plant-pathogen interactions. While other studies have aimed to understand the molecular mechanisms controlling switchgrass resistance to switchgrass rust (Zhu *et al.*, 2013; Serba *et al.*, 2015), none of these studies has mined the currently available switchgrass genome (v 1.1) for potential NB-LRR resistance gene homologs. In this research, a homology-based computational approach was used to identify 1,011 unique RGHs in the switchgrass genome. Approximately 266% more RGHs were identified than in a similar study that detected switchgrass RGHs from EST sequences and PCR-based cloning (Zhu *et al.*, 2013). However, the total number of RGHs in switchgrass may change with further refinement of the switchgrass genome; although, the percentage of RGHs identified in this study is similar to rice (Singh *et al.*, 2015).

Structural analysis of the 1,011 switchgrass RGHs provided useful insights into their putative molecular functions. Almost all of the major features expected in plant NB-LRRs were identified (Meyers *et al.*, 1999; McHale *et al.*, 2006), indicating a robust immunity potential is present in switchgrass. However, further studies are needed to validate the sub-cellular localization and functions of these proteins.

The NB domain of plant R protein has been shown to act like a molecular switch and function in signal transduction pathways. 578 of the 1,011 switchgrass RGHs contained a full-length NB domain while the remaining sequences lacked one or more highly conserved motifs (P-loop/WalkerA, Kinase 2, RNBS, and GLPL; (Meyers *et al.*, 2003). This could be explained by the assembly of the switchgrass genome (v 1.1), which consists of 18 pseudomolecules and

several thousand additional contigs ranging in size from 1,000 to 88,021 bp (Goodstein *et al.*, 2012). Incomplete NB-LRR gene sequences in switchgrass could be a result of incomplete duplications or transversions, incomplete assembly or annotations, or may actually be pseudogenes. Some NB-LRR pseudogenes have been shown to code for non-functional or truncated protein products (Marone *et al.*, 2013). Interestingly, the evolution of a pseudogene at the *Pid3* gene locus has been found to promote disease in susceptible rice cultivars (Shang *et al.*, 2009).

Sequence duplication and divergence is also prominent in NB-LRR genes. Phylogenetic analysis of the switchgrass and *Brachypodium* NB-LRRs found that the majority of the *Brachypodium* NB-LRRs have been conserved in switchgrass. Several switchgrass RGHs, including ten that were identified to contain a protein kinase domain, however, appear to have emerged after the two species diverged.

Unique protein domains other than the NB-LRRs were identified in 40 of the switchgrass RGHs in this study. Some of these domains have also been identified in the NB-LRRs of other plant species, and could play roles similar to ones reported recently in Arabidopsis. In Arabidopsis the *RRS1-R* gene encodes NB-LRR protein that contains a WRKY domain that acts as a decoy to intercept effector molecules secreted by *Pseudomonas syringae* pv. *pisi* and *Ralstonia solanacearum* (Sarris *et al.*, 2015).

Two switchgrass RGHs, Pavir.Fa01782 and Pavir.Hb00484, were found to have a C-terminal domain homologous to thioredoxin proteins. One switchgrass RGH, Pavir.Bb01048, was predicted to have a C-terminal glutaredoxin domain. Thioredoxin and glutaredoxin proteins participate in oxidation/reduction reactions and have been associated with increased disease resistance in tobacco (Tada *et al.*, 2008; Sun *et al.*, 2010) and increased disease susceptibility in

Arabidopsis (La Camera *et al.*, 2011). Therefore, the NB-LRR containing a thioredoxin domain may function in disease resistance by reducing pathogen-induced oxidative stresses. Since glutaredoxin has been shown to promote disease resistance (La Camera *et al.*, 2011), the presence of a glutaredoxin domain in a NB-LRR disease resistance gene may function as a decoy similar to the data described by Sarris *et al* (2015).

A total of five switchgrass RGHs were predicted to contain unique domains that are involved in DNA binding. One switchgrass gene, Pavir.J03445, was found to contain a WD40 domain in the C-terminal. Plant genes that contain WD40 domains have been shown to be differentially regulated during pathogen infection, suggesting that these genes may be important regulators of defense-related responses (Narusaka *et al.*, 2004; Wang *et al.*, 2012a). Another switchgrass gene, Pavir.J20878, which contains an N-terminal B3 DNA-binding domain, was also found to contain a C-terminal WRKY DNA-binding domain. This further supports a role for Pavir.J20878 in DNA binding and transcription regulation.

Aside from DNA-binding, several other smaller categories were identified. One switchgrass RGH, Pavir.Hb01174, is predicted to function in sugar binding. This particular RGH contains a C-terminal jacalin-like lectin binding domain. Jacalin-like lectin domains bind carbohydrates, mainly mannose and galactose, and have been shown to play an important role in disease resistance (Xiang *et al.*, 2011). For example, the *RTM1* gene of Arabidopsis encodes a protein that contains a jacalin-like lectin domain and this protein is critical for inhibiting long-distance movement of the tobacco etch virus (Chisholm *et al.*, 2000). Additionally, three switchgrass RGHs, Pavir.Ib02384, Pavir.Ib02433, and Pavir.J18369, are predicted to function in protein trafficking and vesicle movement. These proteins contain a domain in their N-terminals that shows strong homology to a major sperm protein of nematodes. A previous report has

identified a similar domain in the N-terminal region of the VAP27 protein of tomato, which has been shown to interact with the Cf9 resistance protein; however, no direct role for VAP27 in disease response has been established (Laurent *et al.*, 2000). To our knowledge, this is the first report of a major sperm protein domain attached to the N-terminal of a NB-containing resistance gene.

Several other domains found in the switchgrass RGHs, have been linked to disease response in other plants. These include calmodulin (Pavir.Hb00190) (Chiasson *et al.*, 2005; Takabatake *et al.*, 2007), WD domain (Pavir.J03445) (Xu & Min, 2011; Biruma *et al.*, 2012), and transposable elements (Pavir.Fb0150) (Essers *et al.*, 2000; Kang *et al.*, 2012). This highlights the diversity and potential for gene diversification of RGHs encoded by the switchgrass genome.

The switchgrass cultivars ‘Alamo’ and ‘Dacotah’ exhibit significantly different disease responses after exposure to switchgrass rust (*Puccinia emaculata*). ‘Alamo’ is relatively resistant to the rust pathogen whereas ‘Dacotah’ is highly susceptible (Frazier *et al.*, 2012). Polymorphisms within RGHs may contribute to the disease resistance phenotype that is observed between ‘Alamo’ and ‘Dacotah’. In our study, we identified 2,634 variants between ‘Alamo’ and ‘Dacotah’, including SNPs, MNPs, indels, and replacements. Approximately 89% of the variants detected were SNPs. The predominance of SNPs could be attributed to the fact that single nucleotide changes in the coding regions of genes are less likely to disrupt the reading frame, which often results in nonsense mutations. SNPs could also explain the disease response phenotypes of ‘Alamo’ and ‘Dacotah’, as they could alter the amino acid sequence of resistance genes and potentially disrupt gene function. Since these SNPs are associated with defense-related

genes, they could be further developed into molecular markers for use in breeding of disease resistance.

In addition to polymorphisms within the RGHs, differential expression of NB-LRR disease resistance genes may contribute to the different phenotypes observed between the two cultivars. It is believed that *R* genes are expressed at relatively low levels in unchallenged plant cells in anticipation of pathogen attack (Hammond-Kosack & Jones, 1997). Indeed, we found that 338 RGHs displayed expression evidence in both cultivars in an unchallenged state, supporting the idea that *R* genes are basally expressed in healthy plant cells (Hammond-Kosack & Jones, 1997). There could be several reasons for the genes that we could not find expression evidence for. First, these genes could be expressed either at extremely low levels in healthy plant cells or at a different developmental stage and thus, they escaped detection at the sequencing coverage and sampling time used in this study. Second, the expression of these genes could be induced upon pathogen detection and they are not basally expressed in healthy plant cells. Finally, some of these genes may be pseudogenes and may not be expressed under any conditions. Further studies are needed to evaluate the expression patterns of the switchgrass RGHs and to determine the exact role, if any, that these genes play in switchgrass disease response.

Developmental regulation of specific RGHs could also contribute to disease resistance phenotypes at different stages of plant growth. RNA-sequencing data from the flag leaves of field grown cultivar ‘Summer’ provided a unique opportunity to examine switchgrass RGH expression over the course of a growing season (Palmer *et al.*, 2015). The first group of developmentally regulated RGHs (module 1, Figure 3-6a) is of particular interest since these genes are up-regulated at the end of July. The end of July and the beginning of August are

optimal times for switchgrass rust infection and thus, these genes may play an important role in immediate defense responses against foliar pathogens like switchgrass rust. Correspondingly, the transcripts for these RGHs decreased over the remaining harvests, supporting the idea that these genes are involved in the earlier stages of disease response. Field-grown switchgrass plants appear to be more susceptible to switchgrass rust as they begin to flower and set seed (data not published). As displayed in modules 5 and 8, 119 of the 755 RGHs (16% of the genes) exhibited peak gene expression in at least one biological replicate during the last sampling point (9/19). The remaining 84% of the genes displayed peak gene expression over the first four sampled time points. Since fewer RGHs showed preferential expression during the later stages of the growing season, these results support the likelihood that switchgrass plants may utilize resources towards other processes, such as flowering and nutrient remobilization, rather than disease resistance in the later stages of development.

CONCLUSION

Our results provide useful insight into the molecular structure, distribution, and expression patterns of members of the NB-LRR gene family in the biofuel crop switchgrass. The homology-based computational method used in this study found approximately 266% more RGHs in switchgrass than what was previously identified. The large number of putative RGHs in switchgrass indicates that it has the potential to initiate a robust defense response against pathogens, like switchgrass rust. In addition, the SNPs identified in this study between the rust-resistant cultivar ‘Alamo’ and the rust-susceptible cultivar ‘Dacotah’ could be developed into molecular markers for use in switchgrass breeding programs. We also identified RGHs that may

contain unique domains such as calmodulin-liking binding, jacalin, and major sperm protein, which has not been previously reported in any other plant species. Recently, it has been shown that unique domains attached to NB-LRR disease resistance genes may act as decoys to intercept pathogen effector molecules and initiate defense responses. More studies are needed to determine the exact role of these genes in switchgrass disease resistance. Finally, we demonstrate that the expression patterns of 755 RGHs are developmentally regulated. To our knowledge, this is the first report that has characterized the developmental regulation of NB-LRR gene on a whole genome-scale level. The results of this study provide valuable insights into the mechanisms governing plant defense and significantly improve our understanding of disease resistance genes in this important biofuel crop.

ACKNOWLEDGEMENTS

The project was supported by USDA-NIFA Grant Number 2011-67009-30133. The project was also partially supported by a Virginia Tech CALS integrative grant and the Virginia Agricultural Experiment Station (VA135872). Work performed by the ARS was supported in part by the Office of Science (BER), U. S. Department of Energy Grant Number DE-AI02-09ER64829, USDA-NIFA Grant Number 2011-67009-30096, and by the USDA-ARS CRIS project 3042-21000-030-00D. The U.S. Department of Agriculture, Agricultural Research Service, is an equal opportunity/affirmative action employer and all agency services are available without discrimination. Mention of commercial products and organizations in this manuscript is solely to provide specific information. It does not constitute endorsement by USDA-ARS over other products and organizations not mentioned. The work conducted by the US Department of

Energy Joint Genome Institute is supported by the Office of Science of the US Department of Energy under contract number DE-AC02-05CH11231.

AUTHOR CONTRIBUTIONS

B. Zhang, GS, and B. Zhao designed the research projects. TF, FX, and AB performed the computational identification and characterization of the RGHs. TF isolated ‘Alamo’ and ‘Dacotah’ DNA and did RNA-seq analysis and SNP validation. TDR isolated RNA from the ‘Summer’ flag leaf samples. NP, CT, and TDR executed the research and data analysis of the ‘Summer’ flag leaf samples. NP performed WGCNA analysis of the ‘Summer’ flag leaf RNA-seq data. KC performed the RNA-seq of the three biological replicates of ‘Alamo’ and ‘Dacotah’. SS, JJ, and JS carried out sequencing, assembly, and annotation of the switchgrass genome. TF, NP, GS, and B.Zhao wrote the manuscript.

REFERENCES

Bai J, Pennill LA, Ning J, Lee SW, Ramalingam J, Webb CA, Zhao B, Sun Q, Nelson JC, Leach JE, Hulbert SH. 2002. Diversity in nucleotide binding site–leucine-rich repeat genes in cereals. *Genome Res.* **12**: 1871-1884.

Biruma M, Martin T, Fridborg I, Okori P, Dixelius C. 2012. Two loci in sorghum with NB-LRR encoding genes confer resistance to *Colletotrichum sublineolum*. *Theor. Appl. Genet.* **124**: 1005-1015.

Bouton JH. 2007. Molecular breeding of switchgrass for use as a biofuel crop. *Curr. Opin. Genet. Dev.* **17**: 553-558.

Bryan GT, Wu KS, Farrall L, Jia Y, Hershey HP, McAdams SA, Faulk KN, Donaldson GK, Tarchini R, Valent B. 2000. A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. *Plant Cell* **12**: 2033-2046.

Cao Y, Ding X, Cai M, Zhao J, Lin Y, Li X, Xu C, Wang S. 2007. The expression pattern of a rice disease resistance gene *Xa3/Xa26* is differentially regulated by the genetic backgrounds and developmental stages that influence its function. *Genetics* **177**: 523-533.

Century KS, Lagman RA, Adkisson M, Morlan J, Tobias R, Schwartz K, Smith A, Love J, Ronald PC, Whalen MC. 1999. Developmental control of *Xa21*-mediated disease resistance in rice. *Plant J.* **20**: 231-236.

Cesari S, Bernoux M, Moncuquet P, Kroj T, Dodds PN. 2014. A novel conserved mechanism for plant NLR protein pairs: the "integrated decoy" hypothesis. *Front. Plant Sci.* **5**: 606.

Chiasson D, Ekengren S, Martin G, Dobney S, Snedden W. 2005. Calmodulin-like proteins from *Arabidopsis* and tomato are involved in host defense against *Pseudomonas syringae* pv. *tomato*. *Plant Mol. Biol.* **58**: 887-897.

Chisholm ST, Mahajan SK, Whitham SA, Yamamoto ML, Carrington JC. 2000. Cloning of the *Arabidopsis RTM1* gene, which controls restriction of long-distance movement of tobacco etch virus. *Proc. Natl. Acad. Sci. USA* **97**: 489-494.

Deslandes L, Olivier J, Peeters N, Feng DX, Khounlotham M, Boucher C, Somssich I, Genin S, Marco Y. 2003. Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. *Proc. Natl. Acad. Sci. USA* **100**: 8024-8029.

Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2**: 953-971.

Essers L, Adolphs RH, Kunze R. 2000. A highly conserved domain of the maize *Activator* transposase is involved in dimerization. *Plant Cell* **12**: 211-224.

Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**: W29-37.

Frazier T, Shen Z, Zhao B, Bush E. 2012. First report of *Puccinia emaculata* infection on switchgrass in Virginia. *Plant Dis.* **97**: 424-424.

Gilley M, Tomaso-Peterson M, Orquera G, Marek S. 2013. First report of rust caused by *Puccinia emaculata* on cultivated switchgrass in Mississippi. *J. Miss. Acad. Sci.* **58**: 197.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**: D1178-D1186.

Gunter LE, Tuskan GA, Wullschleger SD. 1996. Diversity among populations of switchgrass based on RAPD markers. *Crop Sci.* **36**: 1017-1022.

Gustafson DM, Boe A, Jin Y. 2003. Genetic variation for *Puccinia emaculata* infection in switchgrass *Crop Sci.* **43**: 755-759.

Hammond-Kosack KE, Jones JDG. 1997. Plant disease resistance genes. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**: 575-607.

He L, Du C, Covaleda L, Xu Z, Robinson AF, Yu JZ, Kohel RJ, Zhang HB. 2004. Cloning, characterization, and evolution of the NBS-LRR-encoding resistance gene analogue family in polyploid cotton (*Gossypium hirsutum* L.). *Mol. Plant-Microbe Interact.* **17**: 1234-1241.

Hirotsu N, Murakami N, Kashiwagi T, Ujiie K, Ishimaru K. 2010. Protocol: a simple gel-free method for SNP genotyping using allele-specific primers in rice and other plant species. *Plant Met.* **6**: 12.

Hirsch RL, TeBeest DO, Bluhm BH, West CP. 2010. First report of rust caused by *Puccinia emaculata* on switchgrass in Arkansas. *Plant Dis.* **94**: 381-381.

Jones DA, Jones J. 1997. The role of leucine-rich repeat proteins in plant defences. *Adv. Bot. Res.* **24**: 89-167.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236-1240.

Jupe F, Witek K, Verweij W, Śliwka J, Pritchard L, Etherington GJ, Maclean D, Cock PJ, Leggett RM, Bryan GJ, Cardle L, Hein I, Jones JDG. 2013. Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.* **76**: 530-544.

Kang YJ, Kim KH, Shim S, Yoon MY, Sun S, Kim MY, Van K, Lee S-H. 2012. Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. *BMC Plant Biol.* **12**: 139.

La Camera S, L'Haridon F, Astier J, Zander M, Abou-Mansour E, Page G, Thurow C, Wendehenne D, Gatz C, Metraux JP, Lamotte O. 2011. The glutaredoxin ATGRXS13 is required to facilitate *Botrytis cinerea* infection of *Arabidopsis thaliana* plants. *Plant J.* **68**: 507-519.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**: 559.

Laurent F, Labesse G, de Wit P. 2000. Molecular cloning and partial characterization of a plant VAP33 homologue with a major sperm protein domain. *Biochem. Biophys. Res. Commun.* **270**: 286-292.

Letunic I, Doerks T, Bork P. 2012. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* **40**: D302-D305.

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

Lupas A, Van Dyke M, Stock J. 1991. Predicting coiled coils from protein sequences. *Science* **252**: 1162-1164.

Marone D, Russo M, Laidò G, De Leonardis A, Mastrangelo A. 2013. Plant nucleotide binding site–leucine-rich repeat (NBS-LRR) genes: active guardians in host defense responses. *Int. J. Mol. Sci.* **14**: 7302-7326.

McHale L, Tan X, Koehl P, Michelmore RW. 2006. Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* **7**: 212.

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

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. 2003. Genome-Wide Analysis of NBS-LRR–Encoding Genes in Arabidopsis. *The Plant Cell* **15**: 809-834.

Monosi B, Wisser RJ, Pennill L, Hulbert SH. 2004. Full-genome analysis of resistance gene homologues in rice. *Theor. Appl. Genet.* **109**: 1434-1447.

Mun J-H, Yu H-J, Park S, Park B-S. 2009. Genome-wide identification of NBS-encoding resistance genes in *Brassica rapa*. *Mol. Genet. Genomics* **282**: 617-631.

Narasimhamoorthy B, Saha MC, Swaller T, Bouton JH. 2008. Genetic diversity in switchgrass collections assessed by EST-SSR markers. *BioEnergy Res.* **1**: 136-146.

Narusaka Y, Narusaka M, Park P, Kubo Y, Hirayama T, Seki M, Shiraishi T, Ishida J, Nakashima M, Enju A, Sakurai T, Satou M, Kobayashi M, Shinozaki K. 2004. *RCH1*, a locus in *Arabidopsis* that confers resistance to the hemibiotrophic fungal pathogen *Colletotrichum higginsianum*. *Mol. Plant-Microbe Interact.* **17**: 749-762.

Nguyen Ba AN, Pogoutse A, Provart N, Moses AM. 2009. NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics* **10**: 202.

Opanowicz M, Vain P, Draper J, Parker D, Doonan JH. 2008. *Brachypodium distachyon*: making hay with a wild grass. *Trends Plant Sci.* **13**: 172-177.

Palmer NA, Donze-Reiner T, Horvath D, Heng-Moss T, Waters B, Tobias C, Sarath G. 2015. Switchgrass (*Panicum virgatum* L) flag leaf transcriptomes reveal molecular signatures of leaf development, senescence, and mineral dynamics. *Funct. Integr. Genomics* **15**: 1-16.

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

Porebski S, Bailey LG, Baum B. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* **15**: 8-15.

Porter CL, Jr. 1966. An analysis of variation between upland and lowland switchgrass, *Panicum Virgatum* L., in central Oklahoma. *Ecology* **47**: 980-992.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. 2012. The Pfam protein families database. *Nucleic Acids Res.* **40**: D290-301.

RC T 2015. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Open access available at: http://cran.r-project.org.

Sarris Panagiotis F, Duxbury Z, Huh Sung U, Ma Y, Segonzac C, Sklenar J, Derbyshire P, Cevik V, Rallapalli G, Saucet Simon B, Wirthmueller L, Menke Frank LH, Sohn Kee H, Jones Jonathan DG. 2015. A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell* **161**: 1089-1100.

Serba DD, Uppalapati SR, Mukherjee S, Krom N, Tang Y, Mysore KS, Saha MC. 2015. Transcriptome profiling of rust resistance in switchgrass using RNA-Seq analysis. *Plant Genome* **8**.

Shang J, Tao Y, Chen X, Zou Y, Lei C, Wang J, Li X, Zhao X, Zhang M, Lu Z, Xu J, Cheng Z, Wan J, Zhu L. 2009. Identification of a new rice blast resistance gene, *Pid3*, by genomewide comparison of paired nucleotide-binding site--leucine-rich repeat genes and their pseudogene alleles between the two sequenced rice genomes. *Genetics* **182**: 1303-1311.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**.

Singh S, Chand S, Singh NK, Sharma TR. 2015. Genome-wide distribution, organisation and functional characterization of disease resistance and defence response genes across rice species. *PLoS ONE* **10**: e0125964.

Stroup JA, Sanderson MA, Muir JP, McFarland MJ, Reed RL. 2003. Comparison of growth and performance in upland and lowland switchgrass types to water and nitrogen stress. *Bioresour. Technol.* **86**: 65-72.

Sun L, Ren H, Liu R, Li B, Wu T, Sun F, Liu H, Wang X, Dong H. 2010. An h-type thioredoxin functions in tobacco defense responses to two species of viruses and an abiotic oxidative stress. *Mol. Plant-Microbe Interact.* **23**: 1470-1485.

Sykes V, Allen F, Mielenz J, Stewart CN, Jr., Windham M, Hamilton C, Rodriguez M, Jr., Yee K. 2015. Reduction of ethanol yield from switchgrass infected with rust caused by *Puccinia emaculata*. *BioEnergy Res.*: 1-9.

Tada Y, Spoel SH, Pajerowska-Mukhtar K, Mou Z, Song J, Wang C, Zuo J, Dong X. 2008. Plant immunity requires conformational charges of NPR1 via S-nitrosylation and thioredoxins. *Science* **321**: 952-956.

Takabatake R, Karita E, Seo S, Mitsuhara I, Kuchitsu K, Ohashi Y. 2007. Pathogen-induced calmodulin isoforms in basal resistance against bacterial and fungal pathogens in tobacco. *Plant Cell Physiol.* **48**: 414-423.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**: 2725-2729.

Tan S, Wu S. 2012. Genome wide analysis of nucleotide-binding site disease resistance genes in *Brachypodium distachyon*. *Comp. Funct. Genomics* **2012**: 12.

Uppalapati S, Serba D, Ishiga Y, Szabo L, Mittal S, Bhandari H, Bouton J, Mysore K, Saha M. 2013. Characterization of the rust fungus, *Puccinia emaculata*, and evaluation of genetic variability for rust resistance in switchgrass populations. *BioEnergy Res.* **6**: 458-468.

Wang H, Wijeratne A, Wijeratne S, Lee S, Taylor C, St Martin S, McHale L, Dorrance A. 2012. Dissection of two soybean QTL conferring partial resistance to *Phytophthora sojae* through sequence and gene expression analysis. *BMC Genomics* **13**: 428.

Wheeler TJ, Clements J, Finn RD. 2014. Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinformatics* **15**: 7.

Xiang Y, Song M, Wei Z, Tong J, Zhang L, Xiao L, Ma Z, Wang Y. 2011. A jacalin-related lectin-like gene in wheat is a component of the plant defence system. *J. Exp. Bot.* **62**: 5471-5483.

Xu C, Min J. 2011. Structure and function of WD40 domain proteins. *Protein Cell* **2**: 202-214.

Yu YG, Buss GR, Maroof MA. 1996. Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. *Proc. Natl. Acad. Sci. USA* **93**: 11751-11756.

Zale J, Freshour L, Agarwal S, Sorochan J, Ownley BH, Gwinn KD, Castlebury LA. 2008. First report of rust on switchgrass (*Panicum virgatum*) caused by *Puccinia emaculata* in Tennessee. *Plant Dis.* **92**: 1710-1710.

Zhang B, Horvath S. 2005. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**: Article17.

Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D. 2004. Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Mol. Genet. Genomics* **271**: 402-415.

Zhu Q, Bennetzen JL, Smith SM. 2013. Isolation and diversity analysis of resistance gene homologues from switchgrass. *G3 Genes Genom. Genet.* **3**: 1031-1042.

