

Genetic modification of switchgrass (*Panicum
virgatum* L.) for improvement of plant
architecture, biomass productivity and sugar
release efficiency for biofuel

A Dissertation Presented for the
Doctor of Philosophy
Degree

The University of Tennessee, Knoxville

Wegi Aberra Wuddineh
August 2015

Acknowledgements

I am indebted to Dr. C. Neal Stewart, Jr. for giving me this opportunity to work with him, for the financial assistance, his continuous support and guidance throughout my stay. If not for his support and encouragement, I wouldn't really be where I am with my study. I am grateful to my committee members Dr. Zong-Ming Cheng, Dr. Andreas Nebenfuhr and Dr. Hem Bhandari for their guidance and support.

My special gratitude is to Mitra Mazarai for her invaluable support and guidance throughout my studies. I would also like to express my sincere appreciation to Wusheng Liu who was always willing to go a mile further to help me with all my questions. I would like to thank Charleson Poovaiah, Jonathan Willis and all members of the Stewart Lab for all their support and encouragement.

I would like to acknowledge the University of Tennessee Agricultural Experiment Station, Knoxville, and the US DOE Bioenergy Science Center (BESC) for funding.

Special gratitude to my loving wife, Tsion for her patience, support and prayer throughout my study.

Lastly, I would like to give thanks to God for the richness of his grace bestowed upon me.

Abstract

Switchgrass (*Panicum virgatum* L.) is a leading candidate bioenergy crop for sustainable biofuel production. To ensure its economic viability, tremendous improvements in switchgrass biomass productivity and recalcitrance to enzymatic saccharification are needed. Genetic manipulation of lignin biosynthesis by targeting transcriptional regulators of higher level domains of lignin biosynthesis and other complex traits could alter several bioenergy-desirable traits at once. A three-pronged approach was made in the dissertation research to target one plant growth regulator and transcription factors to alter plant architecture and cell wall biosynthesis.

Gibberellin (GA) catabolic enzymes, GA 2-oxidases (GA2oxs), were utilized to alternatively modify the lignin biosynthesis pathway as GA is known to play a role in plant lignification. Constitutive overexpression of switchgrass C₂₀ [C20] GA2ox genes altered plant morphology and modified plant architecture by increasing the number of tillers. Moreover, transgenic plants exhibited reduced lignin especially in leaves accompanied by 15% increase in sugar release (glucose).

The Knotted1 (PvKN1) TF, a putative repressor of lignin biosynthesis genes, was identified and evaluated for improving biomass characteristics of switchgrass for biofuel. Its ectopic overexpression in switchgrass altered the expression of genes in the lignin, cellulose and hemicellulose biosynthesis, and GA signalling pathways. Consequently, transgenic lines displayed altered growth phenotypes particularly at early stages of vegetative development and moderate changes in lignin content accompanied by improved sugar release by up to 16%.

The APETALA2/ ethylene responsive factor (AP2/ERF) TFs are key putative targets for engineering plants not only so they can withstand adverse environmental factors but also confer modified cell wall characteristics. To facilitate this, a total of 207 switchgrass AP2/ERF TFs comprising 3 families (AP2, ERF and related to API3/VP (RAV)) were identified. Sequence analysis for conserved putative motifs and expression pattern analysis delimited key genes for manipulation of switchgrass. To that end, the PvERF001 TF gene was ectopically overexpressed resulting in improved biomass yield and sugar release efficiency.

The transgenic plants and knowledge produced in this research will be used to create new lines of switchgrass with combined novel traits to address needs in biofuel production and sustainable plant cultivation to enable the development of the bioeconomy.

Table of contents

Chapter 1: Introduction	1
1.1 The challenges and prospects of lignocellulosic biofuel.....	1
1.2 Genetic engineering of lignin biosynthesis	3
1.3 Using transcriptional regulators to engineer lignin biosynthesis	4
1.4 Engineering plants for modified plant architecture, stress tolerance and biomass yield	6
References	10
Chapter 2: Identification and overexpression of <i>gibberellin 2-oxidase (GA2ox)</i> in switchgrass (<i>Panicum virgatum</i> L) for improved plant architecture and reduced biomass recalcitrance	18
2.1 Abstract	19
2.2 Introduction.....	20
2.3 Experimental procedures.....	22
2.3.1 Plant materials and growth conditions	22
2.3.2 Transgene candidate identification, vector construction and plant transformation.....	23
2.3.3 RNA extraction and qRT-PCR	24
2.3.4 Phloroglucinol staining	24
2.3.5 Lignin content and composition by py-MBMS	24
2.3.6 Sugar release	24
2.3.7 Data analysis	25
2.4 Results.....	25
2.4.1 <i>In silico</i> analysis of <i>GA2ox</i> gene family	25
2.4.2 Expression patterns of the switchgrass C ₂₀ <i>GA2ox</i> genes	26
2.4.3 Generation of switchgrass transgenic lines overexpressing <i>PvGA2ox5</i> and <i>PvGA2ox9</i>	26
2.4.4 Overexpression of <i>PvGA2ox5</i> on growth phenotypes in switchgrass.....	27
2.4.5 Exogenous application of GA on transgenics	27
2.4.6 Effect of <i>PvGA2ox5</i> overexpression on lignin content and composition.....	28
2.4.7 Effect of <i>PvGA2ox5</i> overexpression on the expression of lignin genes	28
2.4.8 Effect of <i>GA2ox</i> overexpression on the sugar release.....	28
2.5 Discussion	29
2.6 Acknowledgements	32
References	34
Appendix.....	42

Figures and tables	42
Chapter 3: Identification and overexpression of a Knotted1-like transcription factor in switchgrass (<i>Panicum virgatum</i> L.) for lignocellulosic feedstock improvement	67
3.1 Abstract	68
3.2 Introduction.....	70
3.3 Materials and Methods.....	73
3.3.1 Plants and growth conditions	73
3.3.2 Gene identification, vector construction and transgenic plant production.....	73
3.3.3 RNA extraction and qRT-PCR	74
3.3.4 Phloroglucinol staining	75
3.3.5 Determination of lignin content and composition by py-MBMS	75
3.3.6 Determination of sugar release	75
3.3.7 Data analysis	76
3.4 Results.....	76
3.4.1 KNOX family of TFs in switchgrass and the identification of PvKN1	76
3.4.2 Expression patterns of <i>PvKN1</i> variants.....	77
3.4.3 Constitutive overexpression of <i>PvKN1</i> in switchgrass	77
3.4.4 Altered growth phenotypes in transgenic plants	78
3.4.5 Overexpression of <i>PvKN1</i> in switchgrass and subsequent changes in the expression of putative target genes in transgenic plants	79
3.4.6 Changes in lignin content and composition in transgenic switchgrass	79
3.4.7 Effect of overexpression of PvKN1 on sugar release efficiency in switchgrass.....	80
3.5 Discussion	80
3.6 Acknowledgements	85
References.....	86
Appendix.....	96
Figures and tables:	96
Chapter 4: Genome-wide analysis of switchgrass AP2/ERF transcription factor superfamily, and overexpression of <i>PvERF001</i> for improvement of biomass characteristics for biofuel.....	124
4.1 Abstract	125
4.2 Introduction.....	127
4.3 Materials and methods	129
4.3.1 Identification of AP2/ERF gene families in switchgrass genome.....	129

4.3.2 Cluster and protein sequence analysis of AP2/ERF TFs	130
4.3.3 Analysis of gene structure and gene ontology (GO) annotation	130
4.3.5 Vector construction and plant transformation.....	131
4.3.6 Plants and growth conditions	132
4.3.7 RNA extraction and quantitative reverse transcription polymerase chain reaction (qRT-PCR)	132
4.3.8 Determination of leaf water loss	133
4.3.9 Analysis of lignin content and composition.....	133
4.3.10 Determination of sugar release.....	133
4.3.11 Statistical analysis	134
4.4 Results	134
4.4.1 Identification of AP2/ERF TFs in switchgrass genome.....	134
4.4.2 Cluster analysis of switchgrass AP2/ERF TFs.....	135
4.4.3 Characterization of AP2/ERF gene structures and conserved motifs	135
4.4.4 Gene ontology annotation	137
4.4.5 Expression pattern of switchgrass <i>AP2/ERF</i> genes	137
4.4.6 Expression profiles of switchgrass AP2/ERF genes in lignified tissues	138
4.4.7 Overexpression of <i>PvERF001</i> in switchgrass have enhanced plant growth and sugar release efficiency.....	139
4.5 Discussion	141
4.5.1 Significance of AP2/ERF TFs for improvement of bioenergy crops.....	141
4.5.2 Sequence-based classification of putative AP2/ERF TFs in switchgrass	141
4.5.3 <i>In silico</i> predicted gene functions and subcellular localization of AP2/ERF TFs in switchgrass	142
4.5.4 Gene and protein sequence diversity of switchgrass AP2/ERF TFs.....	142
4.5.5 Diverse expression profiles of switchgrass AP2/ERF TFs and functional implications.....	144
4.5.6 Overexpression of <i>PvERF001</i> improved biomass productivity and sugar release efficiency in switchgrass.....	146
4.6 Acknowledgements	148
References.....	149
Appendix.....	160
Figures and tables	160
Chapter 5: Conclusions	208
Vita.....	211

List of Tables

Table 2-1 List of primers used in this study.....	60
Table 2-2 Comparison of the amino acid sequences among switchgrass C ₂₀ PvGA2ox proteins.	62
Table 2-3 Morphology and biomass yields of transgenic lines overexpressing <i>PvGA2ox5</i> and non-transgenic control (WT) plants..	63
Table 2-4 Sugar release by enzymatic hydrolysis in transgenic and non-transgenic control (WT) lines ...	64
Table 2-5 List of the locus names and GenBank accession numbers of sequences used in this study.	65
Table 3-1 List of primers used in this study.....	115
Table 3-2 Percent identity matrix of the switchgrass KNOX transcription factors.	118
Table 3-3 Features of switchgrass KNOX genes compared to the maize <i>ZmKN1</i> gene.....	119
Table 3-4 Morphology and biomass yields of transgenic switchgrass lines overexpressing <i>PvKN1</i> and non-transgenic control (WT) plants	120
Table 3-5 List of the locus names and/or GenBank accession numbers of the sequences used in this study.	121
Table 4-1 List of primers used in this study.....	187
Table 4-2 Characteristic features of switchgrass AP2/ERF Transcription factor gene family members identified in <i>Panicum virgatum</i>	190
Table 4-3 Summary of the AP2/ERF superfamily genes in switchgrass, rice, <i>Populus</i> and <i>Arabidopsis</i>	199
Table 4-4 Chromosomal distribution of AP2/ERF superfamily in switchgrass.....	200
Table 4-5 Lists of conserved motifs discovered using the MEME Suite version 4.7.0.....	201
Table 4-6 Morphology and biomass yields of transgenic switchgrass lines overexpressing <i>PvERF001</i> and non-transgenic control (WT) plants	205
Table 4-7 Sugar release by enzymatic hydrolysis in transgenic switchgrass overexpressing <i>PvERF001</i> and non-transgenic control (WT) lines.....	206
Table 4-8 List of the locus names and/or GenBank accession numbers of the rice and <i>Arabidopsis</i> sequences used in Figure 4-10.....	207

List of Figures

Figure 2-1 Cluster tree of the C ₁₉ and C ₂₀ GA2ox proteins	43
Figure 2-2 Multiple amino acid sequence alignment of switchgrass C ₂₀ GA2ox proteins and their closest homologs along with two C ₁₉ GA2ox proteins	45
Figure 2-3 Expression patterns of the three putative C ₂₀ GA2ox genes among tissues and developmental stages in switchgrass as determined by qRT-PCR.....	47
Figure 2-4 Molecular characterization of transgenic switchgrass plants overexpressing the <i>PvGA2ox5</i> gene.....	48
Figure 2-5 Representative <i>PvGA2ox5</i> and <i>PvGA2ox9</i> overexpressing transgenic lines showing dwarf phenotypes compared to non-transgenic controls.....	49
Figure 2-6 qRT-PCR analysis of <i>PvGA2ox9</i> overexpressing transgenic and non-transgenic controls (WT) lines from RNA isolated from the top internode of each plant	50
Figure 2-7 qRT-PCR analysis of <i>PvGA2ox5</i> -overexpressing transgenic and non-transgenic lines.....	51
Figure 2-8 Morphology of plants overexpressing <i>PvGA2ox5</i>	52
Figure 2-9 <i>PvGA2ox5</i> overexpressing rice plants showing extremely dwarf phenotypes as compared to the wild type control	53
Figure 2-10 Foliar spray with 100 μ M GA ₃ application restored normal growth in transgenic lines.....	54
Figure 2-11 Histochemical detection of lignin in leaves of <i>PvGA2ox5</i> overexpressing and wild type lines in light microscopy.....	55
Figure 2-12 Lignin content and S/G ratio of transgenic and non-transgenic (WT) lines as determined via py-MBMS.	56
Figure 2-13 Relative expression of lignin biosynthetic genes in transgenic and non-transgenic lines as determined by qRT-PCR.....	57
Figure 2-14 The relative expression of lignin biosynthetic genes in transgenic dwarf, semi-dwarf and non-transgenic (WT) lines as determined by qRT-PCR.....	58
Figure 2-15 Cluster analysis of putative GA2ox genes from monocots	59
Figure 3-1 Cluster analysis of KNOX family of transcription factors (TFs) for the deduced amino acid sequences from switchgrass along with already-characterized proteins from both monocots and dicots. .	97
Figure 3-2 Multiple amino acid sequence alignment of the C-termini of class I and class II family of KNOX TFs.....	98

Figure 3-3 Gene structures of class I and II KNOX transcription factor family coding genes from <i>Arabidopsis</i> , switchgrass (<i>Panicum virgatum</i>) and maize (<i>Zea mays</i>).	100
Figure 3-4 Expression patterns of <i>PvKN1a</i> and <i>PvKN1b</i> in different plant tissues as determined by qRT-PCR.	101
Figure 3-5 Molecular characterization of transgenic switchgrass plants overexpressing the <i>PvKN1a</i> gene	102
Figure 3-6 Comparison of morphological phenotypes in transgenic switchgrass overexpressing <i>PvKN1a</i> and <i>PvKN1b</i>	103
Figure 3-7 Shoot development and leaf morphology of transgenic <i>PvKN1a</i> - overexpressing switchgrass plants	104
Figure 3-8 Relative transcript levels of the <i>PvKN1a</i> in transgenic and non-transgenic (WT) switchgrass plants	105
Figure 3-9 Transgenic SA37 (switchgrass clonal line) plants showing abnormal phenotypes throughout various stages of development	106
Figure 3-10 Overexpression of <i>PvKN1</i> in rice altered shoot but not root development.....	107
Figure 3-11 The relative expression of putative target genes of <i>PvKN1</i> in transgenic switchgrass vs the non-transgenic (WT) plants as determined by qRT-PCR	108
Figure 3-12 Histochemical detection of lignin in leaves of <i>PvKN1</i> overexpressing and non-transgenic control lines in light microscopy	110
Figure 3-13 Lignin content and S/G ratio of transgenic and non-transgenic switchgrass lines as determined via py-MBMS.	111
Figure 3-14 Sugar release by enzymatic hydrolysis in transgenic and non-transgenic control (WT) lines	112
Figure 3-15 Dendrogram showing the relatedness of the <i>PvKN1</i> target genes (<i>PvLAC1</i> , <i>GA20ox1</i> and cellulose and hemicellulose biosynthetic genes) indicated in Figure 3-11.....	113
Figure 4-1 An unrooted phylogenetic tree of switchgrass AP2/ERF proteins.	161
Figure 4-2 The schematic representation of protein and gene structures of switchgrass ERF subfamily.	162
Figure 4-3 The schematic representation of protein and gene structures of switchgrass DREB subfamily.	165
Figure 4-4 The schematic representation of protein and gene structures of switchgrass AP2 and RAV families.....	167