Zipfel C, Robatzek S, Navarro L, Oakeley EJ, Jones JDG, Felix G, Boller T. 2004. Bacterial disease resistance in *Arabidopsis* through flagellin perception. *Nature* **428**: 764-767.

Table 3-1. RNA-seq analysis of 'Alamo' and 'Dacotah' used for variant detection and gene expression analysis.

One sample of 'Alamo' and one sample of 'Dacotah' were sequenced at a deep coverage (36X and 20X, respectively) and were used for variant detection. Three biological replicates of 'Alamo' and 'Dacotah' were used for gene expression analysis and were sequenced between 3.7X and 5.2X coverage.

Sample	Sequencing	Initial Reads	Trimmed Reads	Mapped Reads	% Reads Mapped
Alamo	101bp-PE	221,647,550	216,244,584	180,996,166	84.35
Dacotah	101bp-PE	125,257,298	121,958,884	99,283,683	82.03
Alamo- 1	50bp-PE	56,036,894	54,211,026	52,494,171	96
Alamo- 2	50bp-PE	45,931,626	44,044,084	42,657,586	96
Alamo- 3	50bp-PE	58,312,466	55,472,698	53,590,453	96
Dacotah- 1	50bp-PE	59,434,186	56,774,376	54,489,840	96
Dacotah- 2	50bp-PE	48,408,374	46,266,102	44,525,787	96
Dacotah- 3	50bp-PE	63,304,978	61,060,688	58,815,617	96

Table 3-2. Allele-specific PCR primers designed for 'Alamo' for validation of eight SNPs identified by RNA-sequencing analysis.

Nucleotides colored in red indicate the 'Alamo'-specific SNP at that particular location in the switchgrass genome (v 1.1). The annealing temperature is the temperature used for each primer pair during the annealing step of PCR.

Switchgrass Gene ID	Location	Base position	'Alamo' specific forward primer	Base position	'Alamo' specific reverse primer	Annealing temp °C
Pavir.Ba02315	Chr02a	33,777,096	ACCCTCCAGTGGAGCGCATG**T**	33,777,812	GAGCTTGGAGAGGTCGCACA**G**	58
Pavir.Ba03659	Chr02a	73,959,094	CATTAGAGCCCATAAATGA**A**	73,959,715	GAGAACCAGCACATCTCCT**T**	52
Pavir.Hb00487	Chr08b	10,403,612	TACTATGCAAATGAGCTCC**A**	10,404,080	ACCAATTGACATTCCAAC**A**	52
Pavir.Hb01688	Chr08b	48,215,706	GCTGTCCCTTGTTGTTGGACG**C**	48,216,432	CGAAAGGTGAAGATCTGGTTC**G**	58

Table 3-3. Allele-specific PCR primers designed for 'Dacotah' for validation of SNPs identified by RNA-sequencing analysis.

Nucleotides colored in red indicate the 'Dacotah'-specific SNP at that particular location in the switchgrass genome (v 1.1). The annealing temperature is the temperature used for each primer pair during the annealing step of PCR.

Switchgrass gene ID	Location	Base position	'Dacotah' specific forward primer	Base position	'Dacotah' specific reverse primer	Annealing temp °C
Pavir.Ba02315	Chr02a	33,777,096	ACCCTCCAGTGGAGCGCATG**G**	33,777,812	GAGCTTGGAGAGGTCGCACA**A**	58
Pavir.Ba03659	Chr02a	73,959,094	CATTAGAGCCCATAAATGA**C**	73,959,715	GAGAACCAGCACATCTCCT**G**	52
Pavir.Hb00487	Chr08b	10,403,612	TACTATGCAAATGAGCTCC**G**	10,404,080	ACCAATTGACATTCCAAC**G**	52
Pavir.Hb01688	Chr08b	48,215,706	GCTGTCCCTTGTTGTTGGACG**A**	48,216,432	CGAAAGGTGAAGATCTGGTTC**C**	58

Table 3-4. Conserved primers used to validate SNPs identified between 'Alamo' and 'Dacotah' by RNA-sequencing analysis.

All SNPs were validated using traditional DNA sequencing except for the SNP predicted for Pavir.Ib01513.

Gene	Location	SNP location	Alamo SNP	Dacotah SNP	Forward primer	Reverse primer
Pavir.J03515	contig04623	4,711	T	G	ATCCTGAAACGCTTGGAGAG	CTGGATAGCAGCTGGACTTG
Pavir.Ha00691	Chr08a	17,432,708	A	G	GACGCAATGGATAGCAGAAG	TGCTGGAGATTTGTCAGCATC
Pavir.Ha00691	Chr08a	17,432,759	A	C	GACGCAATGGATAGCAGAAG	TGCTGGAGATTTGTCAGCATC
Pavir.J36509	contig60676	1,882	A	C	TGTAGATTGGTCAGCTCGCC	CAATTGCAATGGTGAGGTGC
Pavir.J36509	contig60676	2,098	G	C	TGTAGATTGGTCAGCTCGCC	CAATTGCAATGGTGAGGTGC
Pavir.J36509	contig60676	2,113	G	T	TGTAGATTGGTCAGCTCGCC	CAATTGCAATGGTGAGGTGC
Pavir.Ba02315	Chr02a	33,779,014	G	A	GCGCTTGGTGGTTTAGATATG	TCTTCGGATCTGAAGCACTC
Pavir.J39382	contig83648	2,037	T	G	TACGCCCAGGAACTTGATAG	TGTAGCACAGCTTTCTGAGC
Pavir.Cb00731	Chr03b	14,688,439	C	T	ATCCTTGGCTAGCTGGTAAC	ATAGTAACCATCAGCAGGCG
Pavir.J37164	contig65733	1649	T	C	AACTTCTGCCTCACCATCG	TCAACCTGCAGTCGCTCATC
Pavir.J37164	contig65733	1700	C	G	AACTTCTGCCTCACCATCG	TCAACCTGCAGTCGCTCATC
Pavir.Ib01513	Chr09b	16,589,008	T	C	TCAAACTCAACACCAAGAGG	ATGGCTCCTTGAAGCTGCC

Table 3-5. Unique domains that were identified in 40 switchgrass NB or NB-LRR proteins.

Category	Unique domain	Protein ID	E-value	Structure
Protein modification	Glutaredoxin	Pavir.Bb01048.1.p	9.80E-08	NB-LRR-Glutaredoxin
	Thioredoxin	Pavir.Fa01782.1.p	3.70E-22	NB-LRR-Thioredoxin
		Pavir.Hb00484.1.p	8.50E-21	NB-LRR-Thioredoxin
	Protein tyrosine kinase	Pavir.Ba02898.1.p	1.50E-26	NB-LRR-PK
	Serine/threonine phosphatases, family 2C, catalytic domain	Pavir.J24356.1.p	2.90E-14	NB-LRR-PP2C
	Serine/Threonine protein kinases, catalytic domain	Pavir.Aa03320.1.p	2.30E-13	PK-NB-LRR
		Pavir.Bb00833.1.p	3.20E-30	PK-NB
		Pavir.Ha00561.1.p	4.10E-26	PK-NB-LRR
		Pavir.Ha00691.1.p	2.10E-97	NB-LRR-PK
		Pavir.Ha01101.1.p	2.40E-23	PK-NB-LRR
		Pavir.Ha01108.1.p	1.70E-25	PK-NB-LRR
		Pavir.Ha01246.1.p	4.00E-21	PK-NB-LRR
		Pavir.Hb00345.1.p	7.80E-16	PK-NB-LRR
		Pavir.Hb00951.1.p	6.20E-28	PK-NB
		Pavir.Hb01026.1.p	3.10E-22	PK-NB
		Pavir.Hb01049.1.p	8.40E-13	PK-NB
		Pavir.Hb01188.1.p	1.40E-28	PK-NB
		Pavir.J00375.1.p	2.40E-25	PK-NB-LRR
		Pavir.J01180.1.p	4.60E-27	PK-NB-LRR
		Pavir.J16540.1.p	1.60E-23	PK-NB-LRR
		Pavir.J23132.1.p	7.30E-19	PK-NB
		Pavir.J36755.1.p	3.30E-12	PK-NB
		Pavir.J37596.1.p	1.30E-14	PK-NB
		Pavir.J39176.1.p	4.00E-14	PK-NB
DNA binding, transcription	B3 DNA binding domain	Pavir.Ba02315.1.p	2.20E-10	B3-NB-LRR
	B3 DNA binding domain WRKY DNA binding domain	Pavir.J20878.1.p	3.10E-07 1.00E-14	B3-NB-LRR-WRKY
	BED zinc finger	Pavir.Fa02339.1.p	5.40E-09	ZF-NB-LRR
		Pavir.Ga00028.1.p	5.40E-09	ZF-NB-LRR
		Pavir.Gb00931.1.p	1.70E-08	ZF-NB-LRR
		Pavir.J19380.1.p	2.50E-12	ZF-NB-LRR
		Pavir.J40131.1.p	7.80E-08	BED-NB
	WRKY DNA-binding domain	Pavir.J33941.1.p	1.80E-14	NB-LRR-WRKY
protein trafficking and vesicle movement	MSP (Major sperm protein) domain	Pavir.Ib02384.1.p	1.10E-09	MSP-NB-LRR
		Pavir.Ib02433.1.p	1.50E-08	MSP-NB-LRR
		Pavir.J18369.1.p	5.10E-13	MSP-NB
Protein-protein interaction	WD domain, G-beta repeat	Pavir.J03445.1.p	5.00E-06	NB-LRR-WD
	hAT family C-terminal dimerisation region	Pavir.Fb01504.1.p	3.50E-08	NB-LRR-hAT
Sugar binding	Jacalin-like lectin domain	Pavir.Hb01174.1.p	5.60E-11	NB-LRR-Jacalin
Signal transduction	Calmodulin binding protein-like	Pavir.Hb00190.1.p	5.10E-79	NB-LRR-Calmodulin
Transposon element	gag-polypeptide of LTR copia-type	Pavir.Aa01444.1.p	8.90E-07	NB-LRR-LTR

Table 3-6. Significant differentially expressed RGHs between 'Alamo' and 'Dacotah'.

A gene was considered significantly differentially expressed if it had a fold change greater than two and an FDR corrected p-value < 0.05.

Gene ID	Chromosome	Gene start location	Gene end location	Alamo RPKM[1] means	Dacotah RPKM[1] means	Fold change[2]	FDR corrected p-value[3]
Pavir.Ha01577	Chr08a	45819137	45820543	0.074	11.565	54.837	6.03E-08
Pavir.Hb00191	Chr08b	3683280	3690800	22.965	4.363	-5.724	1.02E-07
Pavir.J41099	contig99714	31	3755	10.721	0.387	-23.222	6.52E-07
Pavir.Ha00254	Chr08a	5275668	5283177	8.387	0.088	-43.065	2.74E-06
Pavir.Ea02999	Chr05a	52240105	52245554	7.763	0.145	-31.66	9.51E-06
Pavir.J18698	contig20439	2509	7946	0.019	10.273	68.028	1.57E-05
Pavir.J36183	contig58243	960	5090	7.474	0.014	-57.923	1.77E-05
Pavir.Ha00258	Chr08a	5331659	5334691	12.753	0.617	-18.44	3.17E-05
Pavir.Ba03626	Chr02a	73291071	73293182	0.139	8.63	30.634	5.28E-05
Pavir.J38310	contig74920	865	4419	0.377	8.784	16.208	5.28E-05
Pavir.J13560	contig156657	131	2675	0.565	8.893	11.899	1.71E-04
Pavir.J35611	contig54273	2061	5449	0.195	5.96	17.525	4.86E-03
Pavir.J01215	contig01325	2932	5562	7.052	1.095	-6.492	5.22E-03
Pavir.J03054	contig03762	12948	18535	4.087	0.096	-20.509	7.55E-03
Pavir.Aa01110	Chr01a	14217597	14230708	3.764	0.03	-26.75	0.013
Pavir.Ha01146	Chr08a	31223571	31227884	0.703	5.777	6.478	0.019
Pavir.Cb00441	Chr03b	7318929	7322714	11.648	26.964	2.068	0.034
Pavir.Bb00035	Chr02b	557471	559270	0.054	3.333	18.041	0.041
Pavir.Ha00828	Chr08a	23240515	23246499	3.354	0.02	-25.367	0.041
Pavir.J29227	contig33532	301	4214	3.316	0.035	-22.844	0.041
Pavir.J02637	contig03134	4596	14505	4.903	0.772	-6.169	0.048

[1]RPKM = reads per kilobase per million of reads mapped
[2]Fold Change = expression level difference between the two samples when using 'Dacotah' as the reference sample
[3]FDR corrected p-value = p-value that the Fold Change is significant based on the corrected False Discovery Rate

Figure 3-1. Numerical and structural representation of the 1,011 switchgrass RGHs identified in this study.

Four different protein structures were found for the 1,011 switchgrass RGHs identified in this study: 1) CC-NB-LRR, 2) CC-NB, 3) NB, and 4) NB-LRR. The coiled-coil domain is depicted in purple, the NB domain in green, and the LRR region in yellow. The placement of the domains below does not reflect accurate molecular distances.

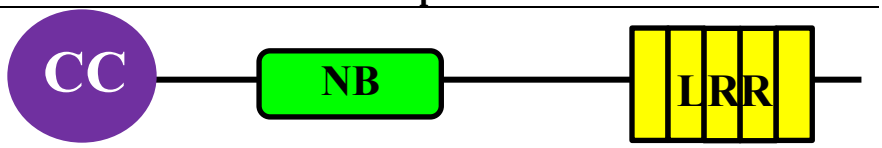
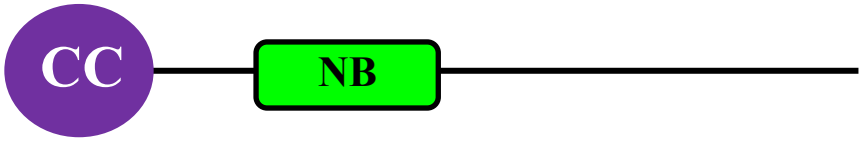
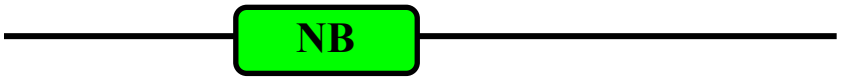
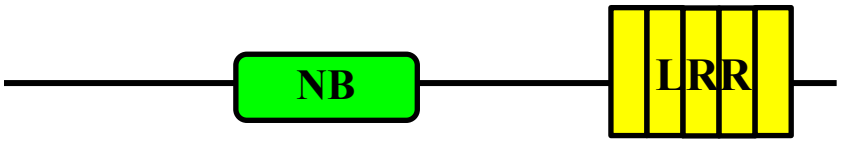
Number of genes	Structure	Structure representation
255	CC-NB-LRR	
150	CC-NB	
179	NB	
427	NB-LRR	

Figure 3-2. Chromosomal distribution of the 1,011 switchgrass RGHs identified in this study.

The switchgrass A and B subgenomes are marked as black and white, respectively. An additional 500 genes were found on unanchored contigs (data not shown).

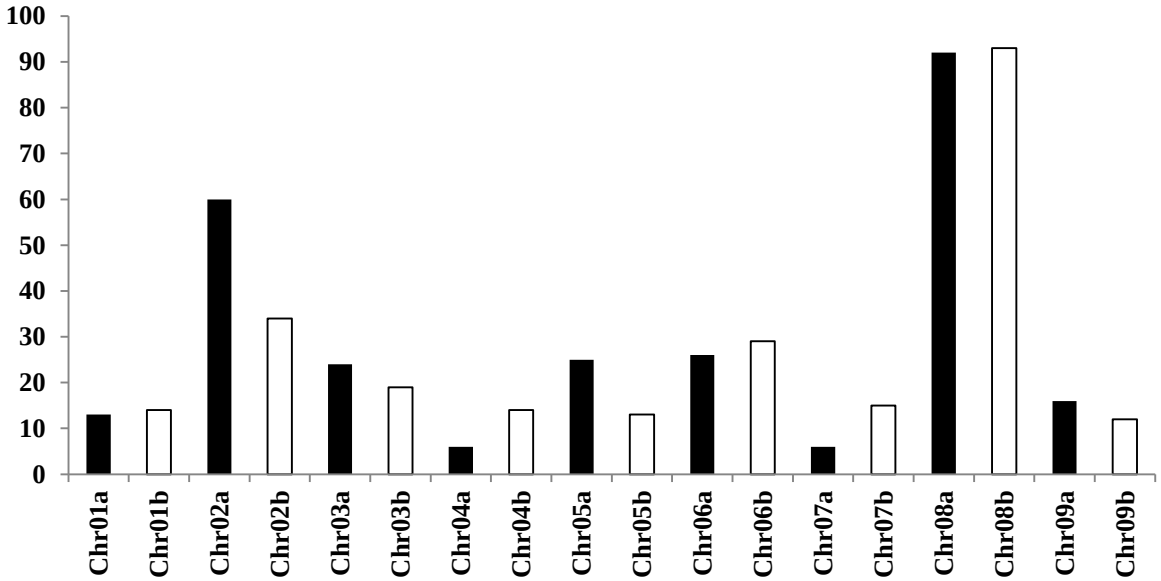


Figure 3-3. Skylign output of the NB region for the 578 switchgrass sequences that contained a full-length NB domain.

The circled regions include the Walker-A/P-loop motif, Kinase 2, Kinase 3/RNBS-B, and GLPL.

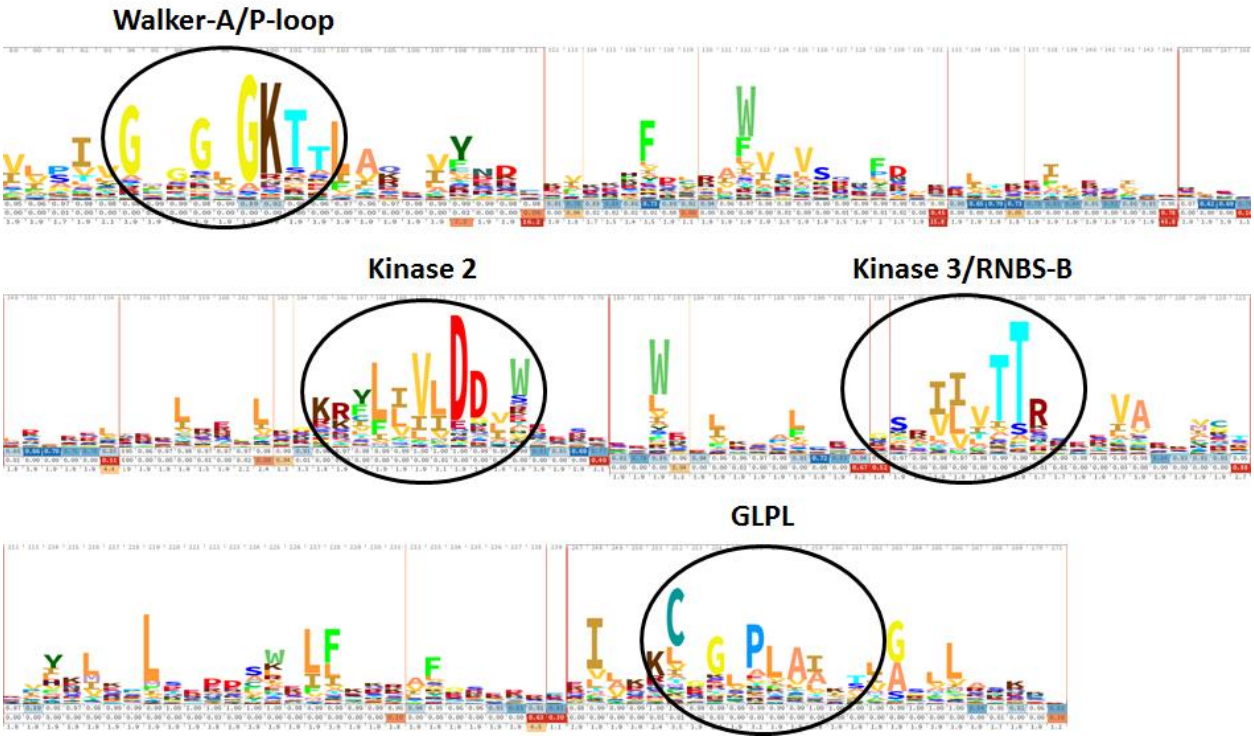


Figure 3-4. Phylogenetic analysis of 578 switchgrass RGHs and 116 *Brachypodium distachyon* RGHs.

The phylogenetic tree was built using MEGA 6 with default settings. Branches for the *Brachypodium* RGHs are highlighted in blue and branches for switchgrass RGHs located on chromosome 8 are highlighted in red. The green dots represent the phylogenetic location for the 10 protein kinase-containing NB-LRR genes that are found on chromosome 8. Numbers on the tree indicate branches with bootstrap values greater than 50.

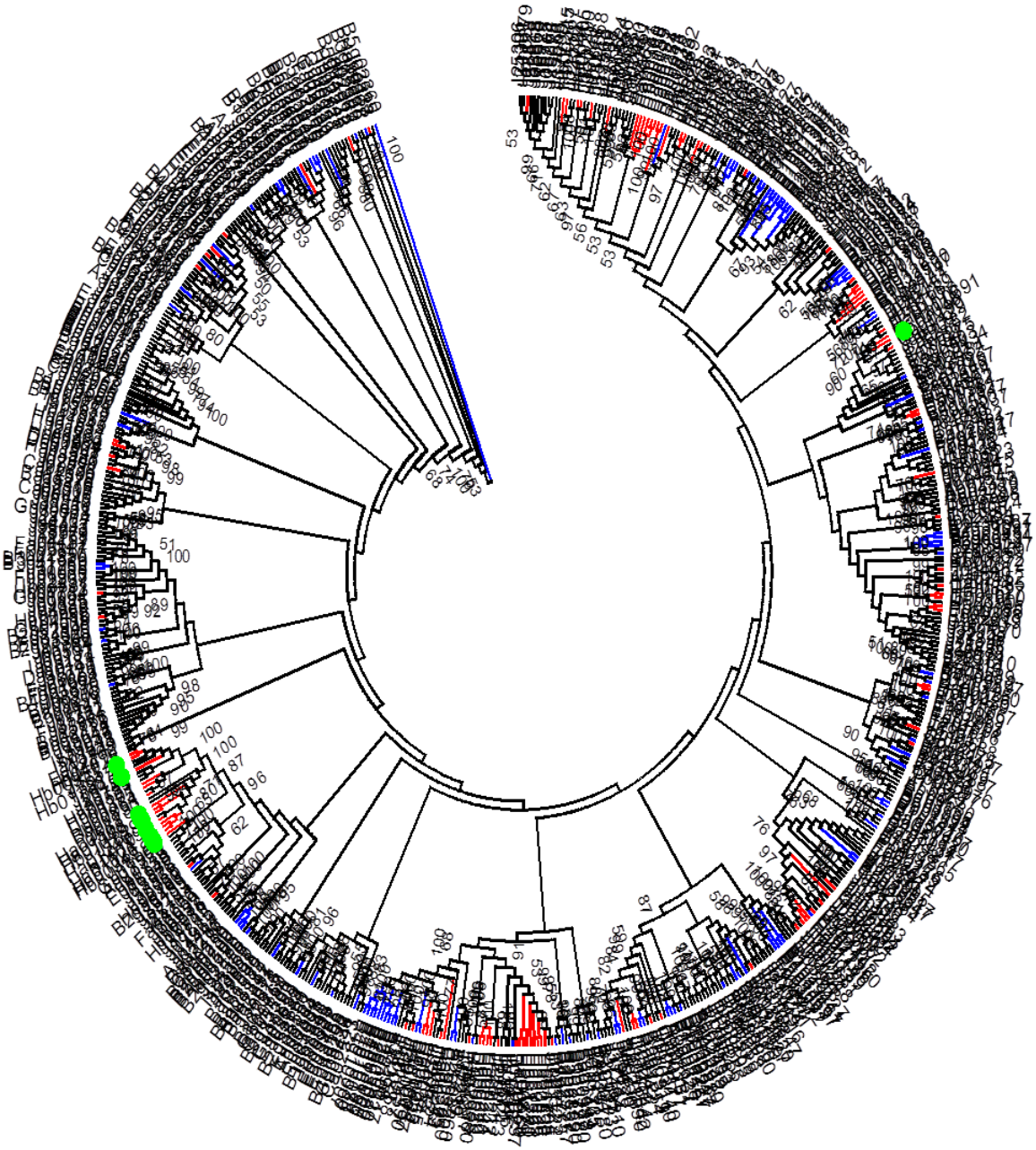


Figure 3-5. Validation of SNPs identified by RNA-sequencing using allele-specific primers.

Pictured are the PCR products using 'Dacotah'-specific primers where a = 'Alamo' DNA and d = 'Dacotah' DNA. 1) 1 kb ladder, 2) Pavir.Ba03659, 3) Pavir.Ba02315, 4) Pavir.Hb01688, 5) Pavir.Hb00487, 6) 1 kb ladder, 7) ELF1α

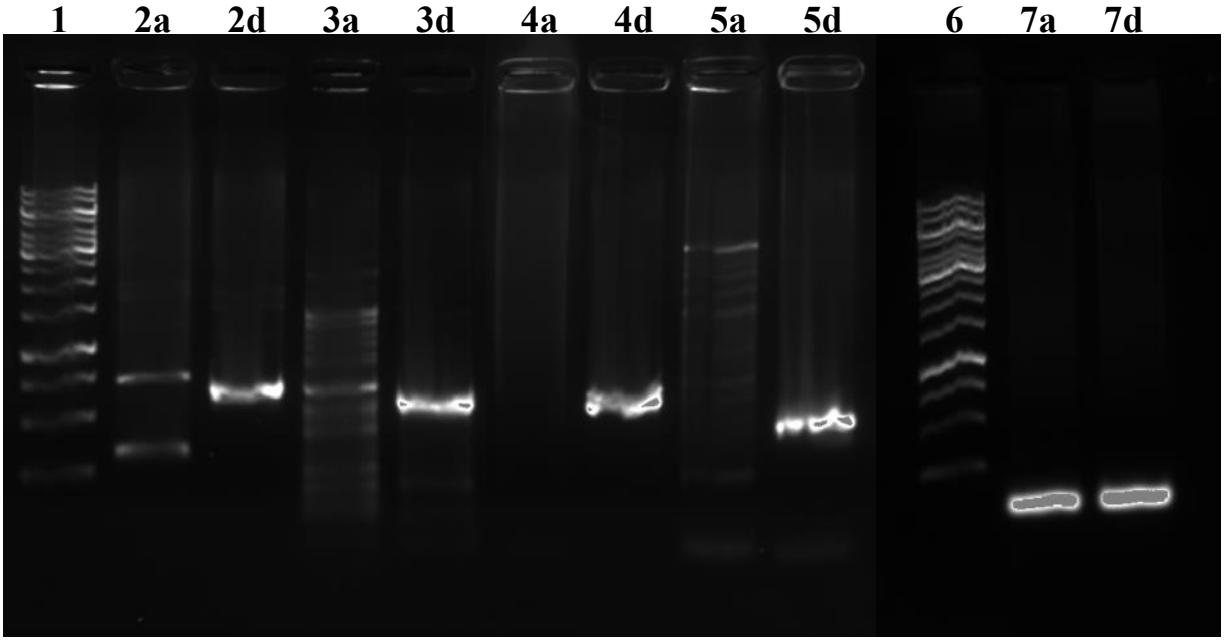
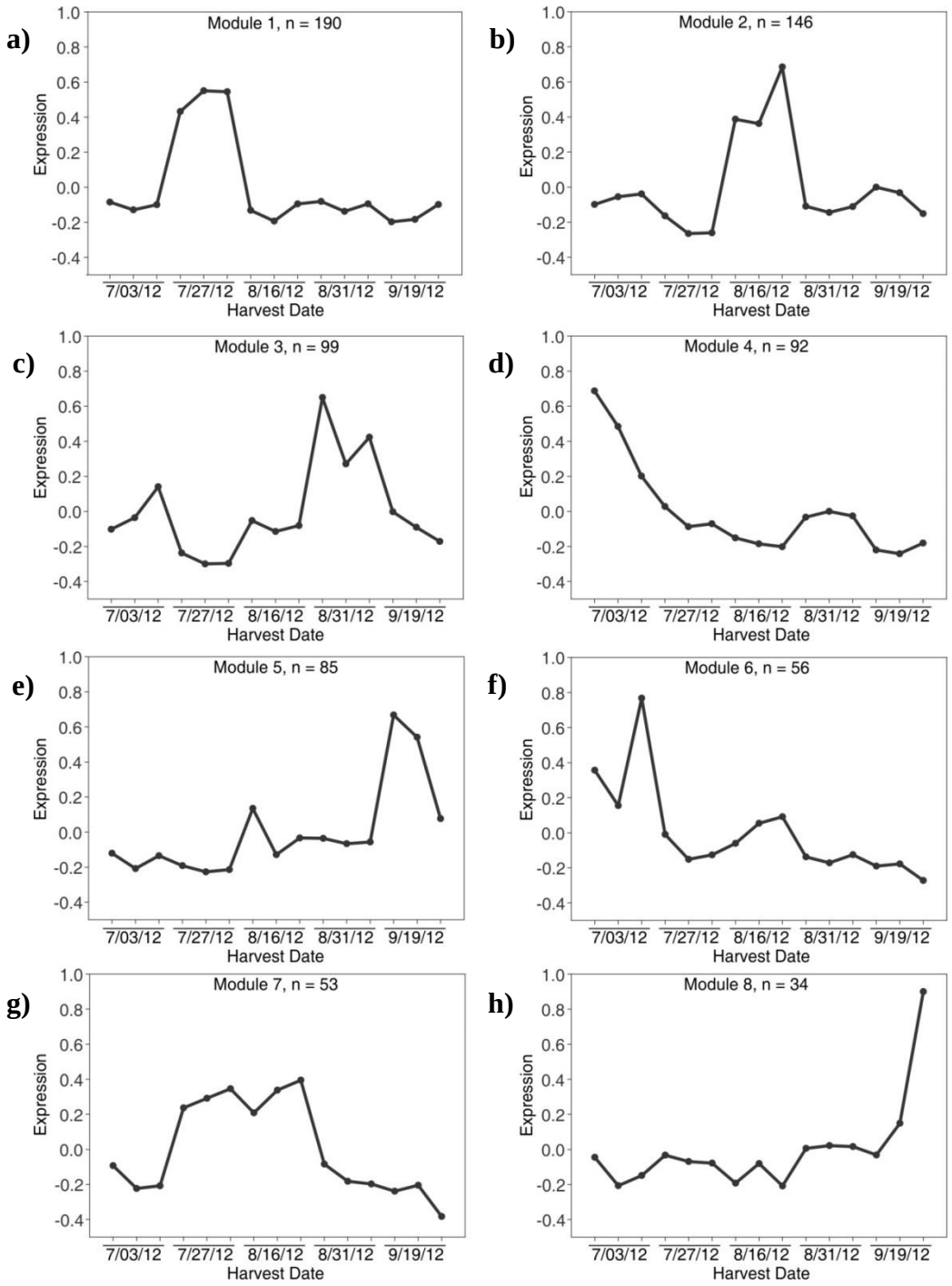


Figure 3-6. WGCNA analysis of developmental gene expression for 755 RGHs in the switchgrass cv. Summer.

Shown are eight modules that correspond to the expression patterns of 755 RGHs identified in this study over five sampling time points. The line represents the overall expression pattern of each co-expression module. The number of genes included in each module is represented by n.



CHAPTER 4: SNP-based linkage map construction from RNA-sequencing and identification of a QTL associated with disease resistance in the bioenergy crop switchgrass

ABSTRACT

Large-scale planting and maintenance of elite switchgrass cultivars will be necessary to fully utilize switchgrass as a bioenergy crop. Breeding of switchgrass cultivars for improved biomass production and/or feedstock quality will likely reduce the genetic diversity that is prevalent in this species. This could lead to devastating disease epidemics caused by switchgrass rust, the most wide spread and predominant foliar fungal pathogen of switchgrass. Therefore, an immediate understanding of the genetic mechanisms underlying tolerance to switchgrass rust is needed to make informed breeding decisions. In this study, we created novel pseudo-F_2 mapping populations of switchgrass that were derived from an initial cross between 'Alamo', a rust-resistant cultivar, and 'Dacotah', a rust-susceptible cultivar. RNA-sequencing was used to identify 1,591,979 SNP molecular markers among individuals of these populations, of which 966 high-quality SNPs were used to create a genetic linkage map. The map created in this study was 3916 cM in length with marker spacing averaging 4.13 cM. In addition, we were able to assign 170 of the previously unanchored contigs in the switchgrass genome (v 1.1) to one of the 18 genetic linkage groups. QTL analysis identified one genetic region spanning from 74 cM to 138 cM on chromosome 2B that may be contributing to rust resistance. Additionally, a casein kinase II protein was also found in this region that may play a role in disease response. This is the first report of using RNA-sequencing from a pseudo-F_2 population to create a genetic linkage map for switchgrass. The results of this study provide useful insight into the molecular mechanisms that

may be contributing to rust disease resistance and provide a platform for future genetic studies in this important bioenergy crop.

INTRODUCTION

Switchgrass is a warm-season prairie grass that can be found growing across the vast plains of North America. In the early 1990s, switchgrass was designated by the United States Department of Energy as an herbaceous crop species worthy of development into a bioenergy crop (Wright & Turhollow, 2010). Switchgrass has many favorable attributes that make it an attractive plant for sustainable biomass production. As a C4 perennial grass, switchgrass is efficient at converting light energy and carbon dioxide into biomass (Madakadze *et al.*, 1998; Field *et al.*, 2008). In addition, some switchgrass cultivars can grow upwards of 3 meters in height (Porter, 1966). The extensive belowground root system of switchgrass is effective at sequestering water and preventing soil erosion, making it an excellent conservation crop (Parrish & Fike, 2005). Once established, switchgrass requires minimal nutrient input and can thrive when grown on marginal lands (Sanderson *et al.*, 2006).

Due to its growing popularity as a model bioenergy crop, genomic resources for switchgrass are rapidly becoming available. In early 2012, the first preliminary version of the switchgrass genome (v 0.0), which was based on the tetraploid AP13 variety of the commercially-grown cv. Alamo, was released (Goodstein *et al.*, 2012). Since then, a second version of the switchgrass genome (v 1.1) has become available that has anchored approximately 52% of the 1,230 Mbp sequencing data onto chromosomes; however, 48% of the sequenced genome remains on unanchored contigs (Goodstein *et al.*, 2012). Other notable genomic

resources include thousands of EST sequences (Tobias *et al.*, 2008; Wang *et al.*, 2012b), a BAC library (Saski *et al.*, 2011), Affymetrix cDNA microarray chip (Zhang *et al.*, 2013), SSR and SRAP molecular markers (Huang *et al.*, 2011; Wang *et al.*, 2011), and numerous RNA-sequencing datasets (Wang *et al.*, 2013; Childs *et al.*, 2014a; Palmer *et al.*, 2015).

Switchgrass is a readily out-crossing species (Martinez-Reyna & Vogel, 2002). There are two major ecotypes of switchgrass, lowland and upland, that vary considerably in their physical characteristics and growth habitats (Porter, 1966). Hybrid crosses between the two ecotypes are possible if the plants are the same ploidy (Martinez-Reyna & Vogel, 2002). To create new cultivars with improved agronomic traits, it will be necessary to successfully apply molecular marker-assisted breeding strategies to switchgrass. With recent advances in next generation sequencing (NGS), several studies have aimed to identify molecular markers and construct genetic linkage maps that can be utilized in switchgrass breeding programs (Liu *et al.*, 2012; Serba *et al.*, 2013; Fiedler *et al.*, 2015). Single nucleotide polymorphisms (SNPs) are useful molecular markers as they are wide spread across the genome, abundant, polymorphic, and easily detectable by NGS technologies (Kumar *et al.*, 2012). A study by Fielder *et al* (2015) used a genotype-by-sequencing approach to identify 4,611 high-quality SNPs in a ‘Kanlow’ x ‘Alamo’ full-sib F_1 population. These SNPs were dispersed across the 18 linkage groups and map lengths for the female and male plants were estimated to be 1770 cM and 2059 cM, respectively (Fiedler *et al.*, 2015). Presently, no studies have identified SNPs and created genetic linkage maps based on crosses between lowland and upland switchgrass ecotypes.

Elite switchgrass cultivars will likely be grown in monoculture in order to fully utilize switchgrass as a bioenergy crop; however, this practice may increase the selection pressure on foliar pathogens, such as switchgrass rust. Highly virulent switchgrass rust pathogens have the

potential to cause devastating disease epidemics and significant reductions in biomass yield. The causal agent of switchgrass rust is *Puccinia emaculata* (Gustafson *et al.,* 2003). Most switchgrass cultivars have been shown to be moderate to highly susceptible to this pathogen, with lowland ecotypes generally being more resistant to rust than their upland counterparts (Gustafson *et al.*, 2003; Uppalapati *et al.*, 2013). Switchgrass rust occurrence is wide spread and infected fields have been reported from Virginia (Frazier *et al.,* 2012) to Oklahoma (Uppalapati *et al.*, 2013). Recently, Skyes *et al* (2015) evaluated the effect that rust infection had on bioethanol production of field-grown 'Alamo' and 'Kanlow' switchgrass plants. Switchgrass plants displaying high disease severity to switchgrass rust produced 21-55% less ethanol than uninfected plants (Sykes *et al.,* 2015). Therefore, it will be necessary to understand the genetic mechanisms controlling disease resistance in switchgrass in order to employ disease-resistant cultivars in switchgrass breeding programs.

For a genetic linkage map to be complete for switchgrass, it must incorporate genetic information from both lowland and upland cultivars. Previously, only two reports have generated linkage maps from lowland x upland crosses and of these reports, none has identified QTLs that may contribute to disease resistance (Missaoui *et al.*, 2005; Serba *et al.*, 2013). In this study, we performed RNA-sequencing on 154 F_2 population individuals that were derived from an initial cross between 'Alamo' and 'Dacotah', and subsequent second crosses between six of the 'Alamo' x 'Dacotah' F_1 progeny. 'Alamo' is a lowland tetraploid switchgrass cultivar that is moderately resistant to switchgrass rust whereas 'Dacotah' is an upland tetraploid switchgrass cultivar that is highly susceptible to switchgrass rust. Reference-guided mapping and SNP calling identified 966 SNPs that were used to assign 170 not previously anchored contigs to one of the 18 linkage groups. QTL analysis for disease response of the 154 F_2 individuals identified

genomic areas that may contribute to disease resistance in this important bioenergy crop. The results of this study help aid our understanding of the genetic mechanisms that may be underlying switchgrass resistance to switchgrass rust.

MATERIALS AND METHODS

Identification of switchgrass F_1 progeny segregating for disease resistance

'Alamo', a rust-resistant lowland cultivar, and 'Dacotah', a rust-susceptible upland cultivar, were crossed to generate a F_1 population segregating for disease resistance. A total of 165 F_1 individuals was planted in the field at the Virginia Tech Kentland Farm Research Center in Blacksburg, VA in 2009. Over the course of two years, the F_1 plants were assessed for variation in key agronomic traits, as well as surveyed for their response to naturally occurring rust infection. Six F_1 plants that varied in their resistance phenotype to switchgrass rust were selected to create pseudo-F_2 populations. The pseudo-F_2 populations created in this study are expected to retain the majority of the genetic information present in 'Alamo' and 'Dacotah'.

Crossing of F_1 plants to generate segregating F_2 populations

All of the switchgrass plants for this study were maintained in the Washington Street greenhouse facility at Virginia Tech. In 2012, three sets of paired F_1 individuals (DA90 x DA120, DA23 x DA139, and DA5 x DA49) were crossed by bagging the plants and keeping the pairs in separate greenhouses. Hybrid seeds were harvested from the female plants (90, 23, and 5) and approximately 100 seeds were obtained for each cross. The hybrid F_2 seeds were planted

in 3.5 x 3.5 inch pots containing MiracleGro Moisture Control soil and maintained in the greenhouse. Overall, the pseudo-F_2 population was comprised of three subpopulations, which consisted of 154 total individual plants. For this study, approximately 50 plants were randomly chosen from each F_2 population for further analysis. Three clonally propagated replicates of each individual plant were planted in the field at the Kentland Farm Research Center in Blacksburg, VA. After two years of growth, phenotypic data were collected for the members of the pseudo-F_2 population.

Switchgrass tissue collection, RNA isolation, and RNA-sequencing

Leaf tissue for 'Alamo' and 'Dacotah' was collected as described in chapter 2. For the F_1 and pseudo-F_2 individuals, the first fully expanded leaf of E2 to E3 stage switchgrass tillers was chosen for analysis. First, the midrib of the leaf blade was removed and discarded. Next, 200 mg of fresh leaf tissue was collected in 2 mL epi-tubes containing a metal BB. The tubes were immediately frozen in liquid nitrogen. Finally, the tissue was ground using a GenoGrinder 2000 (SPEX Sample Prep, Metuchen, NJ, USA) into a fine powder and stored at -80°C until further analysis.

Total RNA was isolated using a modified TRIzol combined with columns method. The quality and quantity of the total RNA were assessed using a NanoDrop D1000 (Nanodrop, Wilmington, DE, USA). RNA samples were also sent for BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) analysis to measure the RNA Integrity Number (RIN). Total RNA samples with a RIN quality score greater than 7 were used for RNA-sequencing (RNA-seq). The

samples were sent to Weill Cornell Medical College (New York, NY, USA) for cDNA library preparation and 101 bp paired-end RNA-seq.

RNA-seq analysis and mapping to the switchgrass reference genome (v 1.1)

FASTQ-mcf (v 1.04.807) (Aronesty, 2011) was used to trim and clean the raw RNA-seq reads. The Illumina adapter sequences were removed and reads that contained an overall base quality score (Q) greater than 30, as well as a minimum length of 50, were kept for downstream analyses. The switchgrass genome (v 1.1) and its annotation resources were downloaded from Phytozome (v 10.3, http://phytozome.jgi.doe.gov/pz/portal.html). Quality trimmed and filtered reads were mapped to the switchgrass reference genome using Bowtie 2 (v 2.1.0) (Langmead & Salzberg, 2012) with default parameters.

SNP calling, filtering, and R/qtl analysis

Bedtools coverage (v 2.21.0-26-g31878f3) function was run to determine the number of reads that spanned each position in the genome (Quinlan & Hall, 2010). Single nucleotide polymorphisms (SNPs) were called using Freebayes (v 0.9.18-3-gb72a21b) (Garrison & Marth, 2012) and the following restrictions: -u, -x, -i, -m 20, -q 30, --min-coverage 8. The SNP information for all samples was combined into a single file selecting for biallelic SNPs that were covered by a minimum of eight reads. Next, the SNPs were filtered to analyze their locations in the F_2 populations. These results were collected in a hapmap file and converted into map making format using a homemade perl script. The map maker file was then imported into R and analyzed using the R/qtl package (Broman *et al.*, 2003; RC, 2015). The genotype probability for the

disease phenotype data was calculated using a step width of 2.0 and an error probability of 0.0001. Next, the map was scanned using a "normal" model, "em" method, and 1,000 permutations to identify QTLs that may correlate with disease resistance.

Disease assays

Puccinia emaculata urediniospores were collected from infected switchgrass leaves at the Kentland Farm Research Center in Blacksburg, VA. The collection of spores was suspended in inoculation buffer (0.05% Hi-Yield spread sticker, 1 mM 1-Nonanol 97%, and 10 mg/L Benlate) and sprayed onto a susceptible 'Dacotah' plant. After inoculation, the 'Dacotah' plant was incubated in a dew chamber in the dark with 100% humidity for 16 h. Approximately 14 d after inoculation, a single pustule of *P. emaculata* was selected and subsequently amplified on susceptible 'Dacotah' plants using the method described above. Following sufficient amplification, this isolate, termed VT2-1, was maintained on susceptible switchgrass plants in the greenhouse.

Members of the three switchgrass mapping populations were clonally split into three biological replicates. Next, each biological replicate was individually potted in a square pot containing MiracleGro Moisture Control soil. The plants were maintained in the greenhouse with a 16 h day/8 h night photoperiod and supplemental lighting as needed. All disease assays were performed in the greenhouse. Freshly collected *P. emaculata* urediniospores were mixed 1:10 with talcum powder and inoculated by hand onto the surface of the first fully expanded leaf of E2 to E3 stage switchgrass tillers. The plants were placed in a chamber with a humidifier and kept under 100% humidity for 16 hours. Ten days post inoculation the leaves were scored for

disease severity according to the scaling previously reported by Gustafson *et al* (2003). An alternative method for disease scoring is outlined in the appendix to Chapter 4.

RESULTS

Creation of 'Alamo' x 'Dacotah' F_2 mapping populations

To better understand the genetic mechanisms underlying disease resistance in switchgrass to switchgrass rust, we crossed the tetraploid lowland cultivar 'Alamo', which is moderately resistant to rust, to the tetraploid upland cultivar 'Dacotah', which is highly susceptible to rust (Figure 4-1). A total of 165 F_1 hybrid plants was regenerated from this cross and subsequently planted in the field for evaluation. There was phenotypic variation among the F_1 progeny, which was shown to be segregating for several important agronomic traits including flag leaf width, flag leaf length, plant height, number of tillers, and overall biomass (data not shown). In addition, the F_1 progeny were found to be segregating for disease resistance to naturally-occurring switchgrass rust infection (Figure 4-2).

Six F_1 progeny were selected for the creation of the 3 pseudo-F_2 populations based on their disease response (Table 4-1). The following three crosses were made to create three subpopulations: DA23, a resistant line was crossed with DA139, a susceptible line; DA90, a susceptible line, was crossed with DA120, a resistant line; and DA5, a susceptible line was crossed with DA49, another susceptible line. Approximately 50 individuals from each population were selected for evaluation with a final total of 154 pseudo-F_2 individuals selected for analysis.

Field data for the pseudo-F_2 populations found that they were also segregating for plant height, number of tillers, and dry biomass (Table 4-2), as well as for disease resistance (data not shown).

Disease assays of F_2 populations under greenhouse conditions

While field disease surveys can help identify rust-susceptible and rust-resistant switchgrass cultivars, there are several parameters contributing to disease severity that cannot be controlled under natural conditions. For instance, *Puccinia emaculata* inoculum is not homogenous and indeed, several haplotypes of *P. emaculata* have been identified from rust samples collected from the same field (Orquera Delgado, 2014). In addition, rust disease severity is affected by the concentration of inoculum, temperature, humidity, and compatibility between the pathogen and the plant host.

In this study, we aimed to eliminate some of the variation that can affect switchgrass response to switchgrass rust. Disease assays on the three F_2 subpopulations were performed in a greenhouse setting under controlled conditions with a pure isolate strain of *P. emaculata* (termed VT2-1). The plants were scored for disease response based on the scale proposed by Gustafson *et al* (2003) (Table 4-3). Since these scores were based on visual observation, we developed an alternative method to quantify the number of pustules per cm^2 (see Appendix to Chapter 4). In all three populations, there appears to be a normal distribution of the disease phenotype (Figure 4-3). This suggests that a major disease resistance gene or QTL could be controlling rust disease resistance to isolate VT2-1.

RNA-sequencing and reference-guided mapping of 154 F_2 individuals

Total RNA was isolated from the 'Alamo' and 'Dacotah' plants, the six F_1 parents, and 154 F_2 individuals and was sent for RNA-sequencing. In total, 2,738,239,396 paired-end sequencing reads were generated for this experiment. The 'Alamo' and 'Dacotah' plants were sequenced at 36X and 20X coverage, respectively. All F_1 and F_2 individuals were sequenced at 1x coverage. After quality trimming and filtering, 2,619,763,658 high-quality paired-end reads were obtained for all three generations. Bowtie 2 (v 2.1.0) was used to map the RNA-sequencing reads to the switchgrass reference genome (v 1.1) using the default parameters (Langmead & Salzberg, 2012).

SNP calling and filtering for high quality SNPs

After mapping, SNPs were called between each individual and the switchgrass reference sequence using freeBayes. SNPs were considered if they were covered by a minimum of eight reads and had a quality score of at least Q30. A total of 1,591,979 SNPs was found between the 154 F_2 individuals and the switchgrass reference. After removing SNPs that were contained within missing data, 55,474 SNP markers were left that could be identified in all of the plants and of these, 22,458 of these were different from the reference. These SNPs were further filtered to select for biallelic SNPs within the pseudo-F_2 individuals. After filtering, 1,072 high-quality SNP markers were found and used for genetic linkage map construction.

Creation of a draft genetic linkage map for 'Alamo' and 'Dacotah'

The molecular marker file was imported into R and the draft genetic linkage map was created using R/qtl (Broman *et al.*, 2003). After checking for segregation distortion, we found that approximately 87 markers could be considered distorted ($p < 0.05$) and so these were removed from the dataset. To separate the markers into 18 linkage groups, a recombination factor of 0.20 and a minimum LOD score of 8 were used. This yielded 23 linkage groups; however, linkage groups 19-23 contained five markers or less (16 markers total) and were thus excluded from the dataset. The resulting 18 linkage groups contained 969 SNP markers. These 18 linkage groups were reordered and then manually curated to determine appropriate SNP marker placement. After manually checking each individual linkage group, an additional three markers were removed from the dataset resulting in a genetic linkage map with 966 high-quality SNPs.

The final 'Alamo' x 'Dacotah' genetic linkage map consisted of 966 SNPs that were spread across the 18 linkage groups for a recombinational distance of 3916 cM (Figure 4-4). Using our map, we were able to assign 170 previously un-anchored contigs of the switchgrass reference genome (v 1.1) to one of the 18 linkage groups. There are 268 markers assigned to these contigs, with 59 of the contigs containing two markers and the remaining 111 contigs containing just one marker. The total amount of previously un-anchored switchgrass assembly that could be anchored to this map is approximately 2 Mbp of the 593 Mbp of switchgrass sequence that remains on unassembled contigs. In addition, chromosome 9B had the most number of SNP markers at 116 whereas chromosome 7B contained the least number of SNP markers at nine (Table 4-4). Aside from chromosome 7B, only one other chromosome contained less than 20 SNP markers (8B with 14). Overall, the SNP markers were distributed across the 18

linkage groups with average spacing between the markers ranging from 2.72 cM to 9.80 cM for chromosomes 5A and 7B, respectively (Table 4-4). Thus, our results indicate that this map could potentially be used to find regions for QTL analysis.