Figure 4-5 Summary of the gene ontology (GO) annotation as defined by blast2go.	168
Figure 4-6 The expression pattern of putative switchgrass AP2/ERF genes at different stages of seed germination	169
Figure 4-7 The expression pattern of putative switchgrass AP2/ERF genes in roots and shoots during vegetative development.	170
Figure 4-8 The expression pattern of putative switchgrass AP2/ERF genes during reproductive development from floret meristem initiation to mature seed development.	172
Figure 4-9 The expression pattern of putative switchgrass AP2/ERF genes in roots and shoot parts including portions of developing internodes and vascular bundles at stem elongation stage 4 (E4).	174
Figure 4-10 Sequence analysis of group V transcription factors in ERF subfamily.	176
Figure 4-11 Representative plants and transgene and endogenous gene expression level in <i>PvERF001</i> overexpressing and non-transgenic (WT) plants.....	177
Figure 4-12 Molecular characterization of transgenic switchgrass plants overexpressing the <i>PvERF001</i> gene.....	178
Figure 4-13 Correlation between relative transcript levels of <i>PvERF001</i> and growth metrics in switchgrass lines	179
Figure 4-14 Comparison of the relative rate of water loss between the transgenic and non-transgenic control lines.....	180
Figure 4-15 The relative expression of putative wax and cutin biosynthetic genes in <i>PvERF001</i> overexpressing transgenic vs the non-transgenic (WT) plants as determined by qRT-PCR	181
Figure 4-16 The relative expression of putative cell wall biosynthetic genes in <i>PvERF001</i> overexpressing transgenic vs the non-transgenic (WT) plants as determined by qRT-PCR	183
Figure 4-17 Lignin content and S/G ratio of transgenic and non-transgenic (WT) lines as determined via py-MBMS	185
Figure 4-18 Histochemical detection of lignin in leaves of <i>PvERF001</i> overexpressing lines compared to non-transgenic control.	186

List of Abbreviations

4CL	4-coumarate: CoA ligase
ANOVA	Analysis of variance
AP2	Apetala2
BELL	BEL1-like homeodomain
BR	Brassinosteroid
C3H	Coumaroyl shikimate 3-hydroxylase
C4H	Coumaroyl shikimate 4-hydroxylase
CAD	Cinnamyl alcohol dehydrogenase
CCoAOMT	Caffeoyl CoA 3-O-methyltransferase
CCR	Cinnamoyl CoA reductase
CDD	Conserved domain database
CDS	Coding DNA sequence
CER6	Eceriferum 6
CESA	Cellulose synthase
COMT	Caffeic acid O-methyltransferase
CSL	Cellulose synthase-like
CWR	Cell wall residues
CYP86A7-1	Cytochrome P450
DREB	Dehydration responsive element binding
ERF	Ethylene responsive element
F5H	Ferulate 5-hydroxylase
FAE1	Fatty acid elongase1
FDH2	Formate dehydrogenase 2
GA2ox	Gibberellin 2-oxidase
GA20ox	Gibberellin 20-oxidase
GO	Gene ontology

HD	Homeodomain
HTH	HOTHEAD protein
KCS1	3-ketoacyl-CoA synthase 1
KNAT1	KN1-like from <i>Arabidopsis</i>
KN1	Knotted1
KNOX	KN1-like homeobox
LAC	Laccase
LACS1	Long-chain acyl-CoA synthase 1
LSD	Least significant difference
MSA	Multiple sequence alignment
NAC	NAM, ATAF1/2 and CUC2
OFP	Orange fluorescent protein
ORF	Open reading frame
PAL	Phenylalanine ammonia-lyase
Py-MBMS	pyrolysis molecular beam mass spectrometry
qRT-PCR	Quantitative reverse transcription polymerase chain reaction
R1	Reproductive developmental stage 1
RAV	Related to API3/VP
SAM	Shoot apical meristems
STM	SHOOT MERISTEMLESS
SWN	Secondary wall NAC
S/G	Syringyl/guaiacyl
TALE	Three amino acid loop extension
TF	Transcription factor
UBQ	ubiquitin
UTR	Untranslated regions
ZmUbi1	<i>Zea mays</i> ubiquitin 1

Chapter 1: Introduction

Cellulose and hemicellulose constitute the major part of biomass on the planet. These cell wall polymers are made up of sugars that can be refined to produce biofuel. Using these valuable resources for sustainable production of biofuel may greatly minimize the environmental impact arising from relying on fossil fuels and secure future global energy demand. However, extracting sugars from the lignocellulosic biomass has been a daunting task due to biomass recalcitrance induced by the crystalline cellulose embedded in lignin and hemicellulose polymers hindering enzymatic hydrolysis of the cell wall carbohydrates to produce sugars for microbial fermentation. Recently, promising progress has been made in tackling the recalcitrance of lignocellulosic biomass. To ensure economic feasibility of biofuel production from lignocellulosic biomass, it would entail safeguarding continuous supply of ample biomass with optimal biomass composition. To meet this demand, tremendous improvements in plant biomass productivity that require less input (water and fertilizers) are essential. Moreover, relentless effort has to be made to improve relevant bioenergy traits by means of modifying plant architecture for higher biomass yield and tolerance against the duress of environmental adversity.

1.1 The challenges and prospects of lignocellulosic biofuel

Currently the need for a clean renewable bioenergy is increasing as a result of unstable prices and supply of fossil fuels as well as negative ecological impact from relying on such energy sources. The use of non-cellulosic biomass from crops such as corn and sugarcane to produce 1st generation biofuel has long been practiced although such activity has been widely criticized as diverting food away from human use and thus leading to global food shortage and a surge in food prices (Runge and Senauer, 2007). However, lignocellulosic biomass from fast-growing high biomass perennial C4 photosynthesis grasses such as switchgrass (*Panicum virgatum* L.) and miscanthus (*Miscanthus x giganteus*) dubbed ‘dedicated bioenergy feedstocks’ was regarded putative candidates to fulfill global energy demands. As a result of alarmingly increasing global climate change and subsequent expansion of drought throughout the world, C4 plants would be even more amenable owing to their higher efficiency of light energy conversion into biomass, high water and nitrogen (N) use efficiencies, ability to grow on marginal lands, higher tolerance

to environmental stresses such as salinity and water logging (Byrt et al., 2011). Moreover, C4 plants are thought to remove less soil nutrients such as nitrogen (N), phosphorus (P) and potassium (K) (Propheter and Staggenborg, 2010), have higher P use efficiency under P-limited conditions (Ghannoum and Conroy, 2007) and could reduce greenhouse gas by virtue of their ability to more effectively assimilate CO₂ compared to C3 plants. C4 plants such as switchgrass are therefore viable candidates as sources of lignocellulosic biofuel.

One of the factors dictating the feasibility of commercial production of liquid transportation biofuels from lignocellulosic bioenergy feedstocks is the quality of plant biomass determined by its composition as this affects the efficiency of enzymatic saccharification of the plant cell wall carbohydrates. The processing of lignocellulosic biomass to produce ethanol by fermentation in general, involves three major steps: acid pre-treatment, saccharification/hydrolysis and fermentation (Ragauskas et al., 2006). Reducing the recalcitrance of lignocellulosic biomass could thus eliminate or significantly reduce the costly pretreatment step as well as the inhibitors of microbial ethanol fermentation such as furfural and 5-hydroxymethylfurfural that are added during acid pretreatment (Liu and Blaschek, 2010). This, in turn, could help to reduce production costs and hence boost the economic competitiveness of the lignocellulosic ethanol production (Lynd et al., 2008). Therefore, overcoming biomass recalcitrance is one of the key targets for the success of lignocellulosic biofuel industry.

Other factors of consideration to transform the economic feasibility of 2nd generation biofuel industry is ensuring continuous supply of sustainable biomass resources, which could be achieved via the use of advanced breeding and/or genetic engineering approaches to develop bioenergy feedstock varieties endowed with beneficial traits such as higher biomass yield, tolerance to various environmental stress factors, modified plant architecture and optimal biomass composition. The last few years have seen rapid developments in genomic and biotechnological resources that facilitate the progress toward improving biofuel feedstocks for these major phenotypic traits (Vanholme et al., 2013; Yang et al., 2013; Phitsuwan et al., 2013; Poovaiah et al. 2014). Recent developments to reduce biomass recalcitrance are encouraging. However, much remains to be done in terms of evaluating the genetic basis of plant architecture,

stress resistance/tolerance and biomass yield in order to ensure the economic viability of lignocellulosic bioenergy feedstocks.

1.2 Genetic engineering of lignin biosynthesis

Lignin is a complex structural polymer whose main function is to provide plants with the necessary mechanical strength by enabling plants to withstand the negative pressure generated in the vessels during transpiration while serving as barriers against biotic agents. Its synthesis occurs via the phenylpropanoid and monolignol pathway from phenylalanine by subsequent hydroxylation, o-methylations and side chain modifications that are catalyzed by a series of enzymes to form three monolignols namely *p*-coumaryl, syringyl and coniferyl alcohols from which the three lignin monomers; *p*-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) are synthesized in the cytosol (Bonawitz & Chapple, 2010; Vanholme et al., 2010). The lignin monomers are then transported into the cell wall by a still poorly understood mechanism and incorporated in to a lignin polymer through oxidative polymerization. Although tremendous progress has been made in elucidating how lignin biosynthesis occurs in various plant species (Bonawitz & Chapple, 2010; Vanholme et al., 2010; Boerjan et al., 2003; Humphreys & Chapple, 2002), discoveries of new enzymes and reaction network have continued. Recently, a new enzyme, caffeoyl shikimate esterase (CSE) that catalyze the hydrolysis of caffeoyl shikimate in to caffeate in the lignin biosynthesis pathway has been reported (Vanholme et al., 2013). This enzyme helps bypass the second hydroxycinnamoyl-CoA shikimate/quinic acid hydroxycinnamoyl transferase (HCT) reaction together with the previously known enzyme 4-coumarate:CoA ligase (4CL) that converts the caffeate to caffeoyl CoA. New discoveries such as this, emphasize that aspects of the lignin biosynthesis pathway might need reevaluation, such as accounting for crosstalk with related pathways such as the chlorogenic acid pathway. Recently, extensive genome-wide analysis of lignin using transcriptomic and phylogenetic analysis revealed candidate genes that are thought to participate in downstream aspects of monolignol biosynthesis in switchgrass (Shen et al., 2013).

It is now well established that lignin content and composition is one of the major contributors to biomass recalcitrance significantly hindering enzymatic saccharification (Chen & Dixon, 2007; Fornalé et al. 2012; Mansfield et al., 2012; Xu et al., 2011; Fu et al., 2011a; Fu et al., 2011b and

Studer et al., 2011). Extensive genetic modification of lignin biosynthesis pathway have been implemented yielding significant improvements in saccharification efficiency as well as ethanol yield in various plant species including switchgrass. For instance, down regulation of the expression of several genes in the lignin biosynthesis pathway resulted in reduced lignin content and increased saccharification efficiency in both dicots and monocots (Chen & Dixon, 2007; Fu et al., 2011a; Fu et al., 2011b; Saathoff, et al., 2011; Xu et al., 2011; Fornale et al., 2012; Vanholme et al., 2013). The changes in lignin content and saccharification efficiency observed in transgenic plants differ between the different lignin biosynthetic genes manipulated. For instance, in alfalfa, of all the lignin biosynthetic genes, downregulation of upstream genes, such as C3H and HCT yielded the least change in lignin production resulting in wildtype levels of sugar release efficiency (Chen & Dixon, 2007). Genetic modification of lignin biosynthesis in switchgrass, by knocking down the expression of key lignin biosynthetic genes (4CL, CAD & COMT) showed a moderate reduction in lignin content and increase in sugar release or ethanol yield in transgenic biomass feedstocks (Fu et al., 2011a; Fu et al., 2011b; Saathoff, et al., 2011; Xu et al., 2011) .

Manipulation of the expression of lignin genes has frequently led to an undesirable phenotype such as severe reduction in biomass yield, particularly in dicots such as alfalfa (Shadle et al., 2007), poplar (Voelker et al., 2010) and also loss of vessel integrity in *Arabidopsis* (Brown et al., 2005). Moreover, the accompanied potential inhibitors of simultaneous saccharification and fermentation (SSF) as evidenced by an increased level of phenolics and aldehydes in *COMT* downregulated switchgrass lines (Tschaplinski et al., 2012) emphasize the need for a new approach to modify the lignin content in the lignocellulosic biomass. Thus, the search for more efficient strategies to improve the quality of lignocellulosic biomass has become a priority as evidenced by the weight of recent researches investigating alternative approaches to genetically manipulate the biosynthesis of lignin via engineering the regulatory networks of secondary cell wall biosynthesis (Shen et al., 2012; Shen et al., 2013; Yang et al., 2013).