Identification of a QTL that may correlate with disease resistance

R/qtl was also used to identify regions in the 'Alamo' x 'Dacotah' map that may correlate with disease resistance. One region on linkage group 3, which corresponds to chromosome 2B, was identified at position 92 cM and exhibited the highest LOD score of 3.2 (Figure 4-5). Interestingly, this SNP was found to be on contig01799:07758, which we were able to anchor to this linkage group with our map. This SNP is located in a gene that encodes for a casein kinase II protein. This position was also the only position found to be significant at an alpha level of 37% (Figure 4-5) and no positions were found to be significant at higher alpha levels of 10% and 5%. To pinpoint the location of the potential QTL on chromosome 2B, we applied a confidence interval of 95% to the data. This narrowed down the region of the genome corresponding to phenotypic activity between markers at 74 cM and 138 cM on chromosome 2B (Figure 4-6). Finally, an effect plot was generated on the data to identify which genotype was responsible for the phenotype observed at this particular marker. The effect plot showed that genotype AA, which is the genotype for 'Dacotah', the susceptible F_0 parent, was more responsible for the higher disease score at this particular location (Figure 4-7). Since a higher disease score correlates to an increase in susceptibility, further studies are needed to determine what genes are present at this location in the genome and the mechanism of susceptibility that this locus confers.

DISCUSSION

Switchgrass is a candidate lignocellulosic biofuel feedstock (McLaughlin *et al.*, 1999). For switchgrass to be efficiently cultivated as a bioenergy crop there will likely be an increase in the number of switchgrass plantings per acreage, and desirable switchgrass cultivars will likely be grown in monoculture. Several diseases have been observed on switchgrass with *Puccinia emaculata*, the causal agent of switchgrass rust, being the most reported and the most wide-spread. While a few reports have shown that *P. emaculata* infection reduces biomass yields in switchgrass (Hopkins *et al.*, 1995; Hagan *et al.*, 2015), only recently has a study by Sykes *et al* (2015) demonstrated that naturally-occurring *P. emaculata* infections can significantly reduce the ethanol yields from susceptible switchgrass cultivars by up to 55% (Sykes *et al.*, 2015). Therefore, the genetic mechanisms underlying resistance to switchgrass rust must be elucidated.

In this study, we created a pseudo-F_2 mapping population that was derived from an initial cross between the rust-resistant lowland cv. Alamo and the rust-susceptible upland cv. Dacotah. Despite their different ecotypes, 'Alamo' and 'Dacotah' are able to create hybrid progeny because they are both tetraploids (Martinez-Reyna & Vogel, 2002). Aside from their response to switchgrass rust (Frazier *et al.*, 2012), 'Alamo' and 'Dacotah' vary in other important physical characteristics including biomass production, flowering time, and drought tolerance (Liu *et al.*, 2015). Thus, the populations developed in this study could be used to determine genes and/or QTLs underlying other important agronomic traits.

The creation of the pseudo-F_2 populations and the genetic linkage map in our study is novel for several reasons. First, while genetic linkage maps have been generated from lowland x upland crosses (Missaoui *et al.*, 2005; Serba *et al.*, 2013), no studies have used 'Alamo' and

'Dacotah'. Second, the maps generated in other studies were based on segregation data from F_1 populations. Switchgrass is a readily out-crossing species and is believed to have a self-incompatibility system similar to that found in other grasses (Martinez-Reyna & Vogel, 2002). As a result, switchgrass has a highly heterozygous genetic background and even members of the same cultivar can vary significantly in their genetic composition (Narasimhamoorthy *et al.*, 2008). The creation of intraspecific hybrids past the F_1 generation is a cost-efficient method that can capture most of the genetic variation present within this species, as well as provide a greater frequency of recombination events that will be useful for constructing comprehensive genetic linkage maps. Indeed, we observed more variation in the agronomic trait phenotypes of the pseudo-F_2 populations than we observed in the F_1 population (data not shown). Therefore, the establishment of more pseudo-F_2 populations will ultimately aid in increasing the resolution of the switchgrass genetic linkage map.

To date, several genetic linkage maps for switchgrass have been generated that have used a variety of molecular markers including SSRs (Liu *et al.*, 2012; Serba *et al.*, 2013), EST-SSRs (Okada *et al.*, 2010; Liu *et al.*, 2013), and RFLPs (Missaoui *et al.*, 2005). Recently, a study by Fielder *et al* (2015) used genotype-by-sequencing (GBS) methods to identify SNP molecular markers in F_1 progeny derived from a cross of two lowland cultivars, 'Kanlow' and 'Alamo'. SNP molecular markers are becoming preferred because they are abundant and wide spread in plant genomes, polymorphic, and easily detectable through NGS technologies (Kumar *et al.*, 2012). In this study, we used RNA-sequencing data to detect segregating SNPs within the pseudo-F_2 populations. SNPs identified by RNA-sequencing are advantageous over those identified by GBS in that they are found in gene coding sequences. Thus, the high-quality SNPs detected in this study may help correlate genotype to phenotype.

From the map created in our study, we were able to assign 170 of the previously unanchored contigs in the switchgrass reference genome (v 1.1) to one of the 18 linkage groups. While this is only approximately 2 Mb of the 593 Mb unanchored sequencing data in the switchgrass reference genome, the assignment of contigs to linkage groups helps to improve the genetic assembly of switchgrass.

Our previous study (Chapter 2) determined that the extent of polymorphism within switchgrass was still not fully captured at RNA-sequencing depths of 20-36X coverage. To further improve the linkage map created in this study, deeper RNA-sequencing coverage can be utilized to increase the number of high quality biallelic SNP markers. In addition, reference-guided RNA-seq identification of SNPs depends on the establishment of a high quality reference genome. The new version of the switchgrass genome (v 2.0) is expected to be released at the end of this year (2015). In the new version, greater than 90% of the genome sequence data for switchgrass has been assigned to one of the 18 linkage groups. The release of v 2.0 of the switchgrass reference genome will greatly facilitate the identification of high quality SNPs and produce greater resolution genetic linkage maps for this species.

Disease resistance in plants can be classified as qualitative, in which resistance to plant pathogens is controlled by one major resistance gene, or as quantitative, where disease resistance is determined by the combination effects of several genes (St.Clair, 2010). Major disease resistance genes functioning in qualitative disease resistance are usually those of the NB-LRR variety (Hammond-Kosack & Jones, 1997). Switchgrass has a large number of NB-LRR genes in its genome (Chapter 3), some of which are believed to contribute to rust disease resistance. Chr08A and Chr08B were found to be enriched with NB-LRR genes (Chapter 3); however, the QTL mapping performed in this study did not identify any significant peaks that correlated with

disease resistance on Chr08. This could be attributed to either the mapping resolution not being high enough or the RNA-sequencing not being deep enough. Additionally, the genetic resistance to this pathogen may be in a different genomic location. However, several smaller peaks were observed indicating that there is some level of genomic activity on this chromosome that may be contributing to disease resistance. Overall, one peak was identified with a 63% significance level around marker 92 cM on Chr02B, which has 34 predicted NB-LRR genes (Chapter 3). This marker is located in a gene that encodes for a casein kinase II protein, suggesting that it has a role in signal transduction. More studies are needed to determine what other genes are located in this region and which of these genes functions in resistance of switchgrass to switchgrass rust.

While major disease resistance genes are easy to identify, qualitative resistance to switchgrass rust could be significantly compromised by a shift in the genetics of the naturally occurring rust population. This would be especially devastating for large acreages of switchgrass plants grown in monoculture. Thus, switchgrass cultivars that have one or more QTLs controlling disease resistance to rust could be more durable and offer broad-spectrum resistance to several different pathogens or even haplotypes of the same pathogen. The employment of QTLs for disease resistance may be more useful for switchgrass breeding of sustainable biomass production.

CONCLUSION

This report is the first in switchgrass to utilize a pseudo-F_2 mapping population to create a genetic linkage map that can be used to identify genes and/or QTLs that may correlate to rust

resistance. We identified 966 high-quality SNP molecular markers that were distributed across the 18 linkage groups of switchgrass with an average spacing of 4.13 cM. QTL analysis for disease resistance indicated that a genomic region underlying the disease resistance phenotype may be located between markers 74 and 138 cM on Chr02B. Further research is needed to determine exactly what gene/genes are in this region and how they are contributing to the susceptibility of switchgrass to rust. This draft genetic linkage map between 'Alamo' and 'Dacotah' provides a solid foundation for additional QTL studies between these two cultivars. In addition, the resolution of this map will be significantly improved with further refinement of the switchgrass genome and deeper RNA-sequencing coverage.

ACKNOWLEDGEMENTS

This project was partially supported by a DOE-USDA grant (DE-FG02-08ER64629) awarded to BZ and JL. This project was also partially supported by a USDA-NIFA award (2011-67009-30133) to BZ.

AUTHOR CONTRIBUTIONS

BZ and JL designed the study. BZ, KM, and JM created the mapping populations. KM measured the agronomic traits for the F_1 and F_2 field-grown populations. GT isolated the *P. emaculata* strain VT2-1 used in this study. TF and CB collected the leaf tissue and performed the RNA isolations. TF and KM split the F_2 populations into three biological replicates for

greenhouse disease assays. TF carried out the RNA-sequencing analysis, SNP calling, disease assays, and identification of potential QTLs in R/qtl. AB created the genetic linkage map in R/QTL. CT manually counted the number of pustules on the infected switchgrass leaves.

REFERENCES

Aronesty E 2011. ea-utils: "Command-line tools for processing biological sequencing data" Durham, NC http://code.google.com/p/ea-utils.

Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**: 889-890.

Childs KL, Nandety A, Hirsch CN, Góngora-Castillo E, Schmutz J, Kaeppler SM, Casler MD, Buell CR. 2014. Generation of transcript assemblies and identification of single nucleotide polymorphisms from seven lowland and upland cultivars of switchgrass. *Plant Genome* **7**.

Fiedler JD, Lanzatella C, Okada M, Jenkins J, Schmutz J, Tobias CM. 2015. High-density single nucleotide polymorphism linkage maps of lowland switchgrass using genotyping-by-sequencing. *Plant Genome* **8**.

Field CB, Campbell JE, Lobell DB. 2008. Biomass energy: the scale of the potential resource. *Trends Ecol. Evol.* **23**: 65-72.

Frazier T, Shen Z, Zhao B, Bush E. 2012. First report of *Puccinia emaculata* infection on switchgrass in Virginia. *Plant Dis.* **97**: 424-424.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**: D1178-D1186.

Gustafson DM, Boe A, Jin Y. 2003. Genetic variation for *Puccinia emaculata* infection in switchgrass *Crop Sci.* **43**: 755-759.

Hagan A, Bowen K, Miller H 2015. An estimate of aboveground dry matter (ADM) yield loss due to rust on switchgrass. Auburn, AL.

Hammond-Kosack KE, Jones JDG. 1997. Plant disease resistance genes. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**: 575-607.

Hopkins AA, Vogel K, Moore K, Johnson K, Carlson I. 1995. Genotypic variability and genotype× environment interactions among switchgrass accessions from the midwestern USA. *Crop Sci.* **35**: 565-571.

Huang L-K, Bughrara SS, Zhang X-Q, Bales-Arcelo CJ, Bin X. 2011. Genetic diversity of switchgrass and its relative species in *Panicum* genus using molecular markers. *Biochem. Syst. Ecol.* **39**: 685-693.

Kumar S, Banks TW, Cloutier S. 2012. SNP discovery through next-generation sequencing and its applications. *Int. J. Plant Genomics* **2012**: 15.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Meth.* **9**: 357-359.

Liu L, Huang Y, Punnuri S, Samuels T, Wu Y, Mahalingam R. 2013. Development and integration of EST–SSR markers into an established linkage map in switchgrass. *Mol. Breed.* **32**: 923-931.

Liu L, Wu Y, Wang Y, Samuels T. 2012. A high-density simple sequence repeat-based genetic linkage map of switchgrass. *G3 Genes Genom. Genet.* **2**: 357-370.

Liu Y, Zhang X, Tran H, Shan L, Kim J, Childs K, Ervin E, Frazier T, Zhao B. 2015. Assessment of drought tolerance of 49 switchgrass (*Panicum virgatum*) genotypes using physiological and morphological parameters. *Biotechnol. Biofuels* **8**: 1-18.

Madakadze IC, Stewart K, Peterson PR, Coulman BE, Samson R, Smith DL. 1998. Light interception, use-efficiency and energy yield of switchgrass (*Panicum virgatum* L.) grown in a short season area. *Biomass Bioenergy* **15**: 475-482.

Martinez-Reyna J, Vogel K. 2002. Incompatibility systems in switchgrass. *Crop Sci.* **42**: 1800-1805.

McLaughlin S, Bouton J, Bransby D, Conger B, Ocumpaugh W, Parrish D, Taliaferro C, Vogel K, Wullschleger S 1999. Developing switchgrass as a bioenergy crop. In: Janick J ed. *Perspectives on new crops and new uses*. Alexandria, VA: ASHS Press.

Missaoui AM, Paterson AH, Bouton JH. 2005. Investigation of genomic organization in switchgrass (*Panicum virgatum* L.) using DNA markers. *Theor. Appl. Genet.* **110**: 1372-1383.

Narasimhamoorthy B, Saha MC, Swaller T, Bouton JH. 2008. Genetic diversity in switchgrass collections assessed by EST-SSR markers. *BioEnergy Res.* **1**: 136-146.

Okada M, Lanzatella C, Saha MC, Bouton J, Wu R, Tobias CM. 2010. Complete switchgrass genetic maps reveal subgenome collinearity, preferential pairing and multilocus interactions. *Genetics* **185**: 745-760.

Orquera Delgado GK. 2014. *Population biology of switchgrass rust (Puccinia emaculata Schw.).* Oklahoma State University.

Palmer NA, Donze-Reiner T, Horvath D, Heng-Moss T, Waters B, Tobias C, Sarath G. 2015. Switchgrass (*Panicum virgatum* L) flag leaf transcriptomes reveal molecular signatures of leaf development, senescence, and mineral dynamics. *Funct. Integr. Genomics* **15**: 1-16.

Parrish DJ, Fike JH. 2005. The biology and agronomy of switchgrass for biofuels. *Crit. Rev. Plant Sci.* **24**: 423-459.

Porter CL, Jr. 1966. An analysis of variation between upland and lowland switchgrass, *Panicum Virgatum* L., in central Oklahoma. *Ecology* **47**: 980-992.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

RC T 2015. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Open access available at: http://cran.r-project.org.

Sanderson MA, Adler PR, Boateng AA, Casler MD, Sarath G. 2006. Switchgrass as a biofuels feedstock in the USA. *Can. J. Plant Sci.* **86**: 1315-1325.

Saski C, Li Z, Feltus F, Luo H. 2011. New genomic resources for switchgrass: a BAC library and comparative analysis of homoeologous genomic regions harboring bioenergy traits. *BMC Genomics* **12**: 1-12.

Serba D, Wu L, Daverdin G, Bahri B, Wang X, Kilian A, Bouton J, Brummer EC, Saha M, Devos K. 2013. Linkage maps of lowland and upland tetraploid switchgrass ecotypes. *BioEnergy Res.* **6**: 953-965.

St.Clair DA. 2010. Quantitative disease resistance and quantitative resistance loci in breeding. *Annu. Rev. Phytopathol.* **48**: 247-268.

Sykes V, Allen F, Mielenz J, Stewart CN, Jr., Windham M, Hamilton C, Rodriguez M, Jr., Yee K. 2015. Reduction of ethanol yield from switchgrass infected with rust caused by *Puccinia emaculata*. *BioEnergy Res.*: 1-9.

Tobias CM, Sarath G, Twigg P, Lindquist E, Pangilinan J, Penning BW, Barry K, McCann MC, Carpita NC, Lazo GR. 2008. Comparative genomics in switchgrass using 61,585 high-quality expressed sequence tags. *Plant Genome* **1**: 111-124.

Uppalapati S, Serba D, Ishiga Y, Szabo L, Mittal S, Bhandari H, Bouton J, Mysore K, Saha M. 2013. Characterization of the rust fungus, *Puccinia emaculata*, and evaluation of genetic variability for rust resistance in switchgrass populations. *BioEnergy Res.* **6**: 458-468.

Wang Y, Zeng X, Iyer NJ, Bryant DW, Mockler TC, Mahalingam R. 2012. Exploring the switchgrass transcriptome using second-generation sequencing technology. *PloS ONE* **7**: e34225.

Wang Y, Zeng X, Peal L, Tang Y, Wu Y, Mahalingam R. 2013. Transcriptome analysis of nodes and buds from high and low tillering switchgrass inbred lines. *PloS ONE* **8**.

Wang YW, Samuels TD, Wu YQ. 2011. Development of 1,030 genomic SSR markers in switchgrass. *Theor. Appl. Genet.* **122**: 677-686.

Wright L, Turhollow A. 2010. Switchgrass selection as a "model" bioenergy crop: a history of the process. *Biomass Bioenergy* **34**: 851-868.

Zhang JY, Lee YC, Torres-Jerez I, Wang M, Yin Y, Chou WC, He J, Shen H, Srivastava AC, Pennacchio C, Lindquist E, Grimwood J, Schmutz J, Xu Y, Sharma M, Sharma R, Bartley LE, Ronald PC, Saha MC, Dixon RA, Tang Y, Udvardi MK. 2013. Development of an integrated transcript sequence database and a gene expression atlas for gene discovery and analysis in switchgrass (*Panicum virgatum* L.). *Plant J.* **74**: 160-173.

Table 4-1. Two year disease survey responses of six 'Alamo' x 'Dacotah' F_1 progeny that were used for creating three pseudo-F_2 mapping populations.

HR = highly resistant, R = resistant, MS = moderately susceptible, S = susceptible, and HS = highly susceptible.

F_1 Individual	2009 Disease Score	2010 Disease Score
DA5	S	HS
DA23	R	R
DA49	HS	S
DA90	HS	HS
DA120	HR	HR
DA139	MS	HS

Table 4-2. Agronomic trait measurements for the three field-grown pseudo-F_2 subpopulations.

Included are the average plant height (cm), average number of tillers, and average dry biomass (g). The number shown represents the average of two to three biological replicates per plant. Missing plants are indicated with a "-".

	Average Plant Height (cm)			Average Number of Tillers			Average dry biomass (g)		
Plant	**90x120**	**5x49**	**23x139**	**90x120**	**5x49**	**23x139**	**90x120**	**5x49**	**23x139**
1	113.5	143.9	93.1	36.5	144.0	94.0	70.0	265.0	60.0
2	116.4	135.2	125.3	286.5	244.7	100.0	372.5	530.0	145.0
3	109.2	123.9	121.9	62.0	181.7	266.0	212.5	428.3	450.0
4	50.8	120.2	147.3	6.0	160.7	162.0	-	380.0	428.3
5	119.1	128.70	67.7	263.3	311.3	113.0	380.0	556.7	45.0
6	-	103.6	-	-	68.0	-	-	106.7	-
7	89.8	122.8	89.3	42.5	297.3	165.0	57.5	546.7	142.5
8	-	87.8	91.4	-	24.7	45.0	-	30.0	45.0
9	-	130.4	120.2	-	377.0	140.0	-	566.7	190.0
10	98.6	117.1	145.6	106.0	130.3	100.0	102.5	216.7	465.0
11	115.2	101.9	117.3	83.0	96.3	228.0	152.5	131.7	320.0
12	121.9	102.5	111.3	194.3	124.7	140.0	238.3	176.7	257.5
13	136.0	139.7	143.9	281.3	223.3	195.5	563.3	393.3	402.5
14	119.4	150.7	99.9	41.0	169.3	55.0	55.0	381.7	60.0
15	-	127.3	109.2	-	119.7	186.0	-	203.3	408.3
16	123.6	116.8	122.8	144.7	100.3	141.0	316.7	226.7	385.0
17	83.4	140.3	106.1	27.5	139.7	75.0	27.5	380.0	101.7
18	103.7	131.5	174.8	215.5	203.3	178.3	290.0	490.0	510.0
19	123.6	-	210.0	146.0	-	185.7	247.5	-	1671.7
20	159.6	135.8	150.3	176.5	298.3	174.3	1020.0	641.7	491.7
21	138.9	124.2	135.2	79.3	270.0	204.7	193.3	633.3	333.3
22	108.4	153.5	129.5	18.0	203.3	80.0	15.0	563.3	220.0
23	95.7	116.6	141.7	25.0	78.0	211.0	30.0	151.7	375.0
24	144.8	154.9	112.0	161.3	197.7	84.7	380.0	528.3	203.3
25	109.2	166.5	85.1	55.0	213.3	32.5	50.0	581.7	77.5
26	-	131.0	129.5	-	256.7	102.5	-	446.7	75.0
27	-	136.6	121.9	-	150.3	280.5	-	310.0	450.0
28	140.5	110.7	149.6	90.0	187.3	261.7	235.0	248.3	558.3
29	-	155.8	148.2	-	351.3	302.0	-	953.3	477.5
30	113.5	140.0	-	36.0	160.7	-	80.0	486.7	-
31	102.5	171.6	-	33.5	271.7	-	47.5	830.0	-
32	135.6	139.7	-	134.0	469.0	-	285.0	846.7	-
33	-	154.9	108.4	-	228.5	98.0	-	402.5	95.0
34	-	128.4	-	-	161.0	-	-	211.7	-
35	-	138.9	-	-	186.0	-	-	325.0	-
36	112.6	134.6	-	99.0	161.3	-	90.0	545.0	-

37	99.6	121.9	145.3	101.0	164.7	178.3	75.0	221.7	316.7
38	-	134.6	134.6	-	188.3	192.0	-	298.3	345.0
39	-	86.4	100.2	-	140.0	190.7	-	105.0	291.7
40	112.0	145.6	149.6	156.3	128.0	342.0	258.3	303.3	600.0
41	75.4	123.9	581.7	35.0	175.0	229.0	47.5	323.3	215.0
42	-	129.0	141.4	-	132.3	169.0	-	296.7	281.7
43	132.1	143.7	114.3	156.3	226.3	142.7	441.7	466.7	231.7
44	-	-	106.1	-	-	62.0	-	-	86.7
45	-	153.3	142.0	-	358.3	342.0	-	811.7	458.3
46	-	134.6	83.4	-	151.0	49.0	-	286.7	45.0
47	148.2	-	150.1	76.0	-	275.0	250.0	-	628.3
48	-	162.1	98.6	-	161.5	87.5	-	542.5	80.0
49	80.0	-	95.7	11.0	-	72.0	15.0	-	90.0
50	102.9	-	123.6	65.5	-	76.5	77.5	-	137.5
51	-	-	169.1	-	-	289.3	-	-	1270.0
52	149.0	145.2	132.1	235.0	226.0	82.0	875.0	490.0	135.0
53	-	-	127.0	-	83.0	-	-	-	282.5
54	-	128.4	-	-	98.0	-	-	196.7	-
55	-	109.6	93.1	-	58.5	78.0	-	67.5	65.0
56	-	123.3	138.3	-	117.7	284.0	-	213.3	638.3
57	-	-	148.6	-	-	175.5	-	-	562.5
58	110.1	-	126.4	183.0	-	99.3	221.7	-	263.3
59	138.4	-	84.7	226.0	-	40.7	447.5	-	35.0
60	-	-	186.0	-	-	255.3	-	-	720.0

Table 4-3. Disease scores for the pseudo-F_2 populations inoculated with switchgrass rust under greenhouse conditions.

Rust-inoculated switchgrass plants were scored 10 dpi from 0 to 9 with 0 having no disease symptoms and 9 showing heavy/abundant sporulation and no necrosis. The number listed is the average of two to three biological replicates ± the standard deviation. "-" indicates missing or incomplete data.