1.3 Using transcriptional regulators to engineer lignin biosynthesis

Lignin biosynthesis is regulated by a network of TFs including top-tier NAC (NAM, ATAF1/2 and CUC2) and second-tier MYB master regulators (Zhong & Ye, 2014). Several NAC and

MYB transcriptional master regulators of secondary cell wall biosynthesis have been identified and functionally characterized, mainly in *Arabidopsis* (Zhao & Dixon, 2011; Zhong & Ye, 2014). The fiber-related NAC secondary wall thickening promoting factors (NST1/2/3) and vascular related NAC domain transcription factors (TF) (VND6/7) together known as the secondary wall NACs (SWN) can activate the entire network of secondary cell wall biosynthesis (Zhao & Dixon, 2011). The different homologs of SWNs regulate secondary wall biosynthesis in different plant tissues via the regulation of the expression of common downstream target MYB TFs, mainly MYB46/83, which, themselves, in turn, regulate a series of lower level MYB TFs (Hussey et al., 2013). Moreover, several transcriptional repressors of lignin biosynthetic genes have also been identified including the AtMYB4 from *Arabidopsis* (Jin et al., 2000) and its functional orthologs such as EgMYB1 from *Eucalyptus* (Legay et al., 2010), LIMYB1 from *Leucaena leucocephala* (Omer et al., 2013), ZmMYB42 from maize (Sonbol et al., 2009) and PvMYB4 from switchgrass (Shen et al., 2012), and also a Knotted1-like homeobox (KNAT1) from *Arabidopsis* (Mele et al., 2003). The regulation of lignin biosynthetic gene expression by MYB TFs is achieved via interaction with the cis-acting AC-elements in the promoter regions of most lignin genes. Some lignin biosynthetic genes such as *ferulic acid 5-hydroxylase (F5H)* have no AC-elements in their promoter region, and hence are directly regulated by SWNs (Zhao & Dixon, 2011).

The use of MYB and other transcriptional regulators of lignin biosynthesis to reduce lignin content and increase sugar release for biofuel is currently being explored (Shen et al., 2012; Baxter et al., 2015; Ambavaram et al., 2011; Zhao et al., 2010a). A recent study involving overexpression of the transcriptional repressor of the lignin biosynthesis pathway (*PvMYB4*) in switchgrass demonstrated a significant reduction in lignin content, improved sugar release efficiency and higher ethanol yield (more than two-fold) in transgenic switchgrass as compared to wild type plant (Shen et al., 2012, 2013; Baxter et al., 2015). Coupling reduced lignification with enhanced cellulose content via modification of its biosynthesis pathway have been demonstrated in a study involving heterologous expression of *Arabidopsis* AP2/ERF TF (*AtSHN2*) in rice (*Oryza sativa*) (Ambavaram et al., 2011). The observed reduction in lignin and increase in cellulose contents in these transgenic plants were presumably achieved via coordinated transcriptional regulation of cell wall biosynthesis-related TFs (SWNs and MYB),

the promoter cis-elements of which have been shown to be directly bound by AtSHN2 (Ambavaram et al., 2011). The rice ortholog (OsSHN) of AtSHN2 was also proposed to have a similar association with cell wall regulatory and biosynthetic pathways based on global coexpression network of rice genes (Ambavaram et al., 2011). It is vital to experimentally confirm whether the homologs of these genes in monocots have similar transcriptional regulatory activity as a global regulator of secondary cell wall biosynthesis by selectively regulating the expression of cellulose and lignin biosynthetic genes. A recent study in sorghum identified a basic helix-loop-helix TF (SbbHLH1) that displayed similar selective effects on the deposition of secondary cell wall polymers when overexpressed in *Arabidopsis* (Yan et al., 2013). Surprisingly, despite the reduction in the expression of most lignin genes, the expression of SWNs was unaffected while that of lower-level MYB master regulators (*MYB46/83*) and other MYBs (*MYB58/63*) were upregulated in transgenic plants expressing *SbbHLH1* (Yan et al., 2013). The inability of MYB activators of the lignin biosynthesis pathway to outcompete *SbbHLH1* in dictating the expression level of lignin biosynthetic genes has been suggested to be associated with stronger interaction of *SbbHLH1* with the promoters of lignin biosynthetic genes or the formation of a complex between the MYB master regulators and *SbbHLH1* TF that may function as a lignin repressor with the resulting reduction in lignin content activating the expression of MYBs via feedback regulation (Yan et al., 2013).

1.4 Engineering plants for modified plant architecture, stress tolerance and biomass yield

Besides modifications of biomass composition for improved saccharification efficiency, improving switchgrass biomass yield is also important to ensure economic feasibility by increasing energy yield/land area, which helps offset the energy utilized and cost incurred for its production (Casler & Vogel, 2014). The current biomass yield potential of C4 bioenergy grasses grown without irrigation is only about 25-33% of the theoretical maximum yield under optimal conditions (Mullet et al., 2014). Molecular and genetic studies have identified various regulators of plant biomass productivity including different phytohormones and transcription factors that act by controlling the activities of vegetative meristems (Demura & Ye, 2010). Manipulation of vegetative development to stimulate biomass productivity could thus be achieved via modification of the activities of these regulators that controls many aspects of plant development

including vegetative to reproductive phase transition. For instance, extending vegetative growth phases, particularly in C4 grasses, has been suggested as a viable strategy for increased biomass accumulation, owing perhaps to allocation of resources to the development of vegetative rather than reproductive organs (Mullet et al., 2014). Overexpression of *FLOWERING LOCUS C* (*FLC*) gene that regulates flowering time has been shown to substantially delay flowering in *Arabidopsis* thereby resulting in increased biomass yield (Salehi et al., 2005). Interestingly, manipulation of the phase transition from vegetative to reproductive stages in alfalfa has been shown not only to increase total biomass productivity but also simultaneously reduce lignin content (Tadege et al., 2014). These studies, hence, highlight the potential of coordinately optimizing the biomass productivity and composition to effectively reduce the total cost and time needed to develop an elite bioenergy feedstock line.

Extended vegetative growth phase for increasing total biomass yield of bioenergy feedstocks such as switchgrass, would require an extended growing season, which could be problematic for lowland switchgrass cultivars that are sensitive to freezing temperatures during winter. Therefore, genetic modification of plants for cold tolerance to survive the winter would be useful to accommodate with this goal (Casler and Vogel, 2014). Transcriptional regulators such as those belonging to the *Apetala 2* (*AP2*)/ ethylene responsive factors (*ERF*) superfamily, which controls the expression of stress response genes to allow plants tolerate such stress factors (Licausi et al., 2013) may be helpful in this regard.

Genetic modification of plant architecture via alterations in leaf structure, tiller number, height and thickness could also contribute to increasing biomass productivity (Stamm et al., 2012). The significance of manipulating plant architecture for increased grain yield has been demonstrated by the semi-dwarf green revolution varieties of wheat and rice (Peng et al., 1999; Sasaki et al., 2002). However, little effort has been made to improve the architecture of bioenergy feedstocks such as switchgrass for increased biomass productivity. It is notable that plant architecture is controlled by the vegetative meristems whose activity is regulated by the signalling pathways involving various phytohormones as well as transcriptional regulators (Reinhardt & Kuhlemeier, 2002; Stamm et al., 2012). Phytohormones such as gibberellin (GA) and cytokinin, as essential regulators of plant growth and development, could play crucial roles in determining plant

architecture. Manipulation of the biosynthetic pathway of these phytohormones or its signalling pathways could hence be utilized to alter plant architecture, which may lead to increased biomass yield and/or better performance under adverse environments. Particularly, GA signalling has been suggested as an attractive target for genetic manipulation of tiller height and number as it acts in a dose-dependent manner (Busov et al., 2008). Manipulation of the biosynthesis of GA or its signalling pathway intermediates has been associated with positive pleiotropic effects in nitrogen assimilation (Nagel and Lambers, 2002), photosynthetic efficiency (Biemelt et al., 2004) and lateral root production (Busov et al., 2006) making it viable for improvement of bioenergy feedstocks. Moreover, GA signalling has been indicated to be involved in the regulation of lignin biosynthetic pathway (Biemelt et al., 2004; Zhao et al., 2010b) and thus may be utilized to alter biomass composition as well.

The overall plant productivity is determined by the combination of factors such as total radiation per unit area, light interception efficiency, and the conversion rate of the intercepted radiation in to biomass. Therefore, manipulation of leaf size or its orientation to enhance light interception efficiency could be a potential target to enhance plant productivity. A recent study suggested that increased leaf area could lead to increased biomass yield as a result of increased light interception efficiency (Dohleman and Long, 2009). Engineering plants for erect leaf phenotypes has been demonstrated by targeting the biosynthesis of brassinosteroid (BR) hormone and showed to enhance light capture efficiency and hence photosynthesis resulting in increased biomass and grain yield in rice (Sakamoto et al., 2006). Overexpression of *Arabidopsis* NAC domain TF gene, *Long Vegetative Phase 1* (*AtLOV1*), in switchgrass also showed altered plant architecture characterized by erect leaf phenotype that was accompanied by increased lignin content and delayed flowering (Xu et al., 2012). Another TF belonging to AP2/ERF superfamily, *AINTEGUMENTA* (*ANT*), has been implicated in the regulation of plant organ size, since its overexpression in *Arabidopsis* was associated with significantly increased leaf size and flowers via proliferation of the number of cells (Krizek, 1999; Mizukami and Fischer, 2000). Other TFs, including auxin-regulated genes controlling organ size (*ARGOS*), and growth regulating factors (*GRF*), have also been similarly shown to have similar growth-promoting effects (Hu et al., 2003, 2006). Therefore, these TFs could also be utilized to genetically alter the leaf organ size in

switchgrass or other bioenergy feedstocks to increase the plants' ability to intercept light thereby enhancing biomass productivity.

The overall goal of this dissertation research was to investigate alternative ways to manipulate the switchgrass biomass composition and other traits of significance for bioenergy to improve the economic viability of biofuel production from lignocellulosic biomass. Specifically, the aims of this study were to improve switchgrass biomass quality by reducing biomass recalcitrance and modifying plant architecture to increase yield on marginal lands intended for switchgrass cultivation, via genetic manipulations of the GA catabolic pathway as well as by the overexpression of putative transcriptional regulators of the cell wall biosynthetic pathway. The other specific objective was to identify switchgrass AP2/ERF superfamily of TFs, assess their putative roles in the regulation of plant growth and development specially biomass composition as well as roles in the regulation of responses to different environmental stimuli and ultimately provide list of potential target genes for biotechnological application in the improvement of bioenergy feedstocks.

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.
- Ambavaram MM, Krishnan A, Trijatmiko KR, Pereira A. 2011. Coordinated activation of cellulose and repression of lignin biosynthesis pathways in rice. *Plant Physiology* 155:916-931.
- Baxter H, Poovaiah C, Yee K, Mazarei M, Rodriguez M, Jr., Thompson O, Shen H, Turner G, Decker S, Sykes RW, Chen F, Davis M, Mielenz J, Davison B, Dixon R, Stewart CN Jr. 2015. Field evaluation of transgenic switchgrass plants overexpressing *PvMYB4* for reduced biomass recalcitrance. *Bioenergy Research*: Doi. 10.1007/s12155-014-9570-1.
- Biemelt S, Tschiersch H and Sonnewald U. 2004. Impact of altered gibberellin metabolism on biomass accumulation, lignin biosynthesis, and photosynthesis in transgenic tobacco plants. *Plant Physiology* 135:254-265.
- Boerjan W, Ralph J and Baucher M. 2003. Lignin biosynthesis. *Annual Review of Plant Biology* 54:519–546.
- Bonawitz ND and Chapple C. 2010. The genetics of lignin biosynthesis: connecting genotype to phenotype. *Annual Review of Genetics* 44:337–363.
- Brown DM, Zeef LA, Ellis J, Goodacre R, Turner SR. 2005. Identification of novel genes in *Arabidopsis* involved in secondary cell wall formation using expression profiling and reverse genetics. *Plant Cell* 17:2281–2295.
- Busov V, Meilan R, Pearce DW, Rood SB, Ma C, Tschaplinski TJ, Strauss SH. 2006. Transgenic modification of *gai* or *rgl1* causes dwarfing and alters gibberellins, root growth, and metabolite profiles in *Populus*. *Planta* 224:288-299.
- Busov VB, Brunner AM, Strauss SH. 2008. Genes for control of plant stature and form. *New Phytologist* 177:589-607.

- Byrt CS, Grof CP, Furbank RT. 2011. C4 plants as biofuel feedstocks: optimizing biomass production and feedstock quality from a lignocellulosic perspective. *Journal of Integrative Plant Biology* 53:120-135.
- Casler MD, Vogel KP. 2014. Selection for biomass yield in upland, lowland, and hybrid switchgrass. *Crop Science* 54:626-636.
- Chen F and Dixon RA. 2007. Lignin modification improves fermentable sugar yields for biofuel production. *Nature Biotechnology* 25:759-761.
- Demura T, Ye Z-H. 2010. Regulation of plant biomass production. *Current Opinion in Plant Biology* 13:298-303.
- Dohleman FG, Long SP. 2009. More productive than maize in the Midwest: How does *Miscanthus* do it? *Plant Physiology* 150:2104-2115.
- Fornalé S, Capellades M, Encina A, Wang K, Irar S, Lapierre C, Ruel K, Joseleau J-P, Berenguer J, Puigdomènech P, Rigau J, Caparrós-Ruiz D. 2012. Altered lignin biosynthesis improves cellulosic bioethanol production in transgenic maize plants down-regulated for *Cinnamyl Alcohol Dehydrogenase*. *Molecular Plant* 5:817-830.
- 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. *Proceedings of the National Academy of Sciences USA* 108:3803-3808.
- Fu C, Xiao X, Xi Y, Ge Y, Chen F, Bouton J, Dixon R, Wang Z-Y. 2011b. Downregulation of *cinnamyl alcohol dehydrogenase (CAD)* leads to improved saccharification efficiency in switchgrass. *BioEnergy Research*. 4:153-164.
- Ghannoum O, Conroy JP. 2007. Phosphorus deficiency inhibits growth in parallel with photosynthesis in a C3 (*Panicum laxum*) but not two C4 (*P. coloratum* and *Cenchrus ciliaris*) grasses. *Functional Plant Biology* 34:72-81.

- Humphreys JM and Chapple C. 2002. Rewriting the lignin roadmap. *Current Opinion in Plant Biology* 5:224-229.
- Hu Y, Xie Q, Chua NH. 2003. The *Arabidopsis* auxin-inducible gene *ARGOS* controls lateral organ size. *Plant Cell* 15:1951-1961.
- Hu Y, Poh HM, Chua N-H. 2006. The *Arabidopsis* *ARGOS-LIKE* gene regulates cell expansion during organ growth. *Plant Journal* 47:1-9.
- Hussey SG, Mizrachi E, Creux NM, Myburg AA. 2013. Navigating the transcriptional roadmap regulating plant secondary cell wall deposition. *Frontiers in Plant Science* 29:325.
- Jin H, Cominelli E, Bailey P, Parr A, Mehrrens F, Jones J, Tonelli C, Weisshaar B, Martin C. 2000. Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in *Arabidopsis*. *The EMBO Journal* 19, 6150–6161.
- Krizek BA. 1999. Ectopic expression of *AINTEGUMENTA* in *Arabidopsis* plants results in increased growth of floral organs. *Developmental Genetics* 25:224-236.
- Licausi F, Ohme-Takagi M, Perata P. 2013. APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. *New Phytologist* 199:639-649.
- Liu ZL and Blaschek HP. 2010. Biomass conversion inhibitors and in situ detoxification. In: *Biomass to biofuels: strategies for global industries*. Vertès AA, Qureshi N, Blaschek HP, Yukawa H. (eds.), Oxford, UK: Blackwell Publishing Ltd.
- Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, Himmel M, Keller M, McMillan JD, Sheehan J, Wyman CE. 2008. How biotech can transform biofuels. *Nature Biotechnology* 26:169–172.
- Mansfield SD, Kang K-Y, Chapple C. 2012. Designed for deconstruction- poplar trees altered in cell wall lignification improve the efficacy of bioethanol production. *New Phytologist* 194, 91-101.