Plant	DA90 x DA120	DA5 x DA49	DA23 x DA139
1	3.7 ± 0.6	4.0 ± 0.0	3.7 ± 1.5
2	4.0 ± 1.7	6.0 ± 2.6	-
3	6.7 ± 1.5	3.3 ± 1.5	5.3 ± 2.3
4	3.3 ± 0.6	6.0 ± 1.7	6.7 ± 1.5
5	4.3 ± 1.2	3.3 ± 1.2	-
6	7.0 ± 1.0	5.3 ± 0.6	-
7	4.0 ± 1.0	4.3 ± 0.6	-
8	3.0 ± 1.0	6.0 ± 2.0	-
9	-	7.7 ± 1.2	7.0 ± 1.0
10	3.7 ± 0.6	4.7 ± 1.2	7.0 ± 2.6
11	2.7 ± 1.2	7.3 ± 0.6	5.0 ± 1.0
12	1.0 ± 1.0	3.7 ± 0.6	-
13	3.0 ± 1.0	4.7 ± 2.1	4.0 ± 1.4
14	3.3 ± 1.5	6.3 ± 0.6	4.0 ± 1.0
15	1.0 ± 1.0	7.3 ± 1.2	3.7 ± 3.5
16	5.0 ± 0.0	4.7 ± 1.2	-
17	4.7 ± 2.1	3.7 ± 1.2	4.7 ± 0.6
18	3.0 ± 1.0	1.3 ± 0.6	1.0 ± 1.0
19	4.7 ± 1.2	-	5.7 ± 1.2
20	6.7 ± 1.2	5.7 ± 1.2	8.7 ± 0.6
21	3.7 ± 0.6	5.7 ± 2.1	7.0 ± 1.0
22	5.5 ± 0.6	5.7 ± 1.5	5.7 ± 0.6
23	5.3 ± 1,2	5.7 ± 1.2	2.7 ± 2.5
24	8.0 ± 1.7	-	3.0 ± 1.2
25	7.0 ± 1.7	4.0 ± 2.0	5.0 ± 0.0
26	5.3 ± 0.6	4.7 ± 0.6	4.5 ± 0.7
27	-	6.7 ± 0.6	4.5 ± 0.7
28	4.0 ± 1.0	3.7 ± 1.2	6.3 ± 0.6
29	6.3 ± 2.1	7.0 ± 0.0	4.3 ± 2.5
30	6.0 ± 1.7	6.7 ± 1.2	-
31	4.3 ± 0.6	6.5 ± 2.1	-
32	0.5 ± 0.7	3.3 ± 1.2	-
33	6.3 ± 1.5	4.0 ± 1.0	6.0 ± 1.0
34	6.3 ± 1.2	5.0 ± 1.0	-
35	3.3 ± 1.5	3.3 ± 0.6	1.3 ± 0.6
36	3.5 ± 2.1	5.3 ± 0.6	-

37	4.0 ± 0.0	6.7 ± 2.3	2.7 ± 1.5
38	8.3 ± 0.6	6.0 ± 2.6	3.3 ± 0.6
39	-	6.7 ± 1.5	2.7 ± 0.6
40	5.0 ±1.0	3.7 ± 2.1	5.0 ± 1.0
41	6.0 ± 2.0	7.7 ± 0.6	2.3 ± 0.6
42	3.3 ± 0.6	8.3 ± 0.6	3.3 ± 2.9
43	2.0 ± 0.0	5.7 ± 1.2	5.0 ± 1.7
44	-	-	4.3 ± 1.5
45	-	5.7 ± 1.5	2.0 ± 0.0
46	-	6.3 ± 0.6	4.3 ± 1.2
47	4.7 ± 1.5	-	1.3 ± 2.3
48	-	5.7 ± 0.6	2.3 ± 0.6
49	2.3 ± 0.6	-	2.3 ± 0.6
50	8.0 ± 1.0	-	3.7 ± 0.6
51	4.3 ± 2.1	-	3.0 ± 1.0
52	1.7 ± 0.6	6.7 ± 1.5	4.3 ± 0.6
53	-	-	2.7 ± 2.5
54	-	5.7 ± 0.6	-
55	-	6.7 ± 0.6	4.3 ± 1.5
56	-	8.0 ± 1.0	1.5 ± 0.7
57	-	-	3.0 ± 1.7
58	3.0 ± 1.7	5.3 ± 0.6	6.0 ± 0.0
59	3.0 ± 1.0	-	3.0 ± 0.0
60	0.5 ± 0.7	-	2.0 ± 0.0

Table 4-4. Summary of the 'Alamo' x 'Dacotah' genetic linkage map for the 18 linkage groups.

A total of 966 high-quality SNPs was used to create the map. cM = distance in centimorgan.

Linkage group	Number of SNP markers	Genetic length (cM)	Ave. spacing between markers (cM)	Max. spacing between markers (cM)
1A	66	383.2	5.9	28.8
1B	61	273.8	4.6	26.1
2A	54	178.4	3.4	39.5
2B	73	216.7	3.0	21.6
3A	49	192.7	4.0	27.0
3B	58	240.2	4.2	19.3
4A	46	320.6	7.1	27.4
4B	51	193.4	3.9	20.6
5A	80	215.2	2.7	30.3
5B	72	325.8	4.6	23.1
6A	40	94.5	2.4	17.5
6B	58	155.8	2.7	14.0
7A	29	137.5	4.9	29.0
7B	9	78.4	9.8	20.0
8A	22	79.3	3.8	13.5
8B	14	60.0	4.6	19.1
9A	68	258.7	3.9	26.1
9B	116	511.7	4.4	39.4
overall	966	3916.0	4.1	39.5

Figure 4-1. 'Alamo' is resistant to switchgrass rust whereas 'Dacotah' is susceptible.

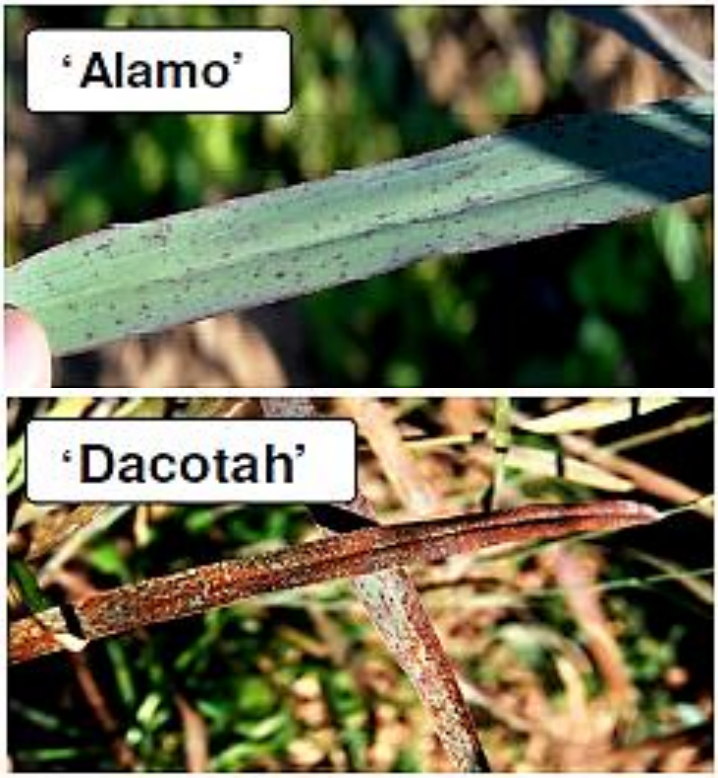


Figure 4-2. Two year disease surveys show that the F_1 progeny of 'Alamo' x 'Dactoah' are segregating for disease resistance.

The following are shown: results of 2009 disease survey (grey, n = 130) and results of 2010 disease survey (black, n =106). HR = highly resistant, R = resistant, MR = moderately resistant, MS = moderately susceptible, S = susceptible, HS = highly susceptible.

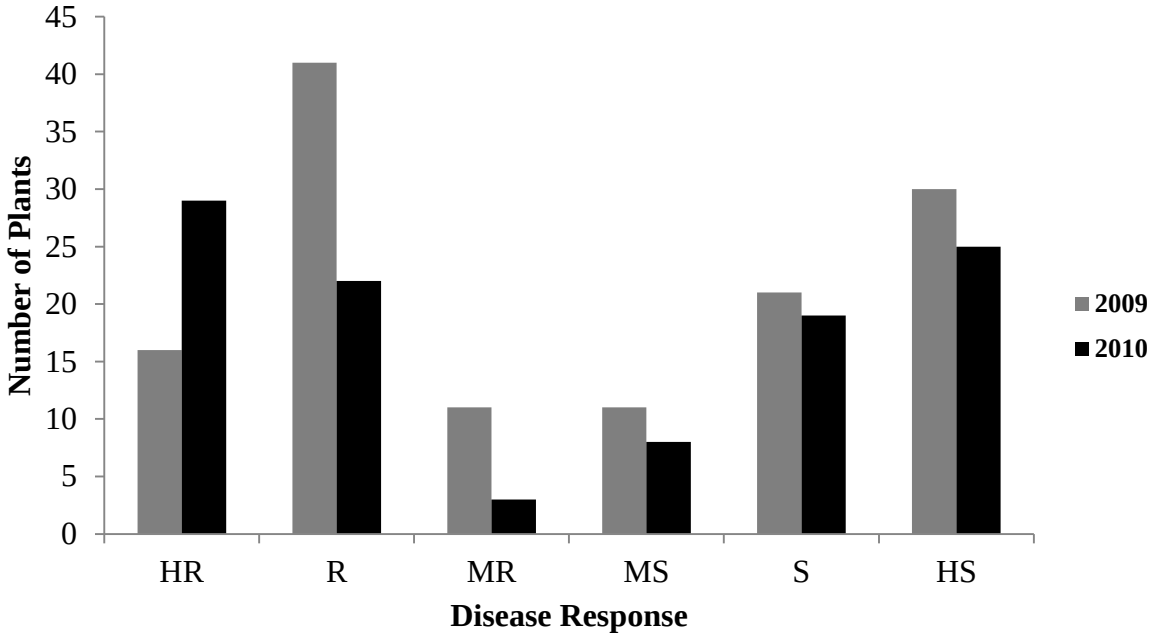


Figure 4-3. Distribution of disease scores for the three individual pseudo-F_2 populations, as well as all 154 individuals, inoculated with switchgrass rust under greenhouse conditions.

a) Disease distribution for the DA90 x DA120 pseudo-F_2 population, b) Disease distribution for the DA5 x DA49 pseudo-F_2 population, c) Disease distribution for the DA23 x DA139 pseudo-F_2 population, e) Disease distribution for all 154 individuals from the three pseudo-F_2 populations.

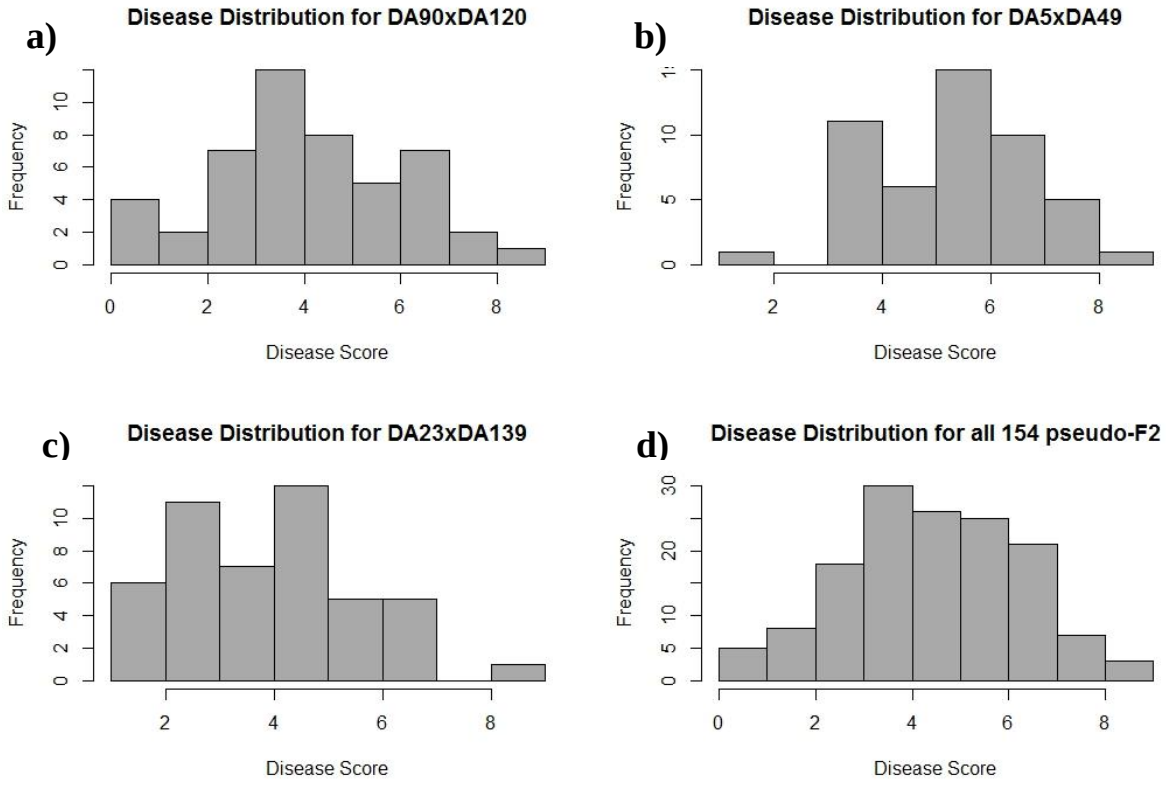


Figure 4-4. Genetic Linkage Map of 'Alamo' x 'Dacotah'.

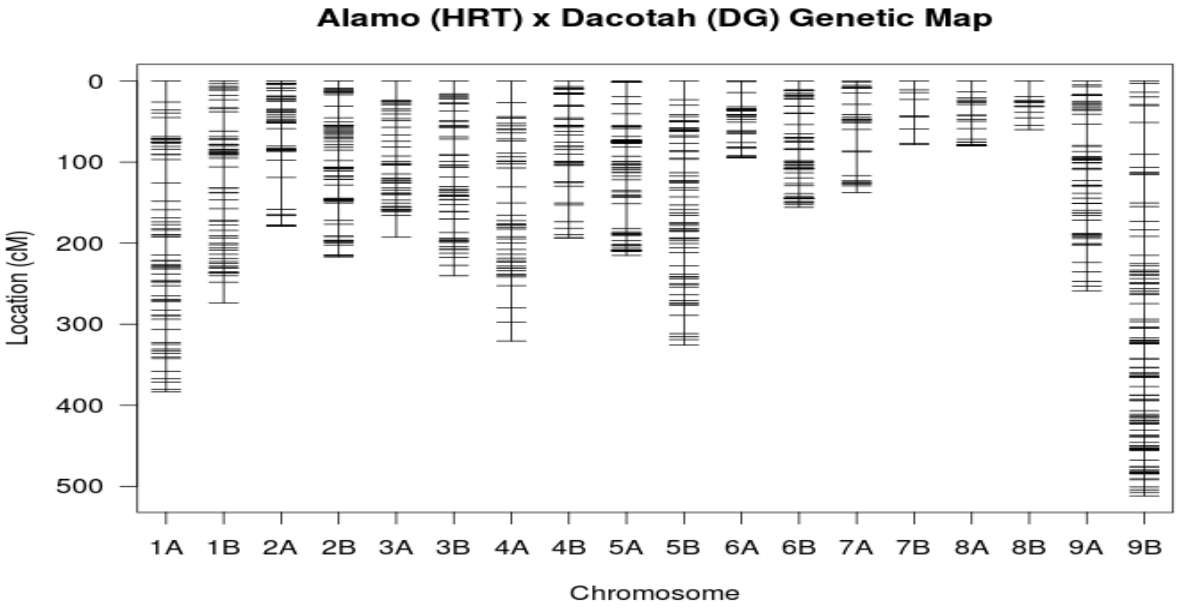


Figure 4-5. QTL analysis for disease resistance in switchgrass.

The dotted line represents the 63% significance threshold. 90% and 95% significant thresholds fall outside the scope of the figure. The chromosome numbers correspond to the 18 different linkage groups used in this study.

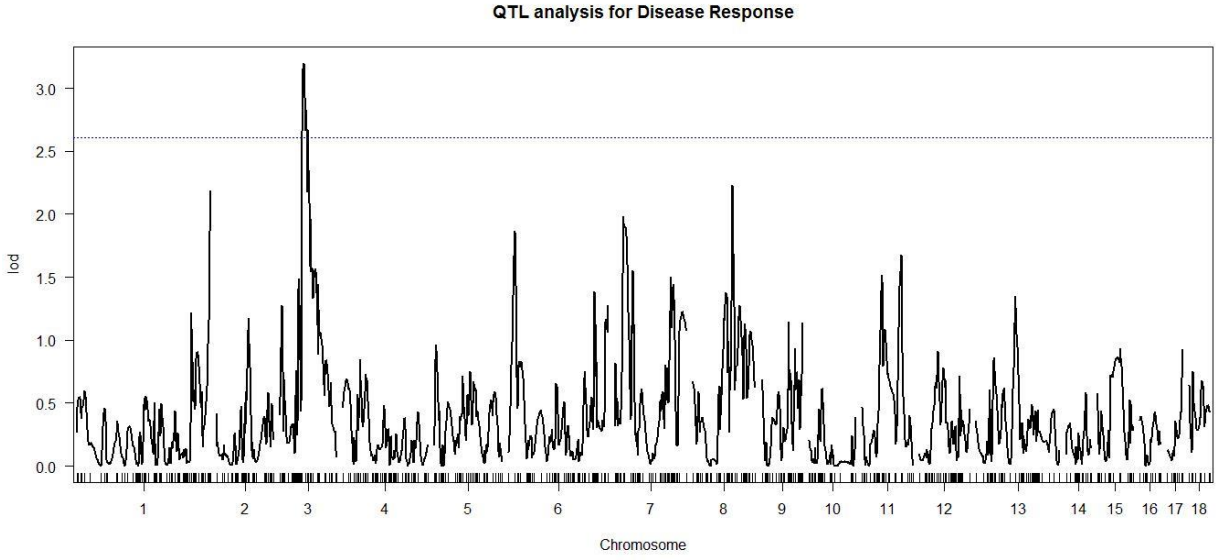


Figure 4-6. Confidence interval plot for the location of the potential disease-related QTL on chromosome 2B.

The green line spans the region of the chromosome that lies within a 95% confidence interval for phenotypic activity.

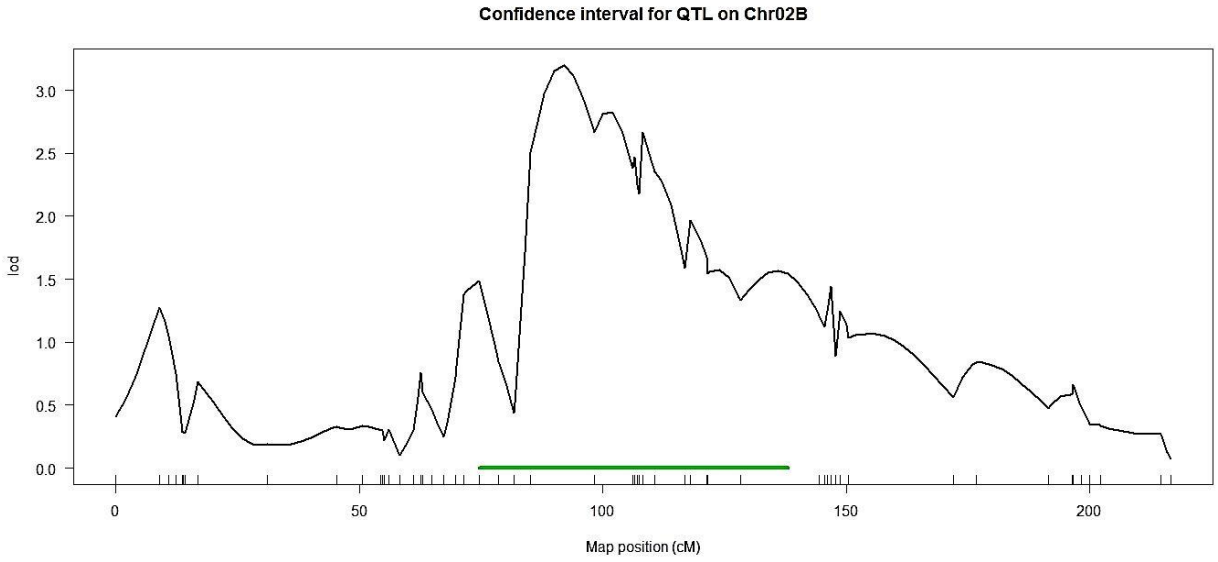
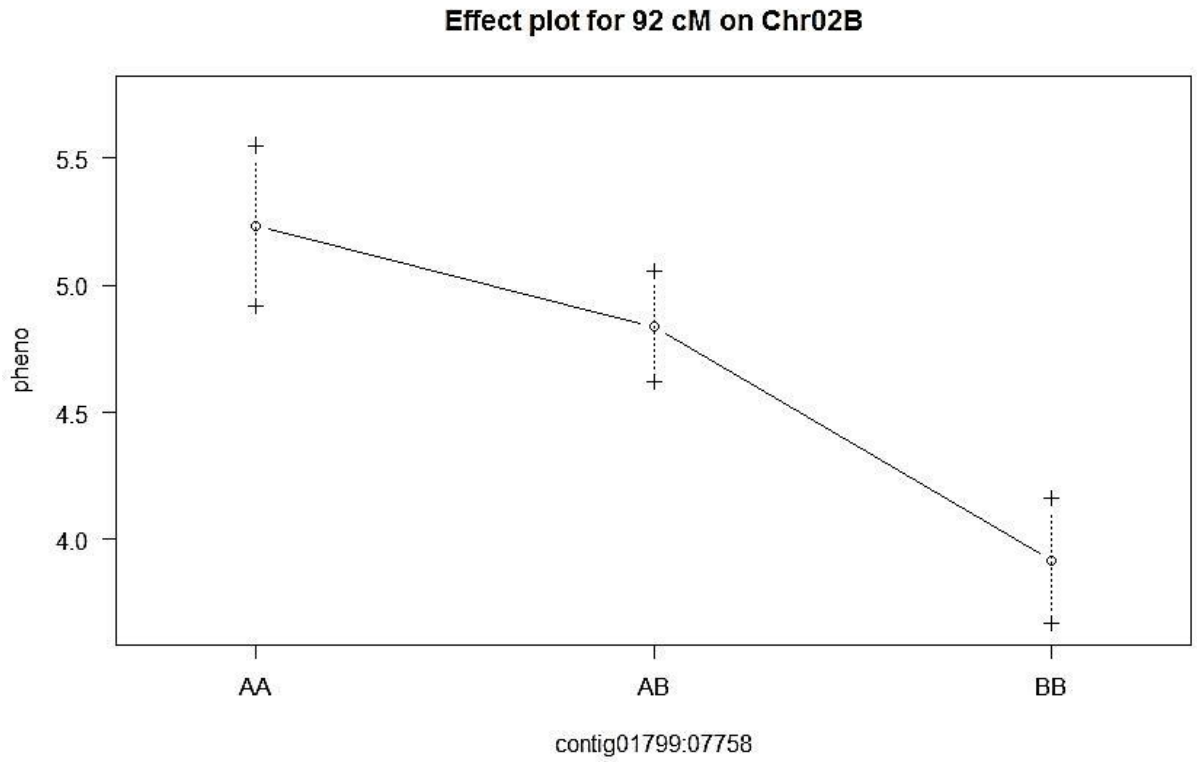


Figure 4-7. Effect plot of SNP marker 92 cM on chromosome 2B to determine which genotype is responsible for the phenotype observed at this location.



APPENDIX TO CHAPTER 4

MATERIALS AND METHODS

The pseudo-F_2 mapping population individuals were screened for their disease response to *P. emaculata* VT2-1 infection as described in Chapter 4. At ten dpi, rust disease severity was assessed based on a previously established scale (Gustafson *et al.*, 2003). The ratings for disease response ranged from 0 to 9 with 0 showing no disease symptoms and 9 showing numerous pustules and heavy abundant sporulation (Figure 4-8); however, the assignment of a disease rating based on leaf appearance may not accurately determine the degree of disease response. To quantitate the level of disease, a set of stencils with three 1 cm^2 sections were created (Figure 4-9) and the number of pustules in the three areas was manually counted. Finally, the number of pustules counted for all three sections was averaged together.

RESULTS

Preliminary results show that there is a correlation between the number of pustules counted and the average disease scores (Table 4-5). Thus, QTL identification was performed using the disease scores that were initially assigned.

Table 4-5. Supplemental Data: Pustule counts of infected switchgrass leaves correlate with disease scores.

Listed below is pustule count data for nine DA90 x DA120 F_2 progenies. Three 1 cm^2 sections were counted for three biological replicates of each line. Numbers listed are the averages ± the standard deviation.

DA90 x DA120 F_2 progeny	Average disease score	Average pustule count
60	0.5 ± 0.5	1.5 ± 1.6
15	1.0 ± 1.0	1.4 ± 2.5
49	2.3 ± 0.6	4.1 ± 4.3
58	3.0 ± 1.7	5.4 ± 5.1
7	4.0 ± 1.0	11.9 ± 7.1
40	5.0 ± 1.0	25.0 ± 14.4
41	6.0 ± 2.0	31.6 ± 21.1
25	7.0 ± 1.7	49.0 ± 31.1
38	8.0 ± 0.6	53.0 ± 13.6

Figure 4-8. Supplemental Data: Disease scoring of the pseudo-F_2 mapping population individuals.

The pseudo-F_2 mapping population individuals were given a disease rating based on the scale established by Gustafson *et al* (2003).

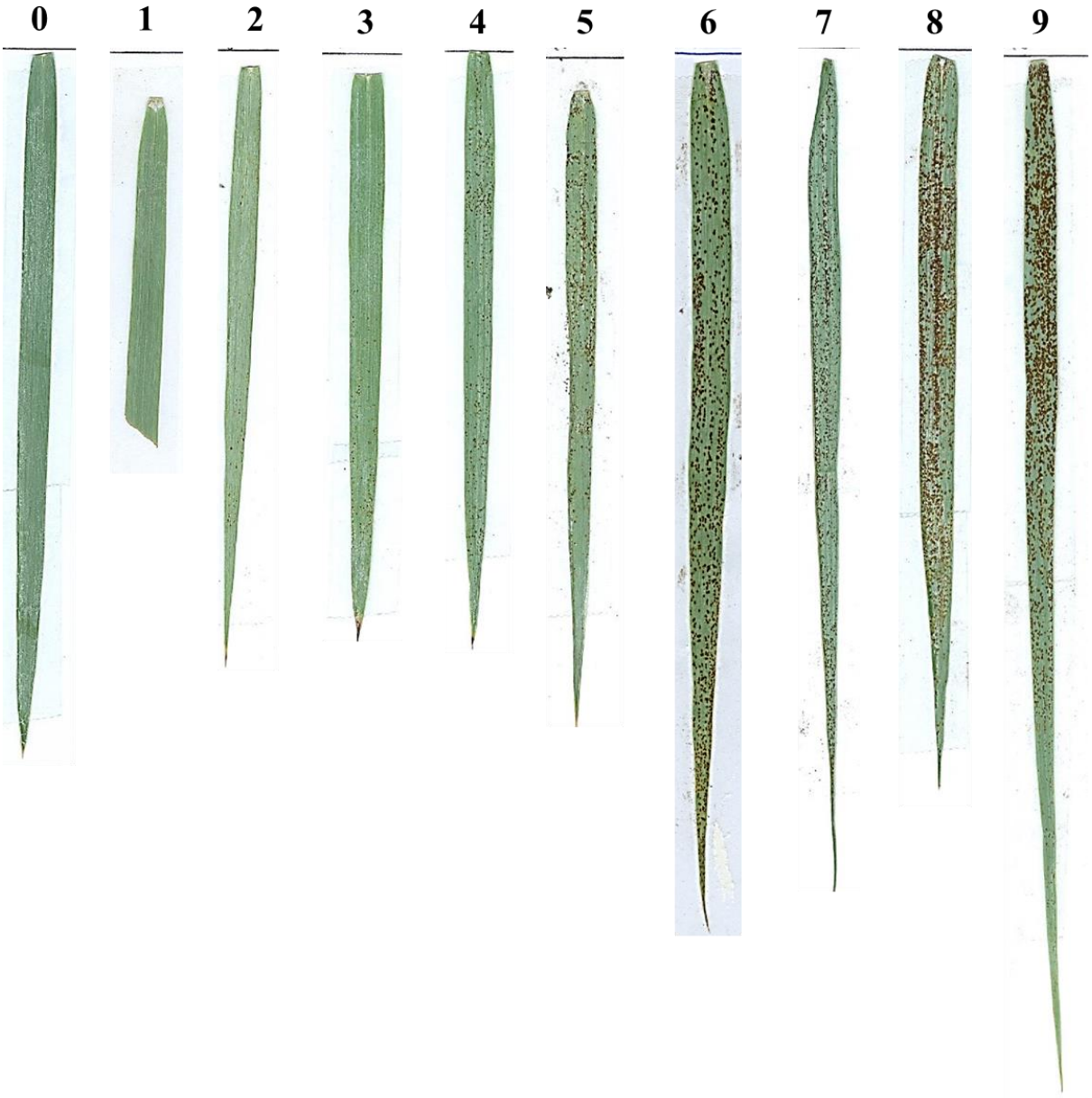
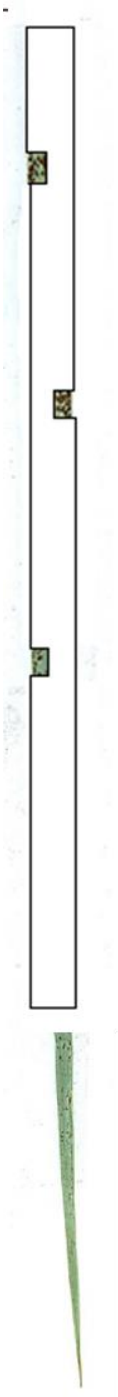


Figure 4-9. Supplemental Data: Creation of stencils to quantitate the disease response of the pseudo-F_2 mapping population individuals.

A set of stencils was created based on the length of the sampled leaves. Shown is a depiction of one stencil with three 1 cm^2 sections cut out. The number of pustules in each section was manually counted and averaged together to quantify the level of disease response.



CHAPTER 5: Overexpression of the Arabidopsis *SHN3* transcription factor affects the cell wall composition and disease resistance of transgenic switchgrass plants

ABSTRACT

Switchgrass is a second generation bioenergy crop that can generate large amounts of renewable biomass. For switchgrass to be efficiently used as a biofuel, the quality of the lignocellulosic feedstock material must be improved. The overexpression or repression of master regulators of cell wall deposition, such as the SHINE (SHN) family of transcription factors, offers a unique opportunity to simultaneously alter numerous cell wall components. Here, we characterize the effects of overexpressing the Arabidopsis *SHN3* transcription factor in switchgrass. In comparison to wild-type plants, three *AtSHN3*-overexpressing switchgrass lines, SHN3 4-1, SHN3 6-3, and SHN3 7-2, are stunted in their growth by 32%, 17%, and 40%, respectively. No significant differences were found in the lignin content of the SHN3 transgenic switchgrass lines; however, SHN3 7-2 plants had a 10% increase in cellulose content over the wild type plants. Moreover, significant differences in several hemicellulose sugar monomers, such as xylose, glucose, and arabinose, were observed for SHN3 5-2, SHN3 7-2, and SHN3 4-1 plants, respectively. Despite slight alterations in all three cell wall components, the *AtSHN3*-overexpressing transgenic switchgrass plants demonstrated similar stem mechanical strength to wild type plants when subjected to tensile-torsion analysis. The SHN3 4-1, SHN3 5-2, and SHN3 6-3 transgenic lines were also more susceptible to switchgrass rust than the wild-type plants. Our findings show that AtSHN3 has the potential to coordinate the deposition of key cell wall components in switchgrass. In addition, we found that slight changes in these components may

significantly impact disease resistance. Since switchgrass lines with increased cellulose and decreased lignin are desirable for bioethanol production, the results of this study provide a foundation for studying how alterations in cell wall deposition affect disease resistance in this important bioenergy crop.

INTRODUCTION

To meet growing energy demands, it is estimated that 22.3 million acres of arable cropland will need to be allocated to biofuel production by the year 2030 (Perlack *et al.*, 2005; Sanderson & Adler, 2008). Perennial forage grasses are an attractive source of sustainable energy and as such, they have been extensively studied as promising second generation bioenergy crops (Sanderson & Adler, 2008). These second generation biofuel feedstocks, such as switchgrass, contain large amounts of lignocellulosic biomass that can provide an inexpensive and abundant source of renewable energy (Naik *et al.*, 2010).

Lignocellulosic feedstock material is comprised of three major components: lignin, hemicellulose, and cellulose. In conjunction with other smaller components, such as pectin and proteins, these molecules function together to form the structural base of the plant cell wall (Jørgensen *et al.*, 2007). The concentrations of lignin, hemicellulose, and cellulose in plant cell walls has been shown to vary among plant species (Jørgensen *et al.*, 2007). Grasses have been found to contain anywhere from 25-40% cellulose, 35-50% hemicellulose, and 10-30% lignin (Sun & Cheng, 2002).

Switchgrass is a C4 perennial grass that is commonly found growing across the vast prairie region of North America. There are two main ecotypes of switchgrass, lowland and

upland, that differ in their habitat preference. Lowlands are typically found growing across the warm southern plains of the United States whereas uplands tend to grow across the northern prairies into the southern parts of Canada (Porter, 1966). There are also morphological differences between the two ecotypes. Lowland ecotypes have thicker stems, wider leaves, and taller tillers that their upland counterparts (Porter, 1966). In addition, the two ecotypes vary significantly in overall biomass production. Lowland varieties have been shown to produce on average 12.9 Mg ha^{-1} of biomass per year while the upland varieties have been shown to produce on average 8.7 Mg ha^{-1} of biomass per year (Wullschleger *et al.*, 2010). Currently, several commercial varieties of switchgrass have been released that are suitable for large-scale sustainable biomass production including lowland varieties 'Alamo' and 'Kanlow', as well as upland cultivars 'Cave-In-Rock' and 'Summer' (McLaughlin & Kszos, 2005; Bouton, 2007).

For switchgrass to be fully utilized as a bioenergy crop, the quality of the lignocellulosic component of the biomass must be improved. Significant effort has been put into identifying elite switchgrass germplasm from already existing cultivars and into developing the best management practices for optimal biomass output (Sanderson *et al.*, 1996; McLaughlin *et al.*, 1999). In addition, traditional breeding methods have been employed to enhance certain characteristics of switchgrass feedstock including biomass production and forage digestibility (Hopkins *et al.*, 1993; Taliaferro & Das, 2002b). Considering the time constraints of current breeding practices, it takes approximately 10 years to develop a new switchgrass cultivar with enhanced characteristics using traditional methods (Sanderson *et al.*, 2006).

Recently, genetic engineering practices have been used to create transgenic switchgrass lines with altered cell wall compositions. Since lignin is a limiting factor in the use of lignocellulosic biomass for bioethanol production, several studies in switchgrass have used

RNAi technology to knock down genes coding for key enzymes in the lignin biosynthesis pathway including 4-coumarate:coenzyme A (4CL) (Xu *et al.*, 2011a), cinnamyl alcohol dehydrogenase (CAD) (Fu *et al.*, 2011b; Saathoff *et al.*, 2011), and caffeic acid *O*-methyltransferase (COMT) (Fu *et al.*, 2011a). Xu *et al* (2011) found that in comparison to the wild type plants, transgenic switchgrass lines with reduced *4CL* activity had a 22% reduction in overall lignin and released 57.2% more fermentable sugar with diluted acid pretreatment. Alternatively, two independent studies found that down-regulating *CAD* in switchgrass results in 23% less lignin and cutin (Saathoff *et al.*, 2011) and 14-22% less lignin (Fu *et al.*, 2011b), respectively. Finally, down-regulation of the *COMT* gene in switchgrass produced up to 38% more ethanol using current biomass fermentation practices (Fu *et al.*, 2011a).

An alternative to directly targeting components of the lignin pathway is to identify transcription factors that may play a role in the regulation of cell wall components. The SHINE (SHN) family of AP2 transcription factors consists of three members and was first reported in Arabidopsis (Aharoni *et al.*, 2004). Arabidopsis *shn* mutants were found to have aberrant deposition of epicuticular wax and altered flower morphology (Aharoni *et al.*, 2004). Shi *et al* (2011) used a microRNA approach to show that SHN proteins also contribute to floral organ development. Recently, the Arabidopsis *SHN2* gene was overexpressed in rice and resulted in transgenic plants with a 34% increase in cellulose content and a 45% decrease in lignin (Ambavaram *et al.*, 2011).

Despite its importance as a promising bioenergy crop, only a handful of studies in switchgrass have aimed to identify transcriptional control mechanisms underlying cell wall deposition (Xu *et al.*, 2012; Shen *et al.*, 2013). In this study, we created transgenic switchgrass plants overexpressing the *AtSHN3* cDNA sequence from Arabidopsis. The transgenic lines

display stunted growth and altered compositions of lignin, cellulose, and hemicellulose in their cell walls. Additionally, we report that overexpressing *AtSHN3* in switchgrass affects disease resistance. The results of this study provide insights into the regulatory mechanisms that control cell wall composition of switchgrass. In addition, these results suggest that changes in the lignocellulosic content of switchgrass may affect disease resistance of this important bioenergy crop.

MATERIALS AND METHODS

Cloning of the AtSHN3 gene

A plasmid containing the *AtSHN3* (TAIR accession: U51209) cDNA was obtained from TAIR-ABRC. Next, the *AtSHN3* cDNA fragment was cloned using a 50 µl PCR reaction with the following components: 25 µl High-Fidelity iProof master mix (Bio-Rad, Hercules, CA, USA), 10 µl plasmid DNA, 10 µl ddH_2O, 2.5 µl 10 µM forward primer (5' caccGAATTCATGGTACATTCGAAGAAGTTCC 3'), and 2.5 µL 10 µM reverse primer (5' CGTCTGCAGGACCTGTGCAATGGATCCAGATC 3'). The PCR reaction was run with an initial denaturation step at 98°C for 3 min, followed by 30 cycles of denaturation at 98°C for 30 sec, annealing at 57°C for 45 sec, and extension at 72°C for 1 min, and then completed with a final extension at 72°C for 7 min. Successful amplification of the PCR product was visualized using a 0.8% agarose gel and the PCR product was purified using a kit.

Insertion of AtSHN3 into the pVT1629 Gateway-compatible binary vector

After the *AtSHN3* PCR product was purified, it was cloned into the pENTR/D-TOPO vector (Invitrogen, Waltham, MA, USA). The pENTR/D-TOPO vector plasmid containing the *AtSHN3* cDNA sequence was then digested with restriction enzymes *Pvu1* and *Nvu1* in order to release the insert, which was subsequently purified using a kit. Next, an LR reaction (Invitrogen, Waltham, MA, USA) was performed between the purified pENTR/D-TOPO/*AtSHN3* product and the pVT1629 destination vector (Xu *et al.*, 2011a). The final construct, pVT1629-*AtSHN3*, was then conjugated into *Agrobacterium* strain *AGL1*.

Switchgrass callus formation and Agrobacterium*-mediated transformation*

Seeds of the HR8 variety of the switchgrass cv. Alamo (Xu *et al.*, 2011b) were dehusked and dehulled by incubating in 60% sulfuric acid on a rotary shaker at room temperature for 1 h. The seeds were washed with distilled water and transferred to a 50 ml conical tube. Next, 50% bleach with 0.01% tween 20 was added to the seeds and the tube incubated on a rotary shaker at room temperature for 1 h. The bleach was discarded in a laminar flow hood and the seeds were washed with distilled water until no bleach odor remained. The seeds were dried on a sterilized paper towel and transferred to callus induction medium. After 4-6 weeks, embryogenic calli were subcultured onto callus induction medium containing 20 g l^{-1} proline. Ten days before transformation, embryogenic calli were subcultured again onto callus induction medium containing proline and 200 µM acetosyringone. The method for *Agrobacterium*-mediated transformation of switchgrass followed that previously described (Li & Qu, 2011). After two rounds of culture on selection medium, transformed calli were transferred to regeneration

medium. Following regeneration and root formation, regenerated plantlets were transplanted into pots containing MiracleGro Moisture Control soil.

DNA extraction and confirmation of transgenic switchgrass plants

Leaves of putative transgenic and wild-type switchgrass plants were collected and immediately frozen in liquid nitrogen. The tissue was ground into a fine powder using a mortar and pestle and transferred to a 50 mL conical tube. DNA was extracted using a modified 2x CTAB protocol as previously described (Porebski *et al.*, 1997). The quality and quantity of the DNA was assessed using agarose gels and a Nanodrop-D1000 (Nanodrop, Wilmington, DE, USA). The switchgrass DNA was then sent to Lofstrand Labs Limited (Gaithersburg, MD, USA) for Southern blot analysis. Briefly, a total of 10 µg of genomic DNA was restriction enzyme digested with Hind*III*. DNA fragments were then separated using gel electrophoresis and probed with a portion of the hygromycin selection gene to detect transgene insertion.

RNA extractions and qPCR analysis

Flag leaves of R3 stage switchgrass tillers were collected from greenhouse grown switchgrass plants and immediately frozen in liquid nitrogen. The tissue was stored at -80°C until further analysis. Tissue samples were collected for three biological replicates of both the transgenic and wild-type plants. Using a mortar and pestle, the leaves were ground into a fine powder and transferred to a 2 mL epi-tube. Total RNA was extracted using the TRIzol reagent (Invitrogen, Grand Island, NY, USA) according to the manufacturer's protocol. The quality and quantity of the RNA was assessed using a Nanodrop-D1000 (Nanodrop, Wilmington, DE, USA).

First strand cDNA was synthesized using a DyNAmo cDNA synthesis kit (Thermo Fisher, Waltham, MA, USA). Briefly, the following components were added to a 0.2 mL PCR tube: 10 µL 2x RT buffer, 2 µL reverse transcriptase enzyme, and 1 µL oligo(dT) primer. Next, 1 µg of RNA was added to each tube along with ddH_2O to make the final reaction volume 20 µL. Reverse transcription was performed with an initial primer extension at 25°C for 10 min, followed by cDNA synthesis at 37°C for 30 min, and concluded at 85°C for 5 min. Following RT-PCR, the reactions were diluted 1:5 by adding 80 µl of ddH_2O to each tube. Quantitative Real Time-PCR (qPCR) was performed using an Applied Biosystems Power SYBR Green PCR Master Mix (Grand Island, NY, USA). The following reagents were added to each reaction: 10 µL of 2x SYBR green mix, 5 µL ddH_2O, 3 µL 1:5 diluted cDNA, 1 µL 10 µM forward primer, and 1 µL 10 µM reverse primer. The same primers for *AtSHN3* gene cloning were also used for qPCR analysis. All reactions were run on an Applied Biosystems 7300 Real Time PCR machine with the following conditions: 1) an initial denaturation and enzyme activation step at 95°C for 10 min and 2) 40 cycles of denaturation (95°C for 30 sec), annealing (60°C for 30 sec), and extension (72°C for 1 min and 30 sec). After the reactions had completed, the threshold was manually set to 3.0 and the data exported for analysis.

Phenotypic characterization of SHN3 transgenic switchgrass lines

In the middle of July 2015, individual E2 to E3 stage tillers from all four transgenic SHN3 switchgrass lines, along with the HR8 control, were clonally propagated and planted in gallon-size pots containing Miracle-Gro ® Moisture Control potting mix. The plants were maintained in a greenhouse at a 16 h photoperiod with supplemental lighting used as needed.

After 3 months of growth, the overall heights, flag leaf lengths, flag leaf widths, and I2 stem widths of 4 R3 stage tillers were measured for three biological replicates of each transgenic line as well as the wild-type control. Finally, all plants were harvested at ground level and weighed to determine fresh biomass yield.

Stem sectioning and histochemical staining for lignin and cellulose

The second internode (I2) of R3 stage tillers was selected for histochemical staining. I2 was characterized as the first full-length stem segment, located between the first and second distinguishable nodes, from the base of the plant. The I2 segments of transgenic and wild-type plants were cut into 40 µm sections using a microtome. Changes in the lignin and cellulose content of the transgenic switchgrass plants were visualized using Weisner reactions and calcufluor staining, respectively. The protocols for these reactions were performed as previously described (Pradhan Mitra & Loque, 2014). The Weisner stained stem sections were visualized using a Zeiss compound light microscope and the calcufluor stained stem sections were visualized using a fluorescence Zeiss AxioImager.M1 microscope mounted with a Zeiss AxioCam MRm.

Measurement of acid soluble and acid insoluble lignin

I2 segments of R3 stage tillers for three biological replicates of each transgenic line, as well as the HR8 control, were dried in an oven at 48°C and then ground into a coarse powder using a coffee grinder. Acid soluble and insoluble lignin content were determined using the procedure established by the National Renewable Energy Laboratory (Sluiter *et al.*, 2005).

Briefly, 300 mg of ground switchgrass samples were added to a pressure tube along with 3 mL of 72% sulfuric acid to hydrolyze the tissue. The tubes incubated at 30°C for 1 h with manual stirring every 5 min. Following incubation, 84 mL of deionized water was added to each tube to dilute the sulfuric acid to a concentration of 4%. The tubes were then autoclaved at 121°C for 1 h. Next, the tubes cooled to room temperature and the mixture was vacuum-filtered through a porcelain crucible. The filtrate, which contained the acid soluble lignin, was collected and diluted to a volume sufficient to obtain a UV absorption value of 0.7-1.0 at 205 nm. The acid insoluble reside, which remained in the porcelain crucible, was dried in an oven at 105°C overnight and then weighed to determine the acid insoluble lignin content.

Measurement of cellulose and hemicellulose

A second set of I2 samples of R3 stage tillers for three biological replicates of control and transgenic plants were also dried in an oven at 48°C. The samples were ground with a 2010 Geno/Grinder at 1500 rpm. The fine powder was then made into alcohol insoluble residue (AIR) and de-starched as described previously (Bartley *et al.*, 2013). The de-starched AIR was used for cellulose and hemicellulose assays. Hemicellulose monosaccharides were released by 4M TFA treatment for 2 h and then measured by HPLC. The pellets after TFA treatment were used for anthrone cellulose assay as described previously (Bartley *et al.*, 2013). Briefly, pellets were hydrolyzed by 72% sulfuric acid to release cellulosic glucose. The cellulosic glucose was quantified by a colorimetric reaction with anthrone reagent and read on a plate reader at OD_{625nm}.

Solvent-submersion tensile-torsion measurements to determine stem stiffness and mechanical strength

Fresh I2 stem segments of R3 stage switchgrass tillers were subjected to solvent-submersion tensile-torsion analysis using a TA Instruments AR G2 rheometer (New Castle, DE, USA). The I2 segments were cut into 2 cm long fragments and then split longitudinally into four different sections. The samples were then fully saturated with ethylene glycol and stored at room temperature for future analysis. On the day of analysis, the samples were secured with tension clamps using 15cNxm torque and 1N static tensile force.

All of the testing steps were operated at a frequency of 0.5Hz and a stress setting of 50,000 Pa. Storage modulus analysis, which is a reflection of stem stiffness, was conducted at room temperature by equilibrating the samples at 25ºC for 5 min and then running the stress sweep. At least three observations were recorded for each sample type. Ultimate fracture tests of ethylene glycol saturated stem samples were conducted in tensile-torsion mode at room temperature. The specimens were clamped at both ends with slight tensile force (1N) to hold the sample vertically straight. Fracture tests were performed under continuous flow conditions with a shear stress increasing from 1E5Pa to 1E8 Pa. The tests were performed four times per sample. Data acquisition was performed in linear mode with a total collection time of 33 min and a total point set at 300. Tests were concluded once the specimens failed.

Disease assays of AtSHN3 transgenic plants

The *AtSHN3*-overexpressing transgenic lines and the HR8 control plant were clonally split into three biological replicates. Each biological replicate was planted in a pot containing MiracleGro Moisture Control soil and grown in the greenhouse under a 16 h photoperiod. Freshly collected *Puccinia emaculata* urediniospores were mixed 1:10 with talcum powder and hand inoculated on the first fully expanded leaf of E2 stage tillers. The plants were placed in a chamber with a humidifier and kept under 100% humidity for 16 h. Ten days post inoculation, the severity of disease was scored according to the scale established by Gustafson *et al* (2003).

Statistical analyses

All statistical analyses were performed using Student's t-tests with a significance level of 0.05.

RESULTS

Creation of AtSHN3 over-expressing transgenic switchgrass plants

Following *Agrobacterium* transformation of somatic embryogenic switchgrass callus, a total of 49 potential *AtSHN3*-overexpressing switchgrass plants were regenerated and transplanted into soil. These 49 plants were derived from 7 independent transformation events. Four plants, representing 4 independent transformation events, were selected for further analysis. DNA samples for all four transgenic lines, as well as the wild-type HR8 control, were sent for Southern blot. The results of Southern blot analysis found that three of the four selected lines contained multiple transgene insertions (Figure 5-1). SHN3 4-1 contained three copies of the

transgene whereas SHN3 5-2 and SHN3 7-2 contained two copies of the transgene. SHN3 6-3 was found to be the only line that had a single insertion copy of the transgene.

AtSHN3-overexpressing transgenic switchgrass are phenotypically different than wild-type plants

Phenotypic differences were compared between greenhouse grown transgenic SHN3 plants and the wild-type HR8 control plants after 3 months of growth. Overall, the *AtSHN3*-overexpressing transgenic plants appeared shorter than the HR8 control (Figure 5-2). Several agronomic traits were measured for all plants in order to evaluate the degree of stunted growth. These included number of tillers, tiller height, leaf length, leaf width, stem size, and overall biomass. The number of tillers produced was not statistically different between the SHN3 4-1 ($p = 0.33$), SHN3 6-3 ($p = 0.14$), and SHN3 7-2 ($p = 0.46$) transgenic lines and the wild-type plants (Table 5-1); however, the SHN3 5-2 transgenic lines did have significantly fewer tillers ($p = 0.01$) than the HR8 control plants. All of the plants in this study were found to contain on average between 6 and 12 tillers per line. Measurement of tiller height found that the SHN3 4-1 ($p < 0.001$), SHN3 6-3 ($p = 0.02$), and SHN3 7-2 ($p < 0.001$) plants were significantly shorter than the wild-type plants (Table 5-1).