Mele G, Ori N, Sato Y, Hake S. 2003. The knotted1-like homeobox gene *BREVIPEDICELLUS* regulates cell differentiation by modulating metabolic pathways. *Genes Development* 17:2088–2093.

Mizukami Y, Fischer RL. 2000. Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. *Proceedings of the National Academy of Sciences, USA* 97:942-947.

Mullet J, Morishige D, McCormick R, Truong S, Hilley J, McKinley B, Anderson R, Olson SN, Rooney W. 2014. Energy Sorghum-a genetic model for the design of C4 grass bioenergy crops. *Journal of Experimental Botany* 65:3479-3489.

Nagel OW, Lambers H. 2002. Changes in the acquisition and partitioning of carbon and nitrogen in the gibberellin-deficient mutants A70 and W335 of tomato (*Solanum lycopersicum* L.). *Plant, Cell & Environment* 25:883-891.

Omer S, Kumar S, Khan B. 2013. Over-expression of a subgroup 4 R2R3 type MYB transcription factor gene from *Leucaena leucocephala* reduces lignin content in transgenic tobacco. *Plant Cell Reports* 32:161-171.

Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM., Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD and Harberd NP. 1999. 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* 400:256-261.

Phitsuwan P, Sakka K, Ratanakhanokchai K. 2013. Improvement of lignocellulosic biomass in planta: A review of feedstocks, biomass recalcitrance, and strategic manipulation of ideal plants designed for ethanol production and processability. *Biomass & Bioenergy* 58:390-405.

Poovaliah CR, Nageswara-Rao M, Soneji JR, Baxter HL, Stewart CN Jr. 2014. Altered lignin biosynthesis using biotechnology to improve lignocellulosic biofuel feedstocks. *Plant Biotechnology Journal* 12:1163-1173.

Propheter JL, Staggenborg S. 2010. Performance of annual and perennial biofuel crops: nutrient removal during the first two years. *Agronomy Journal* 102:798-805.

- Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ, Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T. 2006. The path forward for biofuels and biomaterials. *Science* 311:484-489.
- Reinhardt D, Kuhlemeier C. 2002. Plant architecture. *EMBO Reports* 3:846–851.
- Runge F and Senauer B. 2007. How biofuels could starve the poor. *Foreign Affairs* 86:41-53.
- Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, Swapan D, Ishiyama K, Saito T, Kobayashi M, Khush GS, Kitano H and Matsuoka M. 2002. Green revolution: a mutant gibberellin-synthesis gene in rice. *Nature* 416:701-702.
- 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* 6:804-807.
- Sakamoto T, Morinaka Y, Ohnishi T, Sunohara H, Fujioka S, Ueguchi-Tanaka M, Mizutani M, Sakata K, Takatsuto S, Yoshida S, Tanaka H, Kitano H, Matsuoka M. 2006. Erect leaves caused by brassinosteroid deficiency increase biomass production and grain yield in rice. *Nature Biotechnology* 24:105-109.
- Salehi H, Ransom CB, Oraby HF, Seddighi Z, Sticklen MB. 2005. Delay in flowering and increase in biomass of transgenic tobacco expressing the *Arabidopsis* floral repressor gene *FLOWERING LOCUS C*. *Journal of Plant Physiology* 162:711-717.
- Shadle G, Chen F, Srinivasa Reddy MS, Jackson L, Nakashima J, Dixon RA. 2007. Down-regulation of *hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase* in transgenic alfalfa affects lignification, development and forage quality. *Phytochemistry* 68:1521-1529.
- Shen H, He X, Poovaiah CR, Wuddineh WA, Ma J, Mann,DGJ, Wang H, Jackson L, Tang Y, Stewart CN Jr., Chen F and Dixon RA. 2012. Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks. *New Phytologist* 193:121-136.

Shen H, Poovaiah CR, Ziebell A, Tschaplinski TJ, Pattathil S, Gjersing E, Engle NL, Katahira R, Pu Y, Sykes RW, Chen F, Ragauskas AJ, Mielenz JR, Hahn MG, Davis M, Stewart CN Jr., Dixon RA. 2013. Enhanced characteristics of genetically modified switchgrass (*Panicum virgatum* L.) for high biofuel production. *Biotechnology for Biofuels* 6:71.

Sonbol F-M, Fornalé S, Capellades M, Encina A, Touriño S, Torres J-L, Rovira P, Ruel K, Puigdomènech P, Rigau J, Caparrós-Ruiz D. 2009. The maize ZmMYB42 represses the phenylpropanoid pathway and affects the cell wall structure, composition and degradability in *Arabidopsis thaliana*. *Plant Molecular Biology* 70:283-296.

Stamm P, Verma V, Ramamoorthy R, Kumar PP. 2012. Manipulation of plant architecture to enhance lignocellulosic biomass. *AoB Plants* pls026. doi:10.1093/aobpla/pls026.

Studer MH, DeMartini JD, Davis MF, Sykes RW, Davison B, Keller M, Tuskan GA, Wyman CE. 2011. Lignin content in natural *Populus* variants affects sugar release. *Proceedings of the National Academy of Sciences USA* 108:6300-6305.

Tadege M, Chen F, Murray J, Wen J, Ratet P, Udvardi M, Dixon R, Mysore K. 2014. Control of vegetative to reproductive phase transition improves biomass yield and simultaneously reduces lignin content in *Medicago truncatula*. *BioEnergy Research* 8:857-867.

Tschaplinski TJ, Standaert RF, Engle NL, Martin MZ, Sangha AK, Parks JM, Smith JC, Samuel R, Jiang N, Pu Y, Ragauskas AJ, Hamilton CY, Fu C, Wang ZY, Davison BH, Dixon RA, Mielenz JR. 2012. Down-regulation of the *caffeic acid O-methyltransferase* gene in switchgrass reveals a novel monolignol analog. *Biotechnology for Biofuels* 5:71.

Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W. 2010. Lignin biosynthesis and structure. *Plant Physiology* 153:895-905.

Vanholme B, Desmet T, Ronsse F, Rabaey K, Breusegem FV, Mey MD, Soetaert W, Boerjan W. 2013a. Towards a carbon-negative sustainable bio-based economy. *Frontiers in Plant Science* 4:174.

Vanholme R, Cesarino I, Rataj K, Xiao Y, Sundin L, Goeminne G, Kim H, Cross J, Morreel K, Araujo P, Welsh L, Hastraete J, McClellan C, Vanholme B, Ralph J, Simpson GG, Halpin C, Boerjan W. 2013b. Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in *Arabidopsis*. *Science* 341:1103-1106.

Voelker SL, Lachenbruch B, Meinzer FC, Jourdes M, Ki C, Patten AM, Davin LB, Lewis NG, Tuskan GA, Gunter L, Decker SR, Selig MJ, Sykes RW, Himmel ME, Kitin P, Shevchenko O, Strauss SH. 2010. Antisense down-regulation of *4CL* expression alters lignification, tree growth, and saccharification potential of field-grown poplar. *Plant Physiology* 154:874–886.

Xu B, Escamilla-Treviño LL, Sathitsuksanoh N, Shen Z, Shen H, Percival Zhang YH, Dixon RA, Zhao B. 2011. Silencing of *4-coumarate: Coenzyme A ligase* in switchgrass leads to reduced lignin content and improved fermentable sugar yields for biofuel production. *New Phytologist* 92:611-625.

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

Yang F, Mitra P, Zhang L, Prak L, Verherbruggen Y, Kim J-S, Sun L, Zheng K, Tang K, Auer M, Scheller HV, Loqué D. 2013. Engineering secondary cell wall deposition in plants. *Plant Biotechnology Journal* 11:325-335.

Yan L, Xu C, Kang Y, Gu T, Wang D, Zhao S, Xia G. 2013. The heterologous expression in *Arabidopsis thaliana* of sorghum transcription factor SbbHLH1 downregulates lignin synthesis. *Journal of Experimental Botany* 64:3021-3032.

Zhao Q and Dixon RA. 2011. Transcriptional networks for lignin biosynthesis: more complex than we thought? *Trends in Plant Science* 16:227-233.

Zhao Q, Gallego-Giraldo L, Wang H, Zeng Y, Ding SY, Chen F, Dixon RA. 2010a. A NAC transcription factor orchestrates multiple features of cell wall development in *Medicago truncatula*. *Plant Journal* 63:100-114.

Zhao XY, Zhu DF, Zhou B, Peng WS, Lin JZ, Huang XQ, He RQ, Zhuo YH, Peng D, Tang DY, Li MF, Liu XM. 2010b. Over-expression of the *AtGA2ox8* gene decreases the biomass accumulation and lignification in rapeseed (*Brassica napus* L.). *Journal of Zhejiang University Science B* 11:471-481.

Zhong R, Ye Z-H. 2014. Complexity of the transcriptional network controlling secondary wall biosynthesis. *Plant Science* 229:193-207.

Chapter 2: Identification and overexpression of *gibberellin 2-oxidase* (GA2ox) in switchgrass (*Panicum virgatum* L) for improved plant architecture and reduced biomass recalcitrance

(A version of this chapter was published in Plant Biotechnology Journal (13(5):636-647) with the following authors; Wegi A. Wuddineh, Mitra Mazarei, Jiyi Zhang, Charleson R. Poovaiah, David G. J. Mann, Angela Ziebell, Robert Sykes, Mark Davis, Michael Udvardi, C. Neal Stewart, Jr.)

Wegi A. Wuddineh designed and performed the experiments except lignin and sugar release assays, analyzed the data, and wrote the manuscript.

2.1 Abstract

Gibberellin 2-oxidases (GA2oxs) are a group of 2-oxoglutarate-dependent dioxygenases that catalyze the deactivation of bioactive GA or its precursors through 2 β -hydroxylation reaction. In this study, putatively novel switchgrass C₂₀ GA2ox genes were identified with the aim of genetically engineering switchgrass for improved architecture and reduced biomass recalcitrance for biofuel. Three C₂₀ GA2ox genes showed differential regulation patterns among tissues including roots, seedlings and reproductive parts. Using a transgenic approach, we showed that overexpression of two C₂₀ GA2ox genes, i.e., *PvGA2ox5* and *PvGA2ox9*, resulted in characteristic GA-deficient phenotypes with dark-green leaves and modified plant architecture to be associated with GA2ox transcript abundance. Exogenous application of GA rescued the GA-deficient phenotypes in transgenic lines. Transgenic semi-dwarf lines displayed increased tillering and reduced lignin content and the syringyl/guaiacyl lignin monomer ratio accompanied by the reduced expression of lignin biosynthetic genes compared to non-transgenic plants. A moderate increase in the level of glucose release in these transgenic lines might be attributed to reduced biomass recalcitrance as a result of reduced lignin content and lignin composition. Our results suggest that overexpression of GA2ox genes in switchgrass is a feasible strategy to improve plant architecture and reduce biomass recalcitrance for biofuel.

Keywords: gibberellin, gibberellin 2-oxidase, semi-dwarf, lignin, biofuel

2.2 Introduction

Gibberellins are a large group of diterpenoid natural products characterized by the presence of tetracyclic 6-5-6-5 ring structure derived from *ent*-gibberellane (Peters, 2012). Approximately 136 GAs have been identified from plants, fungi and bacteria. GAs are structurally classified into two groups, namely, C₂₀s and C₁₉s based on the number of carbon atoms in their ring structure. Only C₁₉ GAs such as: GA₁, GA₃, GA₄ and GA₇ are known to be biologically active (Hedden & Phillips, 2000; Hedden & Thomas, 2012; Peters, 2012). GA plays major roles in the regulation of various developmental and growth processes that have enormous implication for agriculture. One of the common functions of GA includes GA-stimulation of shoot elongation that has been extensively utilized for genetic improvement of cereal crops (Harberd et al., 1998; Schwechheimer & Willige, 2009; Hedden & Thomas, 2012). A very well-known example is found in the semi-dwarf and high-yielding Green Revolution varieties of wheat (‘Lerma Rojo 64’ and ‘Sonora 64’) and rice (IR8) developed through breeding, which are now attributed to mutations in either GA signalling pathway intermediates (*Reduced height1*, *Rht1*) (Peng et al., 1999) or gibberellin biosynthetic genes (rice *semi-dwarf1*, *sd1*) (Sasaki et al., 2002). Moreover, recent studies have shown that the level of bioactive GA is negatively correlated with plant tillering and adventitious root development especially among cereal grains (Lo et al., 2008), whereas it is positively correlated with flower and seed formation (Sakamoto et al., 2003; Schomburg et al., 2003; Rieu et al., 2008a; El-Sharkawy et al., 2012). Additionally, GA has been implicated with increased lignin deposition in eudicots (Biemelt et al., 2004; Zhao et al., 2010). To date, there are no reports on how lignification in monocots is affected by GA.

Bioactive GAs regulation appears to be tightly controlled in plant tissues via rates of GA biosynthesis and deactivation (Olszewski et al., 2002; Yamaguchi, 2008). The biosynthesis of GA has been characterized in several plant species. In general, there are three main enzyme classes that are involved in GA biosynthesis. First, the terpene cyclases catalyze the first two cyclization steps from the linear geranylgeranyl diphosphate (GGDP) to the cyclic *ent*-kaurene. Second, the cytochrome P450 mono-oxygenases catalyze the formation of the first GA, GA₁₂ (Helliwell et al., 2001). The last step of GA biosynthesis is catalyzed by a group of 2-oxoglutarate-dependent dioxygenases namely, GA 20-oxidase (GA20ox) and GA 3-oxidase

(GA3ox), which are localized in the cytosol (Olszewski et al., 2002; Sun & Gubler, 2004; Hedden & Thomas, 2012). A distinct group of 2-oxoglutarate-dependent dioxygenases known as GA2oxs, however, irreversibly catalyze the deactivation of bioactive GA or its precursors via 2- β hydroxylation (Thomas et al., 1999; Olszewski et al., 2002).

Several C₁₉ GA-catalyzing GA2oxs have been identified and functionally characterized in various plant species including *Arabidopsis*, pea, rice, poplar, runner bean, pumpkin and wheat (Thomas et al., 1999; Lester et al., 1999; Martin et al., 1999; Sakamoto et al., 2001; Busov et al., 2003; Solfanelli et al., 2005; Radi et al., 2006; Appleford et al., 2007; Dijkstra et al., 2008; Lo et al., 2008; Rieu et al., 2008b). Transgenic overexpression of C₁₉ GA2ox has caused growth and reproductive abnormalities in rice and *Arabidopsis* hindering their potential use in crop improvement without the use of tissue-specific promoters for targeting their expression to non-reproductive organs (Sakamoto et al., 2003; Lee et al., 2014). Alternatively, a small group of C₂₀ GA2ox families comprising rice GA2ox6 (Lo et al., 2008; Huang et al., 2010), *Arabidopsis* GA2ox7/8 (Schomburg et al., 2003) and spinach GA2ox3 (Lee & Zeevaart, 2005) were shown to catalyze the 2- β hydroxylation of C₂₀ GAs (GA₁₂ and GA₅₃) to form inactive GAs (GA₁₁₀ and GA₉₇), respectively (Lee & Zeevaart, 2005). Transgenic expression of the genes coding for these GA2ox proteins were shown to result in not only less severe dwarf phenotypes but also positively affected root growth with less effect on floral and seed development in rice (Sakamoto et al., 2003; Lo et al., 2008). Heterologous overexpression of *AtGA2ox8* in canola resulted in the development of semi-dwarf lines with normal seed yield but significantly higher seed weight and increased seed oil content (Zhao et al., 2010). Moreover, overexpression of *AtGA2ox1* in tobacco (Biemelt et al., 2004) and *AtGA2ox8* in canola (Zhao et al., 2010) reduced both lignification and the expression of lignin biosynthetic genes via reduction of the bioactive GA in the plant. A very recent study has also shown that *OsGA2ox5* overexpression was associated with improved resistance to high salinity (Shan et al., 2014).

Modifications of plant architecture to develop compact semi-dwarf plants with more tillers per plant, and higher biomass, might have enormous potential for the improvement of bioenergy feedstocks (Jakob et al., 2009), with GA being an obvious phytohormone to target for altering plant architecture (Stamm et al., 2012). One of the major hurdles of lignocellulosic biofuel feedstock development is the biomass recalcitrance (resistance of the cellulose and hemicellulose

in the plant biomass to breakdown into fermentable sugars). Reduced lignin content and/or lignin composition through genetic engineering of lignin biosynthetic genes and transcriptional regulators of lignin biosynthesis has shown promising results in addressing the recalcitrance issue (Fu et al., 2011; Xu et al., 2011; Shen et al., 2012; Shen et al., 2013; Baxter et al., 2014). Thus, manipulation of the level of bioactive GA in the plant may also provide an alternative strategy to reduce biomass recalcitrance as GA has already been indicated to play a crucial role in the regulation of plant lignification in several eudicots (Biemelt et al., 2004; Zhao et al., 2010).

Switchgrass (*Panicum virgatum* L.) is a leading candidate for lignocellulosic biofuel feedstocks because of its high biomass yield, resistance to stress conditions, high nutrient-use efficiency, fast growth and ability to thrive on marginal soil conditions (Yuan et al., 2008). To the best of our knowledge, there are no published results on the switchgrass GA catabolic pathways and the genes involved in catalyzing these reactions. Despite the enormous potential that manipulation of GA catabolic pathway has on the improvement of bioenergy feedstocks, little effort has been made to exploit these benefits. Therefore, the purpose of this study was to investigate the impact of genetic manipulation of GA catabolic pathway genes via overexpression of the C₂₀ GA2ox genes on plant architecture, the lignin content, and hence the biomass yield and recalcitrance in switchgrass. In this study, we identified two GA2ox genes in the lignocellulosic biofuel feedstock switchgrass. Stable transgenic switchgrass plants were produced whereby overexpression of *PvGA2ox* yielded plants with reduced biomass recalcitrance by decreasing lignin content and composition. Our results indicate that plant architecture and other biomass characteristics such as lignification and biomass recalcitrance could be optimized for sustainable energy production by targeting GA biosynthesis.

2.3 Experimental procedures

2.3.1 Plant materials and growth conditions

Plants were grown in growth chambers under standard conditions (16 h day/8 h night light at 24°C, 390 $\mu\text{E m}^{-2} \text{s}^{-1}$) and watered three times per week, including weekly nutrient supplements with Peter's 20-20-20 fertilizer. Transgenic and non-transgenic 'Alamo' ST1 clone lines were propagated from a single tiller in three replicates for measuring growth parameters. For

expression pattern analysis, root, leaf blade, leaf sheath, internode, and panicle samples were collected from tillers at R1 developmental stage while the remaining samples were collected from two-week-old seedlings, E1 (elongation stage with one internode) crown and inflorescence meristem of tillers at E5 (elongation stage with five internodes) stage for assaying transgene transcript abundance (Moore et al., 1991; Shen et al., 2009). Each sample was snap frozen in liquid nitrogen and macerated with mortar and pestle in liquid nitrogen. Alternatively, samples were stored at -80°C for subsequent maceration. The macerated samples were used for RNA extraction as described below.

2.3.2 Transgene candidate identification, vector construction and plant transformation

The TBLASTN program was run to identify the homologous gene sequences in all available switchgrass expressed sequence tag (EST) databases using the amino acid sequences of AtGA2ox8 (At4g21200), OsGA2ox5 (LOC_Os07g01340.1), OsGA2ox6 (LOC_Os04g44150.1) and OsGA2ox9 (LOC_Os02g41954.1). Phylogenetic trees and multiple sequence alignment (MSA) analysis were used to identify the most closely related genes for cloning. For overexpression of *PvGA2ox5* and *PvGA2ox9*, the open reading frame (ORF) of the genes were isolated from cDNA of the ST1 clone of switchgrass ‘Alamo’ cultivar using individual gene specific primers flanking the ORF of each gene. Both genes were cloned into pCR8 entry vector for sequencing, and sub-cloned into pANIC-10A expression vector (Mann et al., 2012) by GATEWAY recombination cloning system. The pANIC-10A has the maize ubiquitin 1 (ZmUbi1) promoter driving the expression of the switchgrass *GA2ox* genes. Embryogenic callus derived from ST1 switchgrass genotype was transformed with the expression vector construct through *Agrobacterium*-mediated transformation (Burris et al., 2009). Antibiotic selection was carried out for about two months on 30- 50 mg/L hygromycin followed by regeneration of orange fluorescent protein (*pporRFP*; OFP) reporter positive callus sections on regeneration medium (Li & Qu, 2011) containing 400 mg/L timentin. Regenerated plants were rooted on MS medium (Murashige & Skoog, 1962) plus 250 mg/L cefotaxime (Grewal et al., 2006) and the transgenic lines were screened based on the presence of the insert and expression of the transgene. Rice transformation was performed using callus derived from mature seeds of rice variety TP309 as described before (Nishimura et al., 2006).

2.3.3 RNA extraction and qRT-PCR

For transgene transcript analysis, total RNA was extracted from leaf and stem samples of transgenic and non-transgenic lines using Tri-Reagent (Molecular Research Center, Cincinnati, OH). The purified RNA (3 µg) was treated with DNase-I (Promega) to remove any potential genomic DNA contaminants. The DNase-treated RNA was used for first-strand cDNA synthesis using High-Capacity cDNA Reverse Transcription kit (Applied Biosystems). qRT-PCR analysis was conducted using Power SYBR Green PCR master mix (Applied Biosystems) according to the manufacturer's protocol. All the experiments were conducted in triplicates. The list of all primer pairs used for qRT-PCR is shown in Table 2-1. Analysis of the relative expression was done by the change in Ct method using ubiquitin (UBQ) (Switchgrass Unitranscript ID: AP13CTG25905) as a reference gene (Shen et al., 2009). No amplification product was observed with all the primer pairs when using only the RNA samples or water instead of cDNA.

2.3.4 Phloroglucinol staining

For lignin staining analysis, leaf samples were collected at R1 developmental stage and cleared in a 2:1 solution of ethanol and glacial acetic acid for five days (Bart et al., 2010). Subsequently, the cleared leaf sample was immersed in 1% phloroglucinol (in 2:1 ethanol/HCl) overnight for staining. Low magnification microphotographs were taken using an Infinity X32 digital camera mounted on Fisher Scientific Stereomaster microscope.

2.3.5 Lignin content and composition by py-MBMS

For the quantification of lignin content and S/G lignin monomer ratio, tillers were collected at R1 developmental stage, air dried for 3 weeks at room temperature, and milled to 1 mm (20 mesh) particle size. Subsequently, lignin content and composition were determined via National Renewable Energy Laboratory (NREL) high-throughput py-MBMS on extractives- and starch-free samples (Sykes et al., 2009; Baxter et al., 2014).

2.3.6 Sugar release

Tiller samples were collected at R1 developmental stage and air-dried for 3 weeks at room temperature before grinding to 1 mm (20 mesh) particle size. Sugar release efficiency was

determined via NREL high-throughput sugar release assays on extractives- and starch-free samples (Studer et al., 2010; Decker et al., 2012; Baxter et al., 2014). Glucose and xylose release were determined by colorimetric assays, and total sugar release is the sum of glucose and xylose released.

2.3.7 Data analysis

Fisher's least significant difference (LSD) procedure was used to perform multiple comparisons between means of treatments using SAS version 9.3 (SAS Institute Inc., Cary, NC). Different letters next to the numbers in the table indicate a statistically significant difference between values at $P \leq 0.05$ level whereas the asterisk on the bars in the figures shows a significant difference from the controls type at $P \leq 0.05$ level as determined by two-sided *t*-test.

2.4 Results

2.4.1 *In silico* analysis of GA2ox gene family

A total of 10 putative GA2ox genes were identified using the respective amino acid sequences of the rice GA2ox counterparts as well as the *Arabidopsis* AtGA2ox8 sequence to TBLASTN query the switchgrass expressed sequence tag (EST) databases. The homologous gene variants of the tetraploid switchgrass (2n=4x=36) GA2ox genes were represented by the two subgenomes 'A' or 'B' of switchgrass (Figure 2-1). The phylogenetic analysis confirmed the presence of two discrete groups of putative GA2ox proteins that belong to either C₁₉ or C₂₀ GA classes. The two classes in the clusters each contained four C₂₀ and six C₁₉ switchgrass GA2ox proteins along with their subgenomic variants. MSA analysis showed that switchgrass C₂₀ GA2ox proteins, namely PvGA2ox5a, PvGA2ox5b, PvGA2ox6a, PvGA2ox6b and PvGA2ox9a shared all three unique motifs with GA2ox proteins of *Arabidopsis* (AtGA2ox7 and AtGA2ox8), spinach (SoGA2ox3), poplar (PtGA2ox9 and PtGA2ox10) and rice (OsGA2ox5, OsGA2ox6 and OsGA2ox9) (Figure 2-2). However, the predicted amino acid sequences of PvGA2ox9b were very similar to PvGA2ox9a but the third motif was missing. These motifs were less conserved in PvGA2ox11, OsGA2ox11 and PtGA2ox11 although they were grouped along with the C₂₀ GA2ox proteins in the phylogenetic tree. Moreover, the pattern of amino acid sequences identity among the putative switchgrass C₂₀ GA2ox variants identified in this study showed less

similarity as compared to the identity between their respective rice homologs (Lo et al., 2008) (Table 2-2).

2.4.2 Expression patterns of the switchgrass C₂₀ GA2ox genes

The quantitative reverse transcription polymerase chain reaction (qRT-PCR) results showed that the putative C₂₀ GA2ox genes exhibited differential expression patterns according to the developmental stages of the plant as well as type of plant organs (Figure 2-3). Specifically, the expression of *PvGA2ox5* was high in the seedling and roots whereas it was very low in the other plant organs and stages of development as compared to both *PvGA2ox6* and *PvGA2ox9*. In contrast, *PvGA2ox6* was expressed at very low levels in the seedling stage. Moreover, *PvGA2ox6* and *PvGA2ox9* expression were largely overlapping in almost all plant samples except in the seedling. The level of expression was higher for *PvGA2ox9* in all the plant organs and stages of development examined.

2.4.3 Generation of switchgrass transgenic lines overexpressing *PvGA2ox5* and *PvGA2ox9*

PvGA2ox5b and *PvGA2ox9a* were cloned from cDNA and constitutively overexpressed in switchgrass under the control of maize ubiquitin promoter (Figure 2-4). Fourteen independent transgenic lines overexpressing *PvGA2ox5* and seven *PvGA2ox9*-overexpressing lines were recovered based on visual screening for expression of the *pporRFP* reporter gene and genomic PCR using primers specific to the transgene and the hygromycin resistance gene (Figure 2-4). Moreover, the peculiar dwarf and dark-green broad leaf phenotypes of transgenic lines made the screening process facile.

The observed degree of dwarfism between transgenic lines overexpressing the two genes was different (Figure 2-5). *PvGA2ox5* overexpressing lines showed an array of dwarfism ranging from extremely dwarf to normal/standard plant height as compared to the non-transgenic control, whereas most of the *PvGA2ox9*-overexpressing lines showed similar degree of dwarfism. The relative expression levels of the transgene in *PvGA2ox9*-overexpressing lines determined by qRT-PCR showed a moderate variation ranging from 1-to 4-fold (Figure 2-6). However, both types of transgenics exhibited similar GA-deficient phenotypes such as dwarfism and dark-green broad leaves, and slow initial growth as compared to the non-transgenic parent. Therefore, we

performed detailed analysis only on the *PvGA2ox5*-overexpressing lines, which appeared to be more promising for bioenergy applications.

2.4.4 Overexpression of *PvGA2ox5* on growth phenotypes in switchgrass

There was a range of relative expression among the *PvGA2ox5* transgenic lines, which appeared to correspond with observed patterns of dwarf phenotypes (Figures 2-5, 2-7). Thus, based on the level of transgene expression and degree of dwarfism, two groups of transgenic lines could be identified: the dwarfs (lines 5, 13 and 14) and the semi-dwarfs (all the remaining lines).

During early developmental stages, transgenic lines had reduced shoot growth while the effect on root growth was less remarkable compared to the non-transgenic controls (Figure 2-8a). A stark difference in internode length was also observed between the dwarf and the semi-dwarf transgenic lines (Figure 2-8b). Moreover, the dwarf lines had leafy growth phenotypes with extremely delayed flowering, even after longer vegetative growth.