Corresponding to the difference in overall height, the leaf lengths of the SHN3 4-1 ($p = 0.03$) and SHN3 7-2 ($p = 0.04$) transgenic lines were statistically different from the control plants (Table 5-1). In addition, the leaf widths of SHN3 4-1 ($p = 0.005$) and SHN3 7-2 ($p = 0.03$) transgenic plants were found to be significantly shorter than the HR8 control (Table 5-1). Only the SHN3 4-1 line, however, exhibited significantly smaller stems ($p = 0.02$) than the HR8

control plants (Table 5-1). An indicator of change in cell wall composition is abnormal lengthening of internode stem segments (Jiang *et al.*, 2012). In this study, we found that the second internode from the base of the plant was significantly shorter for the SHN3 4-1 ($p = 0.002$) and SHN3 7-2 plants ($p = 0.002$) (data not shown). Despite their stunted growth, only the SHN3 4-1 ($p = 0.01$) and SHN3 6-3 ($p = 0.01$) transgenic lines produced significantly less biomass than the wild-type plants under greenhouse conditions (Table 5-1).

Transgene copy insertion number correlates with AtSHN3 gene expression in transgenic switchgrass plants

Since the SHN3 4-1, SHN3 5-2, and SHN3 7-2 plants have multiple copies of the transgene (Figure 5-1), qPCR was performed to determine if there was a correlation between transgene copy number and *AtSHN3* gene expression. In comparison to SHN3 6-3, which has a single copy of the transgene, we found that the level of *AtSHN3* gene expression increased with increasing transgene copy number (Figure 5-3). SHN3 7-2 and SHN3 4-1 exhibited greater levels of transgene expression. These two lines were the two shortest among all of the transgenic lines and have two and three copies of the transgene, respectively. Thus, differences in the expression levels of *AtSHN3* may be contributing to the stunted growth phenotype observed in SHN3 7-2 and SHN3 4-1 switchgrass plants. SHN3 5-2 and SHN3 6-3 plants appear to be phenotypically similar and not surprisingly, only a modest difference in transgene expression was observed between these two lines.

AtSHN3-overexpressing switchgrass plants show an aberrant deposition of lignin and cellulose

Overexpression of *AtSHN2* in transgenic rice plants resulted in an overall increase in cellulose and a decrease in lignin (Ambavaram *et al.*, 2011). In this study, we used I2 segments of R3 stage greenhouse grown switchgrass tillers to determine if similar results could be observed with the *AtSHN3*-overexpressing transgenic switchgrass lines. This segment was chosen because it is closest to the ground and therefore carries more of the weight load of the above ground biomass. Weisner staining of I2 stem sections revealed that some of the *AtSHN3* transgenic lines appeared to have slightly less lignin content than the wild-type control (Figure 5-4). In addition, calcufluor staining of I2 stem sections showed that the majority of the *AtSHN3* transgenic lines also had a greater concentration of cellulose than the HR8 plants, with SHN3 5-2 being the only line that resembled the wild type phenotype (Figure 5-4).

AtSHN3-overexpressing plants have similar lignin and cellulose content to wild-type plants

Weisner staining of I2 sections of transgenic and wild-type switchgrass stems suggested that overexpression of *AtSHN3* in switchgrass alters lignin content. Therefore, I2 stem fragments were subjected to sulfuric acid hydrolysis to determine if the amount of acid soluble and acid insoluble material, as well as overall lignin content, differed in the stems of transgenic and wild-type plants. While variation could be observed between the transgenic switchgrass lines and the control plants, acid soluble and acid insoluble lignin contents were not significantly altered between the two ($p > 0.05$) (Table 5-2). Since the individual lignin composition measurements were not statistically different, it was unsurprising to find that the overall total lignin content

between transgenic and wild-type switchgrass plants was also not significantly different (Table 5-2).

The cellulosic glucose content of I2 stem fragments was measured to determine if the transgenic lines had an increase in cellulose. Significant variation between the transgenic lines and the control plants was observed with SHN3 7-2 plants as they had approximately 10% more cellulose than the HR8 plants ($p = 0.03$) (Figure 5-5). For the individual hemicellulosic components, SHN3 4-1 plants were found to have 12% less arabinose ($p = 0.02$), 50% less galactose ($p = 0.03$), and 21% less D-glucuronate ($p = 0.03$) than the wild-type control (Figure 5-6). Additionally, SHN3 5-2 transgenic plants had 43% less glucose ($p = 0.01$), 64% less galacturonic acid ($p = 0.01$), and 26% less D-glucuronate ($p = 0.01$) than the wild-type HR8. The SHN3 6-3 plants differed from the HR8 control in regards to glucose (33% less, $p = 0.03$), xylose (14% more, $p = 0.01$), and galacturonic acid (50% less, $p = 0.03$). Finally, SHN3 7-2 transgenic plants were found to have 31% more arabinose ($p = 0.005$), 90% more xylose ($p = 0.001$), and 20% less D-glucuronate ($p = 0.02$) than HR8 control plants (Figure 5-6).

Overexpression of AtSHN3 in switchgrass does not alter the stiffness or mechanical strength of switchgrass stems

A change in the lignocellulosic composition of the plant cell wall could alter the strength of the stem, which helps the plant maintain an upright growth habit and to withstand abiotic stress such as wind. Storage modulus tests were conducted to measure if any of the potential changes in cell wall composition affected the stiffness of the *AtSHN3*-overexpressing stems. This was performed by applying an oscillating stress to the sample and measuring the responding strength.

The results of this analysis found that there were variations between the transgenic switchgrass lines and the wild-type control; however, these variations were not significant at the 0.05 level (Table 5-3).

In addition to the stiffness, we wanted to test if changing the composition of the cell wall affected the overall mechanical strength of the transgenic switchgrass stems. To accomplish this, fracture tests were performed on switchgrass stem sections in which continuously increasing levels of torsion force were applied to the stem sections until the stems broke. From the fracture tests, two parameters were observed that correlated to the overall mechanical strength of the stem: 1) the slope of the linear region, which reflects the stiffness of the stem and 2) the breaking point, which correlates to the strength of the stem. The results from both the linear region and breaking point analyses found that there was biological variation in the stiffness and strength of transgenic switchgrass stems in comparison to the wild type control (Table 5-4); however, these results were not significantly different ($p > 0.05$).

AtSHN3- overexpressing switchgrass plants are more susceptible than wild-type plants to switchgrass rust

The plant cell wall is the first physical barrier encountered by plant pathogens upon initiation of infection. Since the *AtSHN3*-overexpressing plants had some biological variation in their lignin and cellulose contents, we wanted to investigate whether or not this would affect disease resistance of switchgrass to switchgrass rust. After inoculating both the transgenic lines and the wild-type control with *Puccinia emaculata* urediniospores, we found that the *AtSHN3*-overexpressing plants were more susceptible to disease than the HR8 control plants (Figure 5-7).

In addition, the SHN3 4-1 ($p = 0.008$), SHN3 5-2 ($p = 0.002$), and the SHN3 6-3 ($p = 0.04$) lines were significantly different from the wild type.

DISCUSSION

Switchgrass is a promising bioenergy crop and switchgrass cultivars that contain reduced amounts of lignin and increased levels of cellulose are desirable for cost-effective and efficient bioethanol production. Coordinated activation and repression of these two cell wall components has been shown to be regulated by a gene belonging to the SHINE (SHN) family of AP2 transcription factors (Ambavaram *et al.*, 2011). There are three members of the SHN family that vary in their developmental and tissue-specific gene expression patterns (Aharoni *et al.*, 2004). Despite their proven role in wax accumulation and cell wall formation (Aharoni *et al.*, 2004; Shi *et al.*, 2013), the function of these genes has not yet been explored in switchgrass.

In this study, the *SHINE3* (*SHN3*) cDNA sequence from Arabidopsis was cloned and transformed into switchgrass. While other reports have reported a glossy phenotype of the leaf surface upon over-expression of *SHN* genes (Aharoni *et al.*, 2004; Buxdorf *et al.*, 2014), this phenotype was not observed in any of the transgenic switchgrass plants created in this study. *AtSHN3*-overexpressing switchgrass plants, however, exhibited stunted growth in comparison to wild-type plants (Figure 5-2). Transgenic tomato plants over-expressing *SlSHN3*, the tomato ortholog of *AtSHN3*, also displayed stunted growth (Buxdorf *et al.*, 2014). Interestingly, the stunted growth phenotype of tomato plants was more severe in *SlSHN3*-overexpressing plants than in *SlSHN1*-overexpressing plants (Al-Abdallat *et al.*, 2014). This suggests that while the

SHN proteins may have similar functions, their tissue-specific expression patterns are essential for proper cell wall development.

The *SHN* genes have been shown to regulate wax deposition on both leaf and fruit cuticle surfaces (Aharoni *et al.*, 2004; Shi *et al.*, 2013). In addition, members of the SHN family function in cell elongation and in the thickening of secondary cell walls (Ambavaram *et al.*, 2011; Shi *et al.*, 2013). Ambavaram *et al* (2011) found that *AtSHN2*-overexpression resulted in a 45% decrease in lignin and a 34% increase in cellulose in transgenic rice plants. No significant changes in lignin content were found for the SHN3 transgenic switchgrass plants in this study. Biological variation of lignin content could be observed, especially with the SHN3 4-1 and SHN3 7-2 switchgrass lines. These two lines also exhibited the highest levels of transgene expression (Figure 5-3) and appeared to have slightly less lignin than the control plants (Table 5-2). Similar to the results in rice, we did find one line, SHN3 7-2, that had a 10% increase in cellulose content in comparison to the wild-type plants (Figure 5-5, Table 5-2). The results of calcufluor staining for cellulose showed that SHN3 7-2 stem sections were substantially brighter than wild type stems (Figure 5-4), supporting that this line has increased cellulose content. A change in the hemicellulose composition was also observed for the *AtSHN3*-overexpressing plants (Figure 5-6). Further studies are needed to determine if the changes in lignin and cellulose in these lines results in enhanced bioethanol production.

Changes in the lignin and cellulose components of the plant cell wall could compromise the plant's ability to withstand extracellular forces associated with abiotic stress such as wind and rain. For instance, *brittle stalk* (*bk2*) mutants of maize contain 19% less cellulose and 3% more lignin than wild-type maize plants (Ching *et al.*, 2006); however, the mechanical strength of the stem is significantly compromised and easily broken with minimal applied force (Ching *et*

al., 2006). In this study, storage modulus and fracture tests were performed on transgenic and wild-type switchgrass lines to assess the stem stiffness and mechanical strength, respectively. These tests were recently developed and optimized for use in plant biology research (Wan et al, data not published). Storage modulus tests can evaluate the stiffness of the stem by analyzing the % strain output as it correlates to a specific stress. Fracture tests utilize tensile-torsion force to apply stress to a sample and the % strain is measured during the linear region and at the sample breaking point. While there was biological variation among all samples in regard to storage modules and fracture breaking points, the results were not statistically significant from the wild-type. Ambavaram *et al* (2011) also found that the mechanical strength of *AtSHN2*-overexpressing transgenic rice lines was not compromised.

Over-expression of *SHN1* genes in transgenic Arabidopsis, tomato, and rice plants conferred a level of drought tolerance that was higher than wild-type plants (Aharoni *et al.*, 2004; Ambavaram *et al.*, 2011; Al-Abdallat *et al.*, 2014). This could be attributed to the accumulation of excess epicuticular waxes on the leaf surface, which contributes to the glossy leaf phenotype, or to the reduced numbers of stomata in the transgenic plants (Al-Abdallat *et al.*, 2014). While this study did not analyze if *AtSHN3* overexpression altered the number of stomata or improved the drought tolerance of transgenic switchgrass plants, the water use efficiency of the transgenic lines did appear to be different from the control plants (data not shown). More studies will need to be conducted to determine if *AtSHN3* enhances drought tolerance in switchgrass.

The first physical barrier that foliar pathogens encounter is the plant cell wall. Thus, we investigated if *AtSHN3* over-expression affected the disease resistance response of switchgrass to switchgrass rust. Three transgenic switchgrass lines, SHN3 4-1, SHN3 5-2, and SHN3 6-3 plants

were found to be significantly more susceptible to the rust pathogen than the wild type control (Figure 5-5). SHN3 7-2 plants were also susceptible to switchgrass rust but not significantly different than wild type HR8 plants. Buxdorf et al (2014) applied toluidine blue stain to the leaves of *Slshn3*-RNAi and *SlSHN3*-overexpressing tomato mutant plants to determine if changes in the expression of this gene affected the permeability of the tomato leaf cuticle. These authors found that the overexpression of *SlSHN3* in tomato leaves allowed the leaves to uptake toluidine blue, suggesting that *SlSHN3*-overexpressing plants contained a more permeable cuticle than the wild type (Buxdorf *et al.*, 2014). Thus, a more permeable cuticle could explain the disease susceptibility phenotype of the *AtSHN3* switchgrass transgenic plants. In addition, changes in the cellulose content and/or the individual sugar monomers present in hemicellulose could lend the *AtSHN3*-overexpressing transgenic plants more susceptible to disease. Interestingly, the *SlSHN3*-overexpression tomato plants appeared to be more resistant to *Botrytis cinerea* and *Xanthomonas campestris* pv. *vesicatoria* infection. Finally, the SHN proteins are transcription factors that may regulate genes other than those involved in wax deposition and cell wall composition. It will be interesting to determine if the aberrant expression of other genes caused by SHN3 overexpression is responsible for promoting susceptibility of the AtSHN3 transgenic switchgrass lines to switchgrass rust.

CONCLUSION

In the future, switchgrass cultivars with cell wall compositions that have enhanced levels of cellulose and decreased levels of lignin will be desired for two reasons: 1) to reduce the cost associated with expensive microbial enzymes that are used to break down the lignocellulosic

biomass and 2) to increase the amount of bioethanol derived from the lignocellulosic material. To our knowledge, this is the first report that has evaluated the effect that the *AtSHN3* transcription factor has on cell wall composition and disease response in switchgrass. Lignin and cellulose measurements for *AtSHN3*-overexpressing transgenic lines were biologically different; however, only a 10% increase in the amount of cellulose for SHN3 7-2 plants was found to be significantly different than the wild type. Further analysis is needed to confirm if SHN3 overexpression in switchgrass does in fact change cell wall composition. In addition, the stunted growth phenotype that is associated with *AtSHN3*-overexpression does not compromise the mechanical strength of the transgenic switchgrass stems. More importantly, SHN3 transgenic switchgrass plants are more susceptible to switchgrass rust than wild type plants. The results of this study provide insight into the role that SHN3 has in switchgrass growth and development. Furthermore, these results suggest that transgenic switchgrass lines with altered cell wall content may be compromised in their resistance to switchgrass rust. The information obtained from this study will help in the creation of switchgrass cultivars with improved feedstock quality and disease resistance.

ACKNOWLEDGEMENTS

This project was supported by the Institute for Critical Technology and Applied Science at Virginia Tech, the Biodesign and Bioprocessing Research Center of the College of Agriculture and Life Sciences at Virginia Tech, and the Virginia Agricultural Experimental Station (VA135872).

AUTHOR CONTRIBUTIONS

TF, CF, and BZ designed the study. TF performed the *Agrobacterium* transformation of switchgrass, agronomic trait measurement, stem sectioning, Weisner and calcufluor staining, qPCR, lignin content measurement, and disease assays. CT helped with agronomic trait measurement. GW and AN performed the solvent-submersion tensile-torsion analysis. FL and LB measured the hemicellulose and cellulose content of the transgenic and wild type switchgrass lines. TF wrote the manuscript.

REFERENCES

Aharoni A, Dixit S, Jetter R, Thoenes E, van Arkel G, Pereira A. 2004. The SHINE clade of AP2 domain transcription factors activates wax biosynthesis, alters cuticle properties, and confers drought tolerance when overexpressed in Arabidopsis. *Plant Cell* **16**: 2463-2480.

Al-Abdallat AM, Al-Debei HS, Ayad JY, Hasan S. 2014. Over-expression of *SlSHN1* gene improves drought tolerance by increasing cuticular wax accumulation in tomato. *Int. J. Mol. Sci.* **15**: 19499-19515.

Ambavaram MMR, Krishnan A, Trijatmiko KR, Pereira A. 2011. Coordinated activation of cellulose and repression of lignin biosynthesis pathways in rice. *Plant Physiol.* **155**: 916-931.

Bartley LE, Peck ML, Kim SR, Ebert B, Manisseri C, Chiniquy DM, Sykes R, Gao L, Rautengarten C, Vega-Sanchez ME, Benke PI, Canlas PE, Cao P, Brewer S, Lin F, Smith WL, Zhang X, Keasling JD, Jentoff RE, Foster SB, Zhou J, Ziebell A, An G, Scheller HV, Ronald PC. 2013. Overexpression of a BAHD acyltransferase, *OsAt10*, alters rice cell wall hydroxycinnamic acid content and saccharification. *Plant Physiol.* **161**: 1615-1633.

Bouton JH. 2007. Molecular breeding of switchgrass for use as a biofuel crop. *Curr. Opin. Genet. Dev.* **17**: 553-558.

Buxdorf K, Rubinsky G, Barda O, Burdman S, Aharoni A, Levy M. 2014. The transcription factor *SlSHINE3* modulates defense responses in tomato plants. *Plant Mol. Biol.* **84**: 37-47.

Ching A, Dhugga K, Appenzeller L, Meeley R, Bourett T, Howard R, Rafalski A. 2006. *Brittle stalk 2* encodes a putative glycosylphosphatidylinositol-anchored protein that affects mechanical strength of maize tissues by altering the composition and structure of secondary cell walls. *Planta* **224**: 1174-1184.

Fu C, Mielenz JR, Xiao X, Ge Y, Hamilton CY, Rodriguez M, Chen F, Foston M, Ragauskas A, Bouton J, Dixon RA, Wang Z-Y. 2011a. Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proc. Natl. Acad. Sci. USA* **108**: 3803-3808.

Fu C, Xiao X, Xi Y, Ge Y, Chen F, Bouton J, Dixon RA, Wang Z-Y. 2011b. Downregulation of cinnamyl alcohol dehydrogenase (CAD) leads to improved saccharification efficiency in switchgrass. *BioEnergy Res.* **4**: 153-164.

Hopkins AA, Vogel KP, Moore KJ. 1993. Predicted and realized gains from selection for in vitro dry matter digestibility and forage yield in switchgrass. *Crop Sci.* **33**: 253-258.

Jiang F, Guo M, Yang F, Duncan K, Jackson D, Rafalski A, Wang S, Li B. 2012. Mutations in an AP2 transcription factor-like gene affect internode length and leaf shape in maize. *PLoS ONE* **7**: e37040.

Jørgensen H, Kristensen JB, Felby C. 2007. Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. *Biofuels, Bioprod. Biorefin.* **1**: 119-134.

Li R, Qu R. 2011. High throughput *Agrobacterium*-mediated switchgrass transformation. *Biomass Bioenergy* **35**: 1046-1054.

McLaughlin S, Bouton J, Bransby D, Conger B, Ocumpaugh W, Parrish D, Taliaferro C, Vogel K, Wullschleger S 1999. Developing switchgrass as a bioenergy crop. In: Janick J ed. *Perspectives on new crops and new uses*. Alexandria, VA: ASHS Press.

McLaughlin SB, Kszos LA. 2005. Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. *Biomass Bioenergy* **28**: 515-535.

Naik SN, Goud VV, Rout PK, Dalai AK. 2010. Production of first and second generation biofuels: A comprehensive review. *Renew. Sust. Energ. Rev.* **14**: 578-597.

Perlack RD, Wright LL, Turhollow AF, Graham RL, Stokes BJ, Erbach DC. 2005. Biomass as feedstock for a bioenergy and bioproducts industry: the technical feasibility of a billion-ton annual supply: DTIC Document.

Porebski S, Bailey LG, Baum B. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* **15**: 8-15.

Porter CL, Jr. 1966. An analysis of variation between upland and lowland switchgrass, *Panicum Virgatum* L., in central Oklahoma. *Ecology* **47**: 980-992.

Pradhan Mitra P, Loque D. 2014. Histochemical staining of *Arabidopsis thaliana* secondary cell wall elements. *J. Vis. Exp.*

Saathoff AJ, Sarath G, Chow EK, Dien BS, Tobias CM. 2011. Downregulation of cinnamyl-alcohol dehydrogenase in switchgrass by RNA silencing results in enhanced glucose release after cellulase treatment. *PloS ONE* **6**: e16416.

Sanderson MA, Adler PR. 2008. Perennial forages as second generation bioenergy crops. *Int. J. Mol. Sci.* **9**: 768-788.

Sanderson MA, Adler PR, Boateng AA, Casler MD, Sarath G. 2006. Switchgrass as a biofuels feedstock in the USA. *Can. J. Plant Sci.* **86**: 1315-1325.

Sanderson MA, Reed RL, McLaughlin SB, Wullschleger SD, Conger BV, Parrish DJ, Wolf DD, Taliaferro C, Hopkins AA, Ocumpaugh WR, Hussey MA, Read JC, Tischler CR. 1996. Switchgrass as a sustainable bioenergy crop. *Bioresour. Technol.* **56**: 83-93.

Shen H, Poovaiah CR, Ziebell A, Tschaplinski TJ, Pattathil S, Gjersing E, Engle NL, Katahira R, Pu Y, Sykes R. 2013. Enhanced characteristics of genetically modified switchgrass (*Panicum virgatum* L.) for high biofuel production. *Biotechnol. Biofuels* **6**: 71.

Shi JX, Adato A, Alkan N, He Y, Lashbrooke J, Matas AJ, Meir S, Malitsky S, Isaacson T, Prusky D, Leshkowitz D, Schreiber L, Granell AR, Widemann E, Grausem B, Pinot F, Rose JK, Rogachev I, Rothan C, Aharoni A. 2013. The tomato *SlSHINE3* transcription factor regulates fruit cuticle formation and epidermal patterning. *New Phytol.* **197**: 468-480.

Sluiter A, Ruiz R, Scarlata C, Sluiter J, Templeton D. 2005. Determination of extractives in biomass. Laboratory Analytical Procedure (LAP). Golden, CO: National Renewable Energy Laboratory.

Sun Y, Cheng J. 2002. Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresour. Technol.* **83**: 1-11.

Taliaferro CM, Das M. 2002. Breeding and selection of new switchgrass varieties for increased biomass production. Oak Ridge, Tennessee: Oak Ridge National Laboratory.

Wullschleger SD, Davis EB, Borsuk ME, Gunderson CA, Lynd LR. 2010. Biomass production in switchgrass across the United States: database description and determinants of yield. *Agron. J.* **102**: 1158-1168.

Xu B, Escamilla-Trevino LL, Sathitsuksanoh N, Shen Z, Shen H, Zhang YH, Dixon RA, Zhao B. 2011a. Silencing of 4-coumarate:coenzyme A ligase in switchgrass leads to reduced lignin content and improved fermentable sugar yields for biofuel production. *New Phytol.* **192**: 611-625.

Xu B, Huang L, Shen Z, Welbaum GE, Zhang X, Zhao B. 2011b. Selection and characterization of a new switchgrass (*Panicum virgatum* L.) line with high somatic embryogenic capacity for genetic transformation. *Sci. Hortic.* **129**: 854-861.

Xu B, Sathitsuksanoh N, Tang Y, Udvardi MK, Zhang J-Y, Shen Z, Balota M, Harich K, Zhang P, Zhao B. 2012. Overexpression of *AtLOV1* in switchgrass alters plant architecture, lignin content, and flowering time. *PLoS ONE* **7**: e47399.

Table 5-1. Comparison of agronomic trait measurements for 3 month old greenhouse grown SHN3 transgenic switchgrass and HR8 control plants.

Total number of observations for tiller number, tiller height, flag leaf length, flag leaf width, and stem width was n = 12 (four observations per three biological replicates). Total number of observations for biomass was n = 3 for each line.

T_0 plants	Tiller number	Tiller height (cm)	Flag leaf length (cm)	Flag leaf width (mm)	Stem width (mm)	Biomass (kg)[a]
HR8	12.0±3.0	126.0±4.8	33.3±4.3	9.3±0.5	4.1±0.2	0.096±0.02
SHN3 4-1	10.5±1.9	85.3±6.2**	25.3±1.2*	7.2±0.4**	3.3±0.3*	0.069±0.01*
SHN3 5-2	6.8±2.5*	126.5±7.6	32.2±4.2	10.1±0.6	4.4±0.2	0.087±0.03
SHN3 6-3	9.2±2.5	104.6±9.1*	23.5±4.5	8.0±0.8	3.9±0.2	0.066±0.01*
SHN3 7-2	10.6±2.7	75.5±0.7**	24.8±2.2*	8.0±0.5*	3.6±0.4	0.073±0.02

[a]= biomass of plant fresh weight
*= statistically different at $p < 0.05$
**= statistically different at $p < 0.01$

Table 5-2. Acid soluble lignin and acid insoluble lignin measurement of greenhouse grown *AtSHN3*-overexpressing transgenic plants and the wild-type control.

Lignin content of I2 stem segments was measured for R3-stage greenhouse grown switchgrass tillers. n = 3 for all samples except for SHN3 4-1, where n =2 for Acid Insoluble Lignin.

Switchgrass line	% Acid soluble lignin	% Acid insoluble lignin	% Total lignin
HR8	15.52 ±1.05	2.19 ± 0.15	17.71±1.04
SHN3 4-1	13.12 ±0.01	2.08 ± 0.24	15.20±0.35
SHN3 5-2	15.84±1.05	2.47 ± 0.13	18.32±1.05
SHN3 6-3	15.11± 0.32	2.11 ± 0.39	17.22±0.71
SHN3 7-2	14.24 ± 0.71	2.12 ± 0.22	16.36±0.86

Table 5-3. Average storage modulus derived from stress sweeps at 25ºC for *AtSHN3*-overexpressing transgenic plants and HR8 wild-type control.