Additional traits among transgenic plants were also altered, such as tiller height, tiller number and internode length (Table 2-3). In particular, the dwarf transgenic lines showed 87-91% reduction in tiller height and 25-47% reduction in number of tillers per plant relative to the non-transgenic controls. Moreover, these lines exhibited over 96% and 97% reduction in the aboveground fresh and dry biomass, respectively. On the other hand, the semi-dwarf lines displayed a 3-38% reduction in tiller height, but had a remarkable increase in the number of tillers per plant by 25-172% as compared to the non-transgenics. Interestingly, up to 35% and 24% increase in fresh and dry biomass weight, respectively, was observed in semi-dwarf transgenic lines relative to controls. More importantly, the dwarf lines exhibited a 54-83% increase in fresh to dry weight ratios, while the semi-dwarf lines also showed up to 41% increase in fresh to dry weight ratios. Moreover, heterologous overexpression of *PvGA2ox5* in rice similarly caused extreme dwarfism with severely stunted growth (Figure 2-9).

2.4.5 Exogenous application of GA on transgenics

To test whether exogenous application of GA₃ could rescue the slow growth phenotypes in the dwarf lines, clones from three of these lines were treated with a 100 µM GA₃ through foliar spray application. Changes in tiller height were recorded weekly for the first two weeks followed

by a fourth measurement taken two weeks later (Figure 2-10a & b). Foliar spray with 100 μ M GA₃ resulted in the recovery of plants as early as three days after application (Figure 2-10a). Further application of GA₃ spray resulted in a rapid recovery in the transgenic lines in a period of about four weeks (Figure 2-10b). Moreover, the growth of transgenic lines was halted when exogenous GA application stopped (Figure 2-10b).

2.4.6 Effect of *PvGA2ox5* overexpression on lignin content and composition

To investigate whether the overexpression of *PvGA2ox5* could have effect on the lignin in switchgrass, histochemical staining (Figure 2-11) and pyrolysis molecular beam mass spectrometry (py-MBMS) analysis (Figure 2-12) was conducted. Because of differences in the growth stages between the dwarf and the non-transgenic control lines, we considered only the semi-dwarf transgenic lines and non-transgenic plants from the same developmental stages for these analyses. Accordingly, the phloroglucinol-HCl staining for lignin in the leaves from the 3rd internodes of semi-dwarf lines at R1 (reproductive stage 1) developmental stage revealed a relative reduction in lignin staining relative to non-transgenics (Figure 2-11). Correspondingly, a quantitative analysis of lignin content in the semi-dwarf transgenic lines by py-MBMS also showed up to 8% reduction in lignin content compared to non-transgenics (Figure 2-12a). Moreover, analysis of syringyl/guaiacyl (S/G) lignin monomer ratio by py-MBMS also showed up to 23% reduction in transgenic lines overexpressing *PvGA2ox5* as compared to non-transgenic control plants (Figure 2-12b).

2.4.7 Effect of *PvGA2ox5* overexpression on the expression of lignin genes

Semi-dwarf lines had significant reduction in expression of most of the lignin biosynthetic genes including *Pv4CL3*, *PvCCR*, *PvC3H*, *PvC4H*, *PvCAD*, *PvF5H* and *PvHCT* (Figure 2-13). Moreover, there were only minor differences in expression levels of lignin biosynthetic genes between dwarf and semi-dwarf transgenic lines except *PvCCR1* (Figure 2-14).

2.4.8 Effect of *GA2ox* overexpression on the sugar release

Sugar release analysis revealed that there was up to a 15% increase in glucose release from the semi-dwarf transgenic lines as compared to non-transgenics. However, the level of xylose sugar release measured in the same lines showed up to 11% reduction relative to controls (Table 2-4).

The total sugar release (glucose + xylose) did not show a significant increase in the semi-dwarf lines relative to controls.

2.5 Discussion

In this study, we identified a total of ten GA catabolic *GA2ox* genes comprising six *C*₁₉ and four *C*₂₀ GAs. Of the four switchgrass putative *C*₂₀ *GA2ox* proteins, three possessed conserved amino acid sequences at all the three motifs shared with the functionally characterized rice, spinach and *Arabidopsis* *C*₂₀ *GA2ox* proteins (Lo et al., 2008; Lee & Zeevaart, 2005; Schomburg et al., 2003). It has been shown in rice that the C terminal motif is particularly important for *C*₂₀ *GA2ox* protein activity (Lo et al., 2008). Thus, it can be deduced that the switchgrass *GA2ox* proteins play similar roles in the GA catabolic pathway. However, the putative switchgrass *C*₂₀ *GA2ox*, *PvGA2ox11*, along with its closest homolog in rice, *OsGA2ox11*, have divergent sequences at these motifs. Whether these *GA2ox* proteins function in GA catabolism remains to be determined. Moreover, the general pattern and number of *GA2ox* genes present in the switchgrass genome corresponds with the previous report in rice (Lo et al., 2008). Based on this information and the phylogenetic analysis of putative *GA2ox* protein sequence in other monocots, GA catabolism is apparently highly conserved among monocots (Figure 2-15).

The presence of multiple *GA2ox* genes in plants could facilitate differential regulation patterns of gene copies among tissues and organs, which has been documented in rice and poplar (Lo et al., 2008; Gou et al., 2011). The expression pattern analysis of the members of switchgrass *C*₂₀ *GA2ox* genes also indicated the existence of organ-specific differential regulation (Figure 2-3). Specifically, the nearly exclusive expression of *PvGA2ox5* in the seedling stage as well as in roots highlight the role of this gene in early plant development, including tiller formation and root growth. Consistent with this role, rice *GA2ox5* expressed mainly at seedling and early tillering stage was associated with enhanced tillering and adventitious root formation (Lo et al., 2008). Future studies should shed light on the functional diversification of other switchgrass *GA2ox* genes relative to the *C*₂₀ *GA2ox*s investigated in this work.

Overexpression of the two switchgrass *C*₂₀ *GA2ox* genes, i.e., *PvGA2ox5* and *PvGA2ox9* dramatically altered plant architecture resulting in shorter plants with dark-green leaves, extremely reduced internode length, more tillers, and delayed flowering (Figures 2-5, 2-8; Table

2-3) consistent with previous observation from overexpression of *GA2ox* in numerous other plant species (Appleford et al., 2007; Dijkstra et al., 2008; Lo et al., 2008; El-Sharkawy et al., 2012; Lee et al., 2014). Foliar application of exogenous GA (GA_3) reversed these dwarf phenotypes as expected (Figure 2-10) (Agharkar et al., 2007; Dijkstra et al., 2008; Zhao et al., 2010; Bhattacharya et al., 2012) indicating that overexpression of *PvGA2ox5* reduced the level of bioactive GA in switchgrass. Moreover, all the observed phenotypes from the overexpression of switchgrass *GA2ox* genes were consistent with that of GA-deficiency phenotypes suggesting that the transgenes code for GA catabolic genes.

The level of dwarf phenotype in lines overexpressing *PvGA2ox5* and *PvGA2ox9* observed in this study was in line with the previous observation in rice expressing the corresponding homologs (Lo et al., 2008). Interestingly, overexpression of *PvGA2ox5* in rice also resulted in extremely dwarf plants, indicating the conserved functionality between the two plant species and gene orthologues. Moreover, the observed difference in the relative effect on the shoot and root growth in *PvGA2ox5* overexpressing lines is indicative of differential regulation of GA levels in the root and shoot by *PvGA2ox5* (Figure 2-8). Similar observation has been reported in rice where overexpression of *OsGA2ox6* showed a reduced shoot growth but not that in roots (Lo et al., 2008). Therefore, based on these results it could be deduced that *PvGA2ox5* and *PvGA2ox9* genes participate in the deactivation of the C_{20} GA proteins thereby reducing the level of bioactive GA and that these genes may be functional orthologs of the rice *OsGA2ox5* and *OsGA2ox9*, respectively.

Of special note, plant growth was inhibited only when *PvGA2ox5* was highly overexpressed, yielding significantly reduced tiller height (89%) and aboveground fresh (97%) and dry biomass (98%) (Table 2-3). The semi-dwarf lines with 7.5-10-fold lower expression of the transgene than the dwarf lines showed only minor differences in both fresh and dry biomass compared to controls (Table 2-3). There was a tradeoff, in some lines, between tiller number and tiller height (Table 2-3). The expression of C_{20} *GA2ox* genes in rice was previously reported to promote plant tillering possibly via alteration of GA signalling thereby modulating the expression of transcription factors (TFs) such as the *O. sativa* homeobox1 (OSH1) and TEOSINTE BRANCHED1 (TB1) (Lo et al., 2008). Whether the same pathway is used with similar TFs regulating tillering in switchgrass remains to be investigated. Taken together, these observations

further support the assertion that the decreased bioactive GA level in the plant may be from GA2ox-induced GA catabolism as reported in *Arabidopsis*, tobacco, rice and other species (Biemelt et al., 2004; Lee & Zeevaart, 2005; Agharkar et al., 2007; Dijkstra et al., 2008; Lo et al., 2008; Huang et al., 2010; Zhao et al., 2010; El-Sharkawy et al., 2012; Lee et al., 2014). The mechanism behind the decreased bioactive GA level inducing dwarf phenotype, as well as reduced plant biomass, highlight the importance of GA in plant cell elongation and division via elimination of DELLA proteins, the inhibitors of growth promoting factors (Plackett et al., 2011; Digby et al., 1964; Cowling and Harberd, 1999; Asahina et al., 2002; Daviere & Achard, 2013).

Another intriguing observation in *PvGA2ox5* overexpressing lines is the relative increase in fresh to dry biomass weight ratio (Table 2-3) accompanied by reduction in lignin content relative to control plants (Figures 2-11 and 2-12). Similar observations were previously reported in eudicots such as canola (Zhao et al., 2010) and tobacco (Biemelt et al., 2004) that overexpressed *GA2ox* genes. But to our knowledge, there is no report on how these parameters are affected in monocots such as switchgrass. Here, we hypothesised that lignin reduction could result in decreased dry biomass owing to the fact that lignin normally constitutes over 20% of switchgrass dry biomass; thus, the fresh to dry biomass ratio could be increased. Moreover, GA has been shown to directly stimulate lignin accumulation in tobacco petioles (Biemelt et al., 2004). Concurrently, the expression of most lignin biosynthetic genes was shown to be reduced in *PvGA2ox5* overexpressing lines. This suggests that the mechanism behind the reduction in lignin content might be via reduction in bioactive GA content leading to restricted stimulation of lignin accumulation via its role in the regulation of the lignin biosynthesis pathway. Similar results were reported in canola (Zhao et al., 2010) and tobacco (Biemelt et al., 2004). The observed reduction in S/G ratio in the semi-dwarf transgenic lines might indicate the selective repression of the genes responsible for lignin monomer synthesis although the relative expression level of most of the genes responsible for the synthesis of the two lignin monomers were found to be significantly lower. Reduced S/G ratio in switchgrass has been reported to be associated with improved saccharification efficiency and ethanol yield (Fu et al., 2011; Baxter et al., 2014). Moreover, our analysis demonstrated that *GA2ox* overexpressing lines with reduced lignin content has equivalent increase in glucose release efficiency as expected although there is a modest reduction in xylose sugar content. Taken together, these results suggest that manipulation

of *GA2ox* gene expression in switchgrass has potential biotechnological applications in the emerging field of bioenergy.

In summary, the switchgrass *C*₂₀ *GA2ox* genes identified in this work have a tremendous potential for the improvement of bioenergy feedstocks for increased biofuel for the following reasons. First, the improved plant architecture characterized by increased tillering and slightly higher plant biomass in the semi-dwarf lines could suit cultivation of switchgrass on marginal lands by providing protection against soil erosion, lodging, and weed colonization. Second, the reduced biomass recalcitrance followed by improved sugar release efficiency in these lines could tremendously benefit the lignocellulosic biofuel industry. Additionally, it has recently been reported that genetic engineering for reduced GA levels could enhance plant resistance to pathogens (Qin et al., 2013) and high salinity (Shan et al., 2014). Whether reduced GA levels in switchgrass via overexpression of *GA2ox* play similar roles should be the target of future investigations as this added value may enhance the potential use of these lines in future plant breeding and transgene stacking for various bioenergy traits. Moreover, our findings provide an alternative strategy for genetic engineering of food crops such as cereal grains and fruit trees for semi-dwarfism for the following reasons. First, lines with desirable phenotypes could be selected based on the required level of transgene expression and the degree of dwarfism. Second, the semi-dwarf transgenic lines overexpressing these genes have normal floral and seed development, which are the most desirable traits in these crops (Schomburg et al., 2003; Lee & Zeevaart, 2005; Lo et al., 2008; Zhao et al., 2010). Thus, as initially shown by the first Green Revolution, it is clear that GA biosynthesis biotechnology stands tall as a candidate for manipulation to benefit bioenergy and other crop applications.

2.6 Acknowledgements

We thank Crissa Doeppke, Geoff Turner, Steve Decker, Melvin Tucker, and Erica Gjersing for their assistance with lignin and sugar release assays. We also thank Susan Holladay for her assistance with data entry into LIMS. This work was supported by funding from the BioEnergy Science Center. The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE

Office of Science. We also thank Tennessee Agricultural Experiment Station for providing partial financial support for WAW.

References

- Agharkar M, Lomba P, Altpeter F, Zhang H, Kenworthy K and Lange T. 2007. Stable expression of *AtGA2ox1* in a low-input turfgrass (*Paspalum notatum* Flugge) reduces bioactive gibberellin levels and improves turf quality under field conditions. *Plant Biotechnol Journal* 5:791-801.
- Appleford NE, Wilkinson MD, Ma Q, Evans DJ, Stone MC, Pearce SP, Powers SJ, Thomas SG, Jones HD, Phillips AL, Hedden P and Lenton JR. 2007. Decreased shoot stature and grain alpha-amylase activity following ectopic expression of a *gibberellin 2-oxidase* gene in transgenic wheat. *Journal of Experimental Botany* 58:3213-3226.
- Asahina M, Iwai H, Kikuchi A, Yamaguchi S, Kamiya Y, Kamada H and Satoh S. 2002. Gibberellin produced in the cotyledon is required for cell division during tissue reunion in the cortex of cut cucumber and tomato hypocotyls. *Plant Physiology* 129:201-210.
- Bart RS, Chern M, Vega-Sanchez ME, Canlas P and Ronald PC. 2010. Rice *Sn16*, a cinnamoyl-CoA reductase-like gene family member, is required for NH1-mediated immunity to *Xanthomonas oryzae* pv. *oryzae*. *PLoS Genetics* 6: e1001123.
- Baxter HL, Mazarei M, Labbe N, Kline LM, Cheng Q, Windham MT, Mann DGJ, Fu C, Ziebell A, Sykes RW, Rodriguez M, Davis MF, Mielenz JR, Dixon RA, Wang ZY and Stewart CN Jr. 2014. Two-year field analysis of reduced recalcitrance transgenic switchgrass. *Plant Biotechnology Journal* 12:914-924.
- Bhattacharya A, Ward DA, Hedden P, Phillips AL, Power JB, and Davey MR. 2012. Engineering gibberellin metabolism in *Solanum nigrum* L. by ectopic expression of gibberellin oxidase genes. *Plant Cell Reports* 31:945-953.
- Biemelt S, Tschiersch H and Sonnewald U. 2004. Impact of altered gibberellin metabolism on biomass accumulation, lignin biosynthesis, and photosynthesis in transgenic tobacco plants. *Plant Physiology* 135:254-265.

- Burris J, Mann DGJ, Joyce B and Stewart CN Jr. 2009. An improved tissue culture system for embryogenic callus production and plant regeneration in switchgrass (*Panicum virgatum* L.). *BioEnergy Research* 2:267-274.
- Busov VB, Meilan R, Pearce DW, Ma C, Rood SB and Strauss SH. 2003. Activation tagging of a dominant gibberellin catabolism gene (*GA 2-oxidase*) from poplar that regulates tree stature. *Plant Physiology* 132:1283-1291.
- Cowling RJ and Harberd NP. 1999. Gibberellins control *Arabidopsis* hypocotyl growth via regulation of cellular elongation. *Journal of Experimental Botany* 50:1351-1357.
- Daviere JM and Achard P. 2013. Gibberellin signalling in plants. *Development*, 140:1147-1151.
- Decker SR, Carlile M, Selig MJ, Doeppke C, Davis M, Sykes RW, Turner G, and Ziebell A. 2012. Reducing the effect of variable starch levels in biomass recalcitrance screening. *Methods in Molecular Biology* 908:181-195.
- Digby J, Thomas TH, and Wareing PF. 1964. Promotion of cell division in tissue cultures by gibberellic acid. *Nature* 203:547-548.
- Dijkstra C, Adams E, Bhattacharya A, Page AF, Anthony P, Kourmpetli S, Power JB, Lowe KC, Thomas SG, Hedden P, Phillips AL and Davey MR. 2008. Over-expression of a *gibberellin 2-oxidase* gene from *Phaseolus coccineus* L. enhances gibberellin inactivation and induces dwarfism in *Solanum* species. *Plant Cell Reports* 27:463-470.
- El-Sharkawy I, El Kayal W, Prasath D, Fernandez H, Bouzayen M, Svircev AM and Jayasankar S. 2012. Identification and genetic characterization of a *gibberellin 2-oxidase* gene that controls tree stature and reproductive growth in plum. *Journal of Experimental Botany* 63:1225-1239.
- Fu C, Mielenz JR, Xiao X, Ge Y, Hamilton CY, Rodriguez Jr. M, Chen F, Foston M, Ragauskas A, Bouton J, Dixon RA and Wang ZY. 2011. Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proceedings of National Academy of Science USA* 108:3803-3808.

- Gou J, Ma C, Kadmiel M, Gai Y, Strauss S, Jiang X and Busov V. 2011. Tissue-specific expression of *Populus* C₁₉ GA 2-oxidases differentially regulate above- and below-ground biomass growth through control of bioactive GA concentrations. *New Phytologist* 192:626-639.
- Grewal D, Gill R and Gosal SS. 2006. Influence of antibiotic cefotaxime on somatic embryogenesis and plant regeneration in indica rice. *Biotechnology Journal* 1:1158-1162.
- Harberd NP, King KE, Carol P, Cowling RJ, Peng J, and Richards DE. 1998. Gibberellin: inhibitor of an inhibitor of...? *BioEssays* 20:1001-1008.
- Hedden P and Phillips AL. 2000. Gibberellin metabolism: new insights revealed by the genes. *Trends in Plant Science* 5:523-530.
- Hedden P and Thomas SG. 2012. Gibberellin biosynthesis and its regulation. *Biochemical Journal* 444:11-25.
- Helliwell CA, Chandler PM, Poole A, Dennis ES and Peacock WJ. 2001. The CYP88A cytochrome P450, ent-kaurenoic acid oxidase, catalyzes three steps of the gibberellin biosynthesis pathway. *Proceedings of Natlional Academy of Science USA* 98:2065-2070.
- Huang J, Tang D, Shen Y, Qin B, Hong L, You A, Li M, Wang X, Yu H, Gu M and Cheng Z. 2010. Activation of *gibberellin 2-oxidase 6* decreases active gibberellin levels and creates a dominant semi-dwarf phenotype in rice (*Oryza sativa* L.). *Journal of Genetics and Genomics* 37:23-36.
- Jakob K, Zhou F and Paterson A. 2009. Genetic improvement of C4 grasses as cellulosic biofuel feedstocks. *In Vitro Cellular & Developmental Biology Plant* 45:291-305.
- Lee D, Lee I, Kim K, Kim D, Na H, Lee I-J, Kang S-M, Jeon H-W, Le P and Ko J-H. 2014. Expression of *gibberellin 2-oxidase 4* from *Arabidopsis* under the control of a senescence-associated promoter results in a dominant semi-dwarf plant with normal flowering. *Journal of Plant Biology* 57:106-116.

- Lee DJ and Zeevaart JA. 2005. Molecular cloning of *GA 2-oxidase3* from spinach and its ectopic expression in *Nicotiana sylvestris*. *Plant Physiology* 138:243-254.
- Lester DR, Ross JJ, Smith JJ, Elliott RC and Reid JB. 1999. Gibberellin 2-oxidation and the *SLN* gene of *Pisum sativum*. *Plant Journal* 19:65-73.
- Li R and Qu R. 2011. High throughput *Agrobacterium*-mediated switchgrass transformation. *Biomass & Bioenergy* 35:1046-1054.
- Lo SF, Yang SY, Chen KT, Hsing YI, Zeevaart JA, Chen LJ and Yu SM. 2008. A novel class of gibberellin 2-oxidases control semidwarfism, tillering, and root development in rice. *Plant Cell* 20:2603-2618.
- Mann DGJ, Lafayette PR, Abercrombie LL, King ZR, Mazarei M, Halter MC, Poovaiah CR, Baxter H, Shen H, Dixon RA, Parrott WA and Stewart CN Jr. 2012. Gateway-compatible vectors for high-throughput gene functional analysis in switchgrass (*Panicum virgatum* L.) and other monocot species. *Plant Biotechnology Journal* 10:226-236.
- Martin DN, Proebsting WM and Hedden P. 1999. The *SLENDER* gene of pea encodes a gibberellin 2-oxidase. *Plant Physiology* 121:775-781.
- Moore KJ, Moser LE, Vogel KP, Waller SS, Johnson BE and Pedersen JF. 1991. Describing and quantifying growth stages of perennial forage grasses. *Agronomy Journal* 83:1073-1077.
- Murashige T and Skoog F. 1962. A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiologia Plantarum* 15:473-497.
- Nishimura A, Aichi I and Matsuoka M. 2006. A protocol for *Agrobacterium*-mediated transformation in rice. *Nature Protocols* 1:2796-2802.
- Olszewski N, Sun TP and Gubler F. 2002. Gibberellin signalling: biosynthesis, catabolism, and response pathways. *Plant Cell* 14 (Suppl.):S61-80.

Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD and Harberd NP. 1999. 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* 400:256-261.

Peters RJ. 2012. Gibberellin phytohormone metabolism. In: Bach T and Rohmer M, (eds.) *Isoprenoid synthesis in plants and microorganisms: new concepts and experimental approaches*. Springer, New York, 233-249.

Plackett ARG, Thomas SG, Wilson ZA and Hedden P. 2011. Gibberellin control of stamen development: a fertile field. *Trends in Plant Science* 16:568-578.

Qin X, Liu JH, Zhao WS, Chen XJ, Guo ZJ and Peng YL. 2013. *Gibberellin 20-oxidase* gene *OsGA20ox3* regulates plant stature and disease development in rice. *Molecular Plant-Microbe Interaction* 26:227-239.

Radi A, Lange T, Niki T, Koshioka M and Lange MJ. 2006. Ectopic expression of pumpkin *gibberellin oxidases* alters gibberellin biosynthesis and development of transgenic *Arabidopsis* plants. *Plant Physiology* 140:528-536.

Rieu I, Eriksson S, Powers SJ, Gong F, Griffiths J, Woolley L, Benlloch R, Nilsson O, Thomas, SG, Hedden P and Phillips AL. 2008. Genetic analysis reveals that C₁₉-GA 2-oxidation is a major gibberellin inactivation pathway in *Arabidopsis*. *Plant Cell* 20:2420-2436.

Rieu I, Ruiz-Rivero O, Fernandez-Garcia N, Griffiths J, Powers SJ, Gong F, Linhartova T, Eriksson S, Nilsson O, Thomas SG, Phillips AL and Hedden P. 2008. The gibberellin biosynthetic genes *AtGA20ox1* and *AtGA20ox2* act, partially redundantly, to promote growth and development throughout the *Arabidopsis* life cycle. *Plant Journal* 53:488-504.

Sakamoto T, Kobayashi M, Itoh H, Tagiri A, Kayano T, Tanaka H, Iwahori S and Matsuoka M. 2001. Expression of a *gibberellin 2-oxidase* gene around the shoot apex is related to phase transition in rice. *Plant Physiology* 125:1508-1516.

Sakamoto T, Morinaka Y, Ishiyama K, Kobayashi M, Itoh H, Kayano T, Iwahori S, Matsuoka M and Tanaka H. 2003. Genetic manipulation of gibberellin metabolism in transgenic rice. *Nature Biotechnology* 21:909-913.

Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, Swapan D, Ishiyama K, Saito T, Kobayashi M, Khush GS, Kitano H and Matsuoka M. 2002. Green revolution: a mutant gibberellin-synthesis gene in rice. *Nature* 416:701-702.

Schomburg FM, Bizzell CM, Lee DJ, Zeveaart JA and Amasino RM. 2003. Overexpression of a novel class of gibberellin 2-oxidases decreases gibberellin levels and creates dwarf plants. *Plant Cell* 15:151-163.

Schwechheimer C and Willige BC. 2009. Shedding light on gibberellic acid signalling. *Current Opinion in Plant Biology* 12:57-62.

Shan C, Mei Z, Duan J, Chen H, Feng H and Cai W. 2014. OsGA2ox5, a gibberellin metabolism enzyme, is involved in plant growth, the root gravity response and salt stress. *PLoS One* 9:e87110.

Shen H, Fu C, Xiao X, Ray T, Tang Y, Wang ZY and Chen F. 2009. Developmental control of lignification in stems of lowland switchgrass variety Alamo and the effects on saccharification efficiency. *BioEnergy Research* 2:233-245.

Shen H, He X, Poovaiah CR, Wuddineh WA, Ma J, Mann, DGJ, Wang H, Jackson L, Tang Y, Stewart CN Jr., Chen F and Dixon RA. 2012. Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks. *New Phytologist* 193:121-136.

Shen H, Poovaiah CR, Ziebell A, Tschaplinski TJ, Pattathil S, Gjersing E, Engle NL, Katahira R, Pu Y, Sykes RW, Chen F, Ragauskas AJ, Mielenz JR, Hahn MG, Davis M, Stewart CN Jr. and Dixon RA. 2013. Enhanced characteristics of genetically modified switchgrass (*Panicum virgatum* L.) for high biofuel production. *Biotechnology for Biofuels* 6:71.

- Solfanelli C, Ceron F, Paolicchi F, Giorgetti L, Geri C, Ceccarelli N, Kamiya Y and Picciarelli P. 2005. Expression of two genes encoding gibberellin 2- and 3-oxidases in developing seeds of *Phaseolus coccineus*. *Plant and Cell Physiology* 46:1116-1124.
- Stamm P, Verma V, Ramamoorthy R and Kumar PP. 2012. Manipulation of plant architecture to enhance lignocellulosic biomass. *AoB Plants* pls026.
- Studer MH, DeMartini JD, Brethauer S, McKenzie HL and Wyman CE. 2010. Engineering of a high-throughput screening system to identify cellulosic biomass, pretreatments, and enzyme formulations that enhance sugar release. *Biotechnology and Bioengineering* 105:231-238.
- Sun TP and Gubler F. 2004. Molecular mechanism of gibberellin signalling in plants. *Annual Review of Plant Biology* 55:197-223.
- Sykes RW, Yung M, Novaes E, Kirst M, Peter G and Davis M. 2009. High-throughput screening of plant cell-wall composition using pyrolysis molecular beam mass spectroscopy. *Methods in Molecular Biology* 581:169-183.
- Thomas SG, Phillips AL and Hedden P. 1999. Molecular cloning and functional expression of *gibberellin 2- oxidases*, multifunctional enzymes involved in gibberellin deactivation. *Proceedings of National Academy of Science USA* 96:4698-4703.
- Xu B, Escamilla-Trevino LL, Sathitsuksanoh N, Shen Z, Shen H, Zhang YH, Dixon RA and Zhao B. 2011. Silencing of *4-coumarate:coenzyme A ligase* in switchgrass leads to reduced lignin content and improved fermentable sugar yields for biofuel production. *New Phytologist* 192:611-625.
- Yamaguchi S. 2008. Gibberellin metabolism and its regulation. *Annual Review of Plant Biology* 59:225-251.
- Yuan JS, Tiller KH, Al-Ahmad H, Stewart NR and Stewart CN Jr. 2008. Plants to power: bioenergy to fuel the future. *Trends in Plant Science* 13:421-429.

Zhao XY, Zhu DF, Zhou B, Peng WS, Lin JZ, Huang XQ, He RQ, Zhuo YH, Peng D, Tang DY, Li MF and Liu XM. 2010. Over-expression of the *AtGA2ox8* gene decreases the biomass accumulation and lignification in rapeseed (*Brassica napus* L.). *Journal of Zhejiang University Science B* 11:471-481.

Appendix

Figures and tables

Figure 2-1 Cluster tree of the C₁₉ and C₂₀ GA2ox proteins of monocots; switchgrass (*Panicum virgatum*) and rice (*Oryza sativa*) and C₂₀ GA2ox proteins of eudicots; *Arabidopsis thaliana*, spinach (*Spinacia oleracea*) and poplar (*Populus trichocarpa*) based on the deduced amino acid sequences. The sequences were aligned using MUSCLE program (Edgar, 2004) and the alignment was curated by Gblocks at the phylogeny.fr (<http://www.phylogeny.fr>) to eliminate poorly aligned positions and divergent regions. The tree was constructed by a neighbour joining method using MEGA6.0 program. Analysis using 1000 bootstrap replicates was performed. The numbers on branches indicate bootstrap support values. The scale bar shows 0.2 amino acid substitutions per site. The locus names of the switchgrass sequences and GenBank accession numbers of the sequences used in this tree are listed in Table 2-5.