The number of repetitions for this experiment is n=2 for all biological samples.

Storage Modulus G' (Pa)	**HR8**	**SHN3 4-1**	**SHN3 5-2**	**SHN3 6-3**	**SHN3 7-2**
Strain	2.12E8	2.18E8	2.19E8	1.50E8	1.81E8
Standard Deviation	1.84E7	1.10E8	2.90E7	1.48E7	2.55E7
p-value		0.06	0.89	0.85	0.34

Table 5-4. Initial linear strength measurement and shear stress at breaking point for the *AtSHN3*-overexpressing transgenic lines and the HR8 control.

The number of repetitions used for this analysis was n=5.

Measurement	**HR8**	**SHN3 4-1**	**SHN3 5-2**	**SHN3 6-3**	**SHN3 7-2**
Initial Linear Strength (Pa)	1.98E6	1.65E6	1.25E6	1.70E6	1.58E6
Standard Deviation	9.30E5	3.10E5	5.61E5	7.65E5	3.92E5
p-value		0.53	0.24	0.65	0.47
Shear stress at breaking point (Pa)	1.93E7	1.61E7	1.70E7	1.44E7	1.75E7
Standard Deviation	5.67E6	2.48E6	3.48E6	4.32E6	8.07E5
p-value		0.39	0.57	0.27	0.61

Figure 5-1. Southern blot confirmation of transgene insertion.

A portion of the hygromycin selection gene was used as a probe. 1 = HR8 negative control, 2 = 1 kb positive standard, 3 = SHN3 4-1, 4 = SHN3 5-2, 5 = SHN3 6-3, 6 = SHN3 7-2.

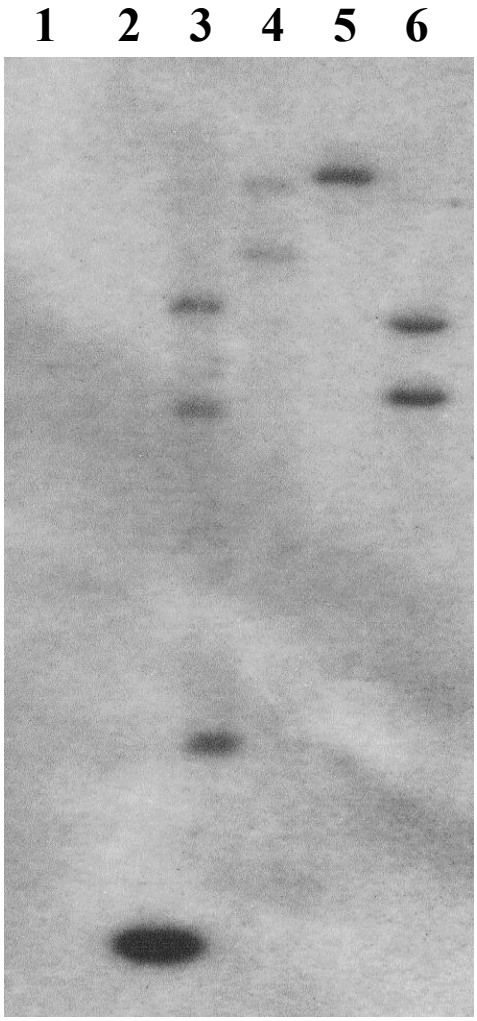


Figure 5-2. *AtSHN3*-overexpressing transgenic switchgrass lines are smaller than wild-type plants.

Listed below are a) HR8 control plant (left) in comparison to SHN3 5-2 (middle) and SHN3 6-3 (right), b) HR8 control plant (left) in comparison to SHN3 4-1 (middle) and SHN3 7-2 (right).

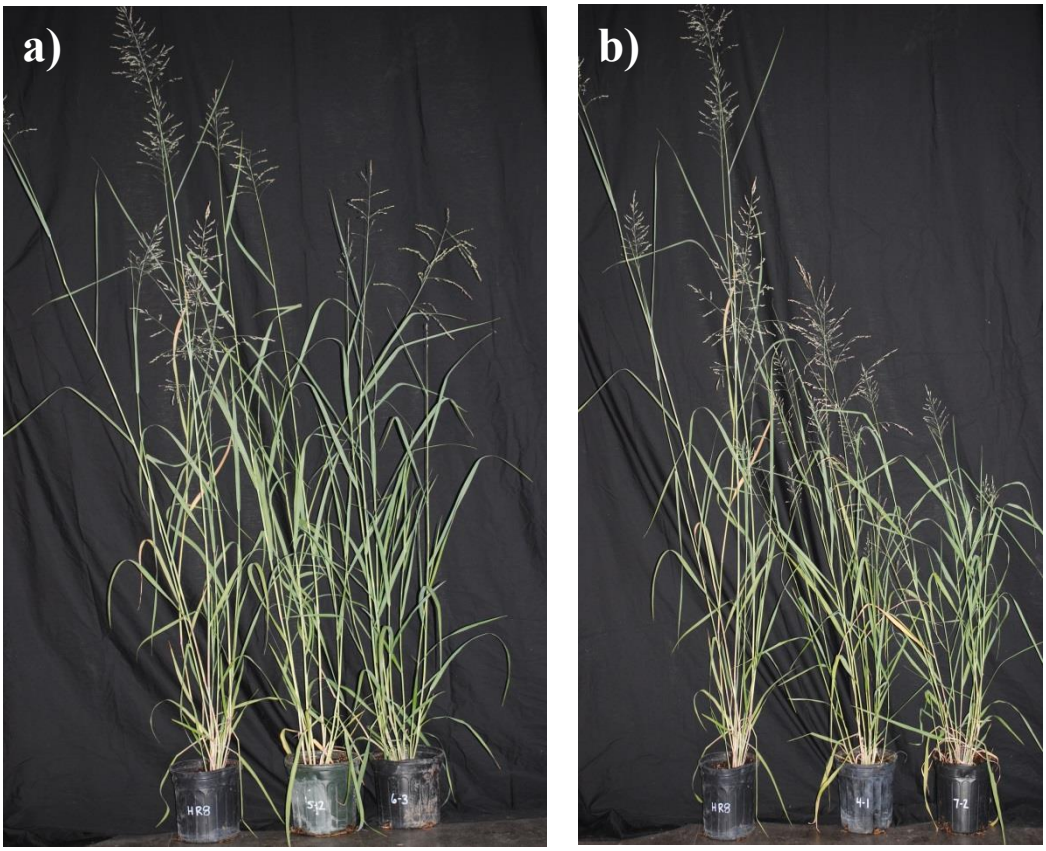


Figure 5-3. qPCR analysis of transgene expression levels in SHN3 transgenic plants.

Expression levels were normalized to the values obtained for SHN3 6-3, which contains one transgene insertion. Three technical replicates of two biological replicates were used for each transgenic line.

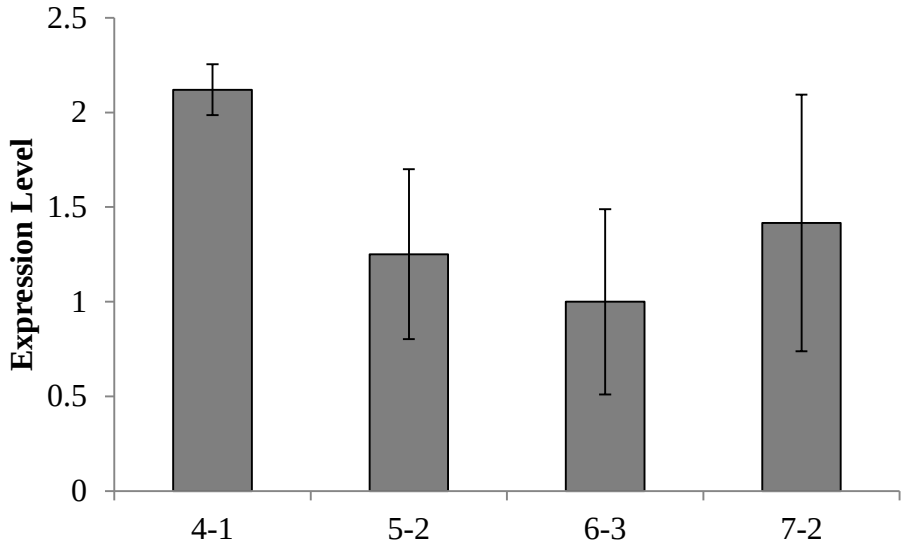


Figure 5-4. Weisner and calcufluor staining of I2 stem sections of four switchgrass transgenic plants overexpressing *AtSHN3* compared to the HR8 control.

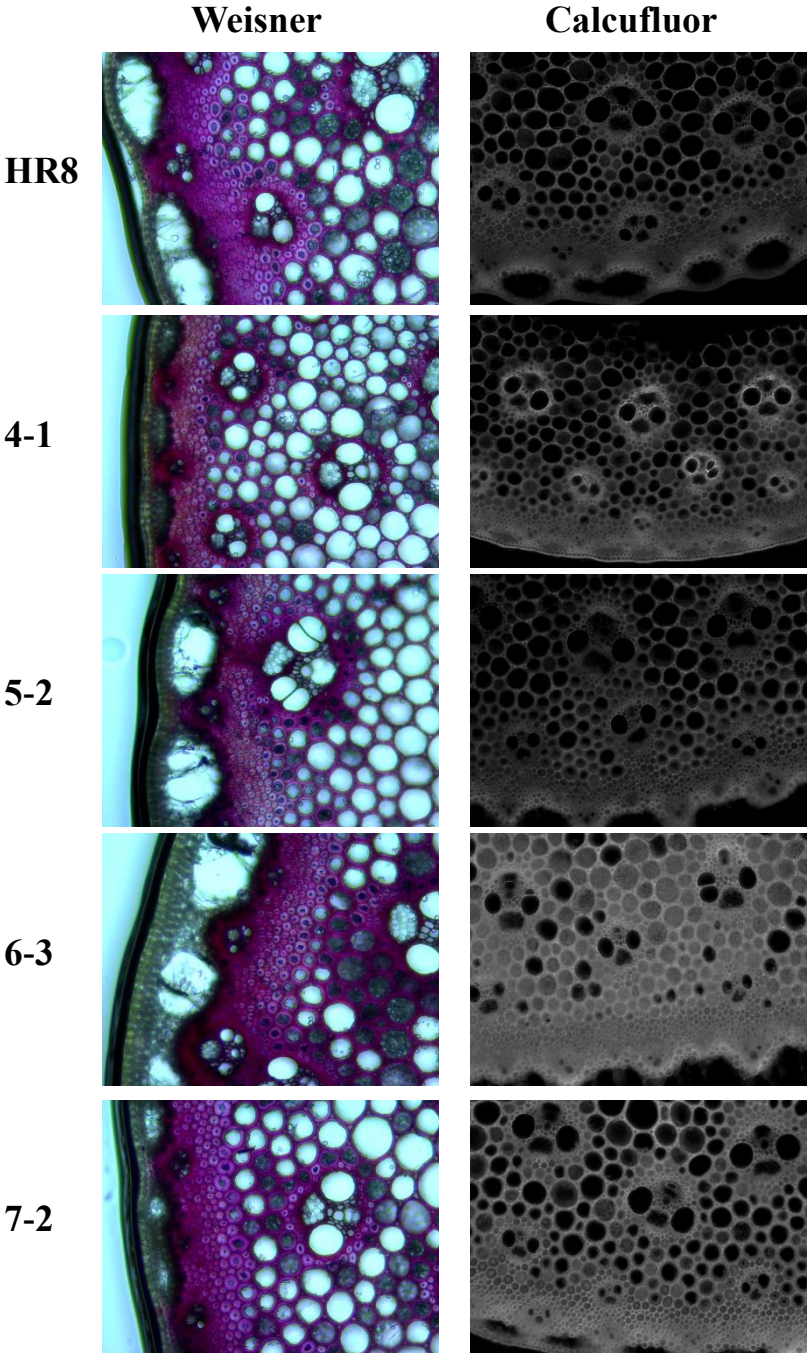


Figure 5-5. Measurement of cellulose content between *AtSHN3*-overexpressing transgenic switchgrass and HR8 wild-type.

Cellulose content of I2 stem segments was measured for R3-stage greenhouse grown switchgrass tillers. n = 3 for all samples. SHN3 7-2 contained significantly more cellulose than the HR8 control (p = 0.03).

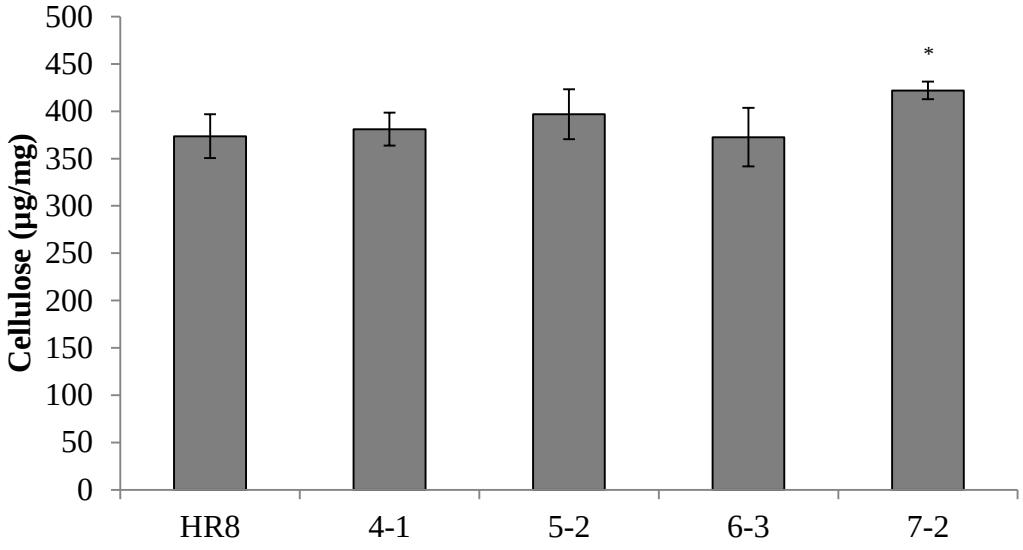


Figure 5-6. Measurement of hemicellulose sugars between *AtSHN3*-overexpressing transgenic switchgrass and HR8 wild-type.

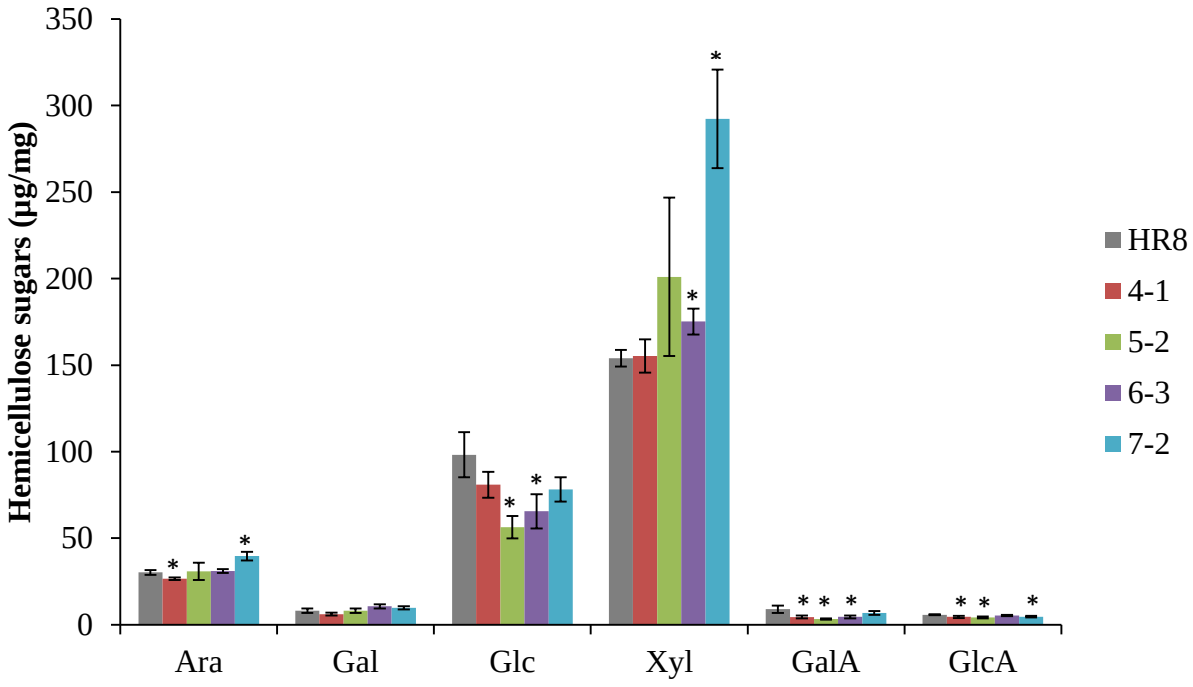
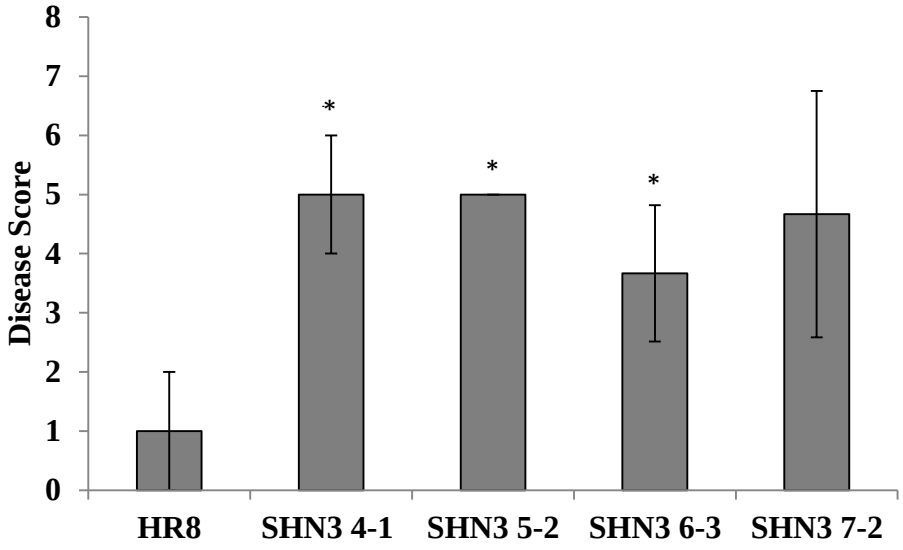


Figure 5-7. Disease assays of AtSHN3-overexpressing transgenic switchgrass plants and wild-type control.

* indicates lines that are significantly different from the wild-type at $p < 0.05$.



CHAPTER 6: Summary and Future Work

In this dissertation, I used bioinformatics and molecular biology tools to accomplish the objectives set forth in Chapter 1, and to comprehensively analyze rust disease resistance in the bioenergy plant switchgrass. Listed below are summaries of the major findings for each objective, as well as directions for future works based off the results of this dissertation.

Objective 1: Explore differences in the transcriptomes of the switchgrass cultivars 'Dacotah' and 'Alamo' (Chapter 2).

Transcriptome analysis was performed for 'Alamo' and 'Dacotah', two switchgrass cultivars that exhibit vastly different phenotypes and disease resistance responses. Since switchgrass is a readily out-crossing species (Martinez-Reyna & Vogel, 2002), there is a high degree of heterozygosity present in the genetic background of existing switchgrass cultivars (Narasimhamoorthy *et al.*, 2008). The results of this study helped determine the level of RNA-sequencing that is necessary in order to fully capture the extent of polymorphism present within switchgrass. In addition, the results of this chapter provided insight into gene expression of these two cultivars. Additional work on this project should focus on dissecting the genes and/or gene networks controlling important agronomic traits such as biomass production, flowering time, and abiotic stress resistance. Moreover, characterizing the SNPs that were identified between the two cultivars in regards to parentage and cultivar relatedness. A deeper understanding of recombination in switchgrass and the mating between cultivars will be essential for successful switchgrass breeding programs.

Objective 2: Identify and characterize switchgrass NB-LRR containing plant disease resistance (*R*) genes (Chapter 3).

Next, a homology-based computational approach was employed to identify NB-LRR disease resistance genes in the switchgrass genome (v 1.1). A total of 1,011 potential NB-LRR disease resistance genes were identified, which is approximately 266% more than a previous study in switchgrass (Zhu *et al.*, 2013). Several of these were found to contain unique domains including calmodulin-like binding, thioredoxin, jacalin-like lectin, and major sperm protein. More studies are needed to determine if these domains are in fact attached to NB-LRR proteins in switchgrass and if so, what their function is in disease response. The SNPs identified in this study could also be developed into molecular markers for use in switchgrass breeding programs. Additionally, this research is the first to analyze developmental regulation of NB-LRR genes on a whole genome level. I found that NB-LRR disease resistances genes are developmentally regulated during switchgrass growth and development. Future work should consider identifying and isolating NB-LRR genes that are highly expressed at times that correlate to optimum rust infection. NB-LRR genes that have been shown to play a role in switchgrass disease resistance can then be inserted into elite switchgrass germplasm, through either biotechnology or traditional breeding methods. Since NB-LRR *R* genes are often times specific to different pathogens and even haplotypes of certain pathogens, pyramiding of NB-LRR genes in commercially important switchgrass cultivars would offer broad spectrum and durable disease resistance.

Objective 3. Evaluate a collection of switchgrass germplasm in response to rust isolate VT2-1 (Chapter 4).

RNA-sequencing was used to identify SNPs contained within three novel pseudo-F_2 mapping populations. A total of 966 molecular markers were used to generate a genetic linkage map that could be used to identify QTLs correlating with disease resistance. Indeed, an area on Chr02B around marker 92 cM was identified as a potential genetic region of interest. Further investigation revealed that this marker is located in a casein kinase II gene, which may function in signal transduction pathways. Future work on this project should identify what other genes are in this region and elucidate their role in switchgrass disease response. In addition, deeper RNA-sequencing of the pseudo-F_2 populations, along with refinement of the switchgrass reference genome, will help create more high-quality SNP markers that can increase the resolution of this map. Moreover, this project used a single pustule isolate of switchgrass (VT2-1) to perform the disease assays. Thus, the QTL identified in this chapter may be specific to that particular haplotype of switchgrass rust. Disease assays with other switchgrass rust haplotypes and other pathogens should be performed to determine additional QTLs that may contribute to disease resistance. These QTLs could then be pyramided into elite switchgrass germplasm to offer broad spectrum and durable disease resistance.

Objective 4. Transform switchgrass with *AtSHN3* and determine the effects of this transgene on the cell wall components (Chapter 5).

AtSHN2-overexpressing rice plants were found contain higher levels of cellulose and reduced levels of lignin, two characteristics that are ideal for maximal bioethanol production from energy crops (Ambavaram *et al.*, 2011). Therefore, transgenic switchgrass lines over-expressing *AtSHN3* were created. Three of the transgenic line created in this study, SHN3 4-1, SHN3 5-2, and SHN3 6-3, were stunted in their growth in comparison to the wild-type plants.

Two of these transgenic lines, SHN3 4-1 and SHN3 7-2, exhibited a slight change in lignin cell wall composition from the wild-type but were not found to be significantly different ($p > 0.05$). One of the lines, SHN3 7-2, was found to have a 10% increase in cellulose in comparison to than the wild-type plants. In addition, there was no significant change in the mechanical strength of the transgenic plants. SHN3 4-1, SHN3 5-2, and SHN3 6-3 transgenic lines were significantly different from the control plants in their response to switchgrass rust ($p < 0.05$). Therefore, slight alterations in the cell wall components of switchgrass have the potential to significantly affect disease resistance to rust. More research is needed to determine the mechanism by which the transgenic lines are more susceptible to disease. SHN3 is a transcription factor and likely regulates additional genes besides those involved in epicuticular wax deposition and cell wall composition. Future work should look at what genes are being regulated by AtSHN3 in switchgrass and determine if the aberrant expression of any of these genes is conferring disease susceptibility. In addition, the native switchgrass SHN genes should be identified and their role in normal switchgrass growth and development should be elucidated.

REFERENCES

Ambavaram MMR, Krishnan A, Trijatmiko KR, Pereira A. 2011. Coordinated activation of cellulose and repression of lignin biosynthesis pathways in rice. *Plant Physiol.* **155**: 916-931.

Martinez-Reyna J, Vogel K. 2002. Incompatibility systems in switchgrass. *Crop Sci.* **42**: 1800-1805.

Narasimhamoorthy B, Saha MC, Swaller T, Bouton JH. 2008. Genetic diversity in switchgrass collections assessed by EST-SSR markers. *BioEnergy Res.* **1**: 136-146.

Zhu Q, Bennetzen JL, Smith SM. 2013. Isolation and diversity analysis of resistance gene homologues from switchgrass. *G3 Genes Genom. Genet.* **3**: 1031-1042.