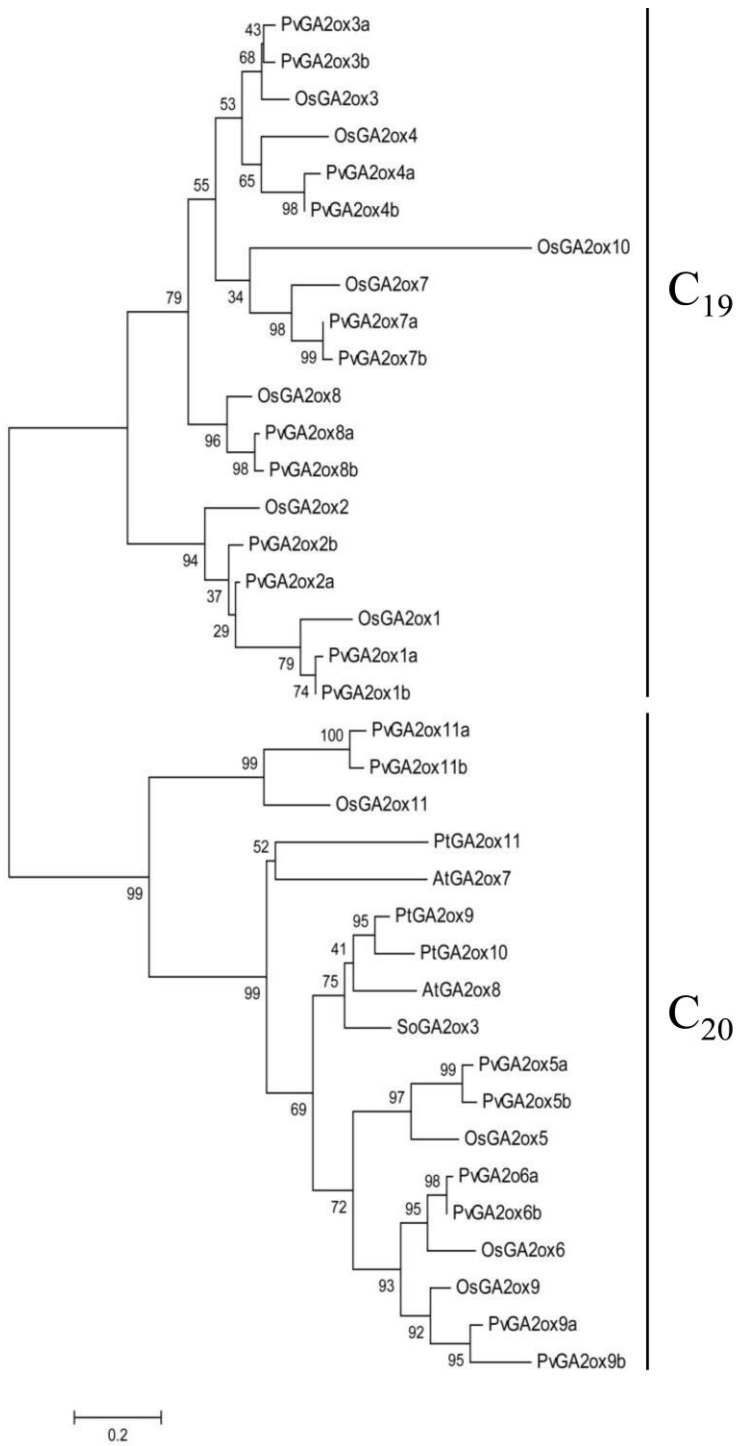


Figure 2-1 Continued.

Figure 2-2 Multiple amino acid sequence alignment of switchgrass C₂₀ GA2ox proteins and their closest homologs along with two C₁₉ GA2ox proteins. The multiple sequence alignment was constructed using the amino acid sequences of C₂₀ GA2ox proteins from switchgrass, rice and *Arabidopsis* as well as two C₁₉ switchgrass GA2ox proteins by MUSCLE program (Edgar, 2004). The three conserved motifs that are unique to the C₂₀ GA2ox proteins are marked by the rectangles.

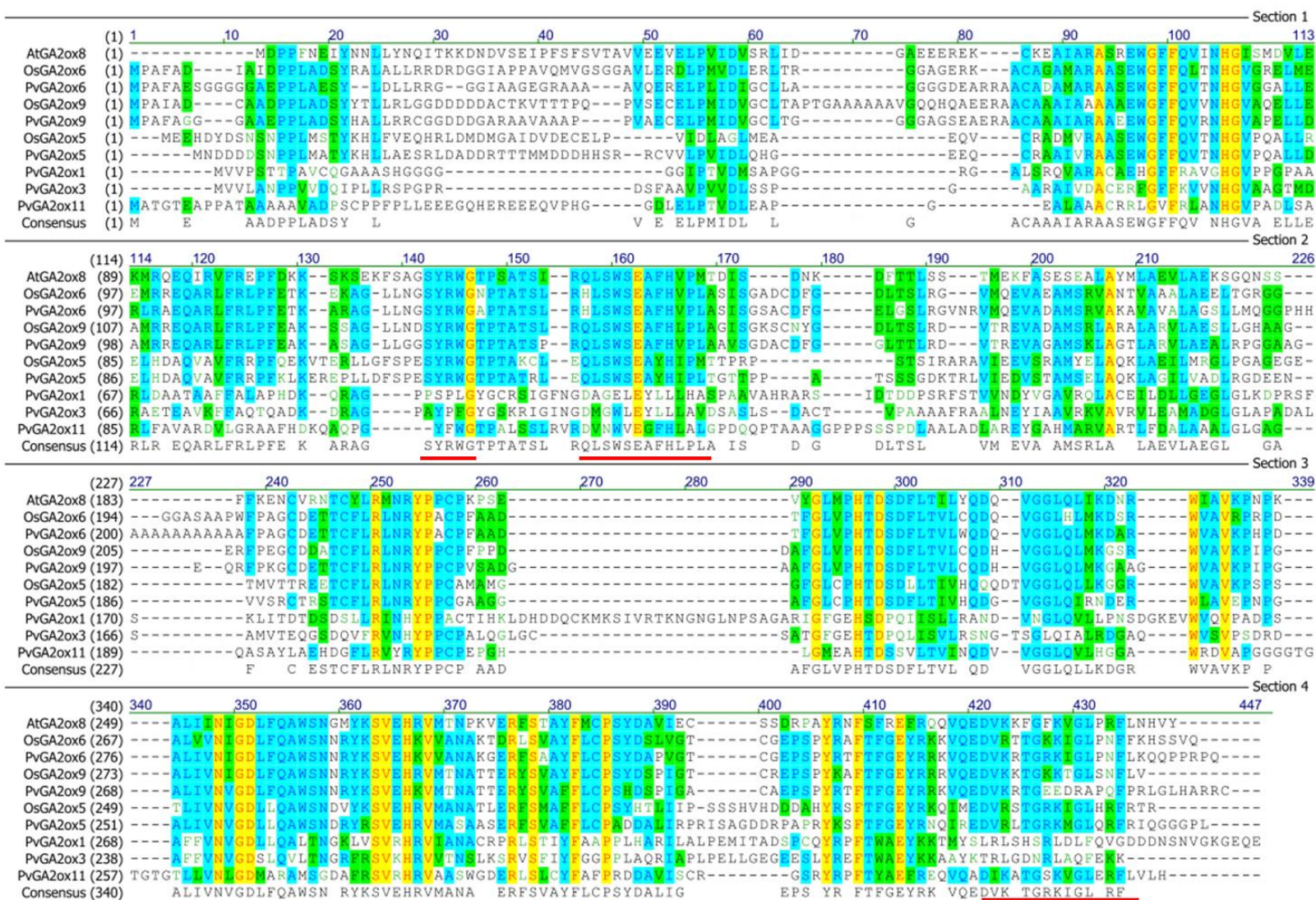


Figure 2-2 Continued.

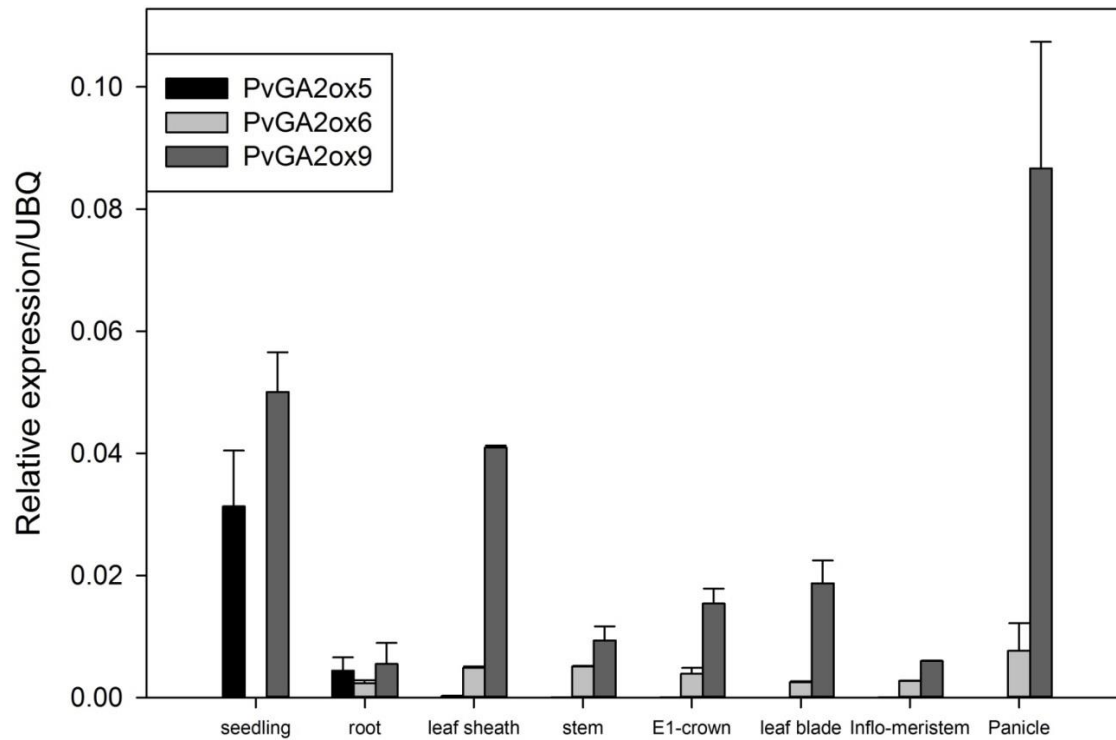


Figure 2-3 Expression patterns of the three putative C_{20} GA2ox genes among tissues and developmental stages in switchgrass as determined by qRT-PCR. The leaf sheath, stem, leaf blade, and panicle samples collected from R1 (reproductive stage 1) tillers, samples of inflorescence meristem from E5 (elongation stage with five internode) stage tillers, 2 wk old seedlings and E1 (elongation stage with one internode) crown were used to get the RNA for qRT-PCR. The dissociation curve for the qRT-PCR products showed that the primers were gene-specific. The relative levels of transcripts were normalized to UBQ. Bars represent mean values of 3 replicates $\pm$ standard error.

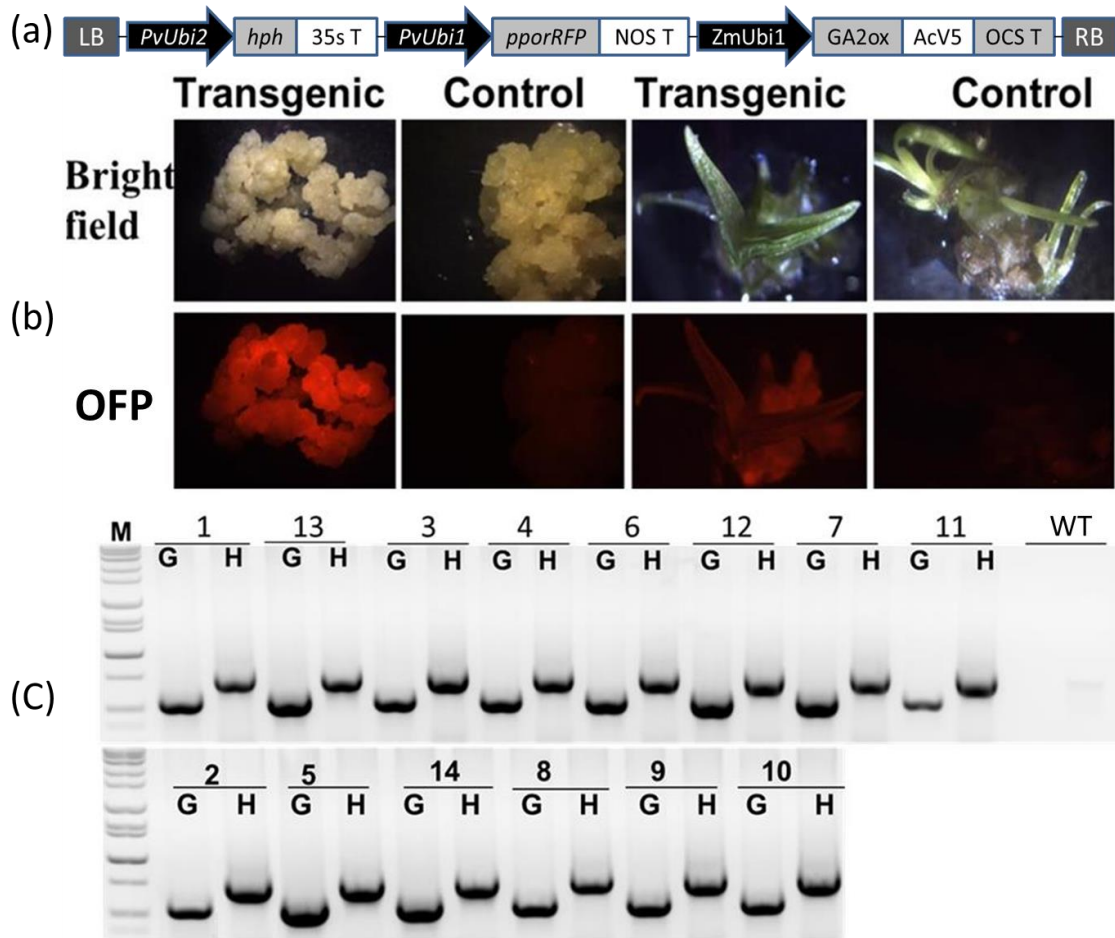


Figure 2-4 Molecular characterization of transgenic switchgrass plants overexpressing the *PvGA2ox5* gene. (a) pANIC10A vector construct used for overexpression of *PvGA2ox*. (b) Orange fluorescence protein (*pporRFP*; OFP) visualization in transgenic callus and plants compared to the non-transgenic control. (c) Genomic PCR confirming the insertion of the transgene (G) and the hygromycin-resistance (H) genes in transgenic lines. No amplification was observed in DNA samples from the non-transgenic (WT) plants.



Transgenic



Transgenic

Figure 2-5 Representative *PvGA2ox5* and *PvGA2ox9* overexpressing transgenic lines showing dwarf phenotypes compared to non-transgenic controls (WT).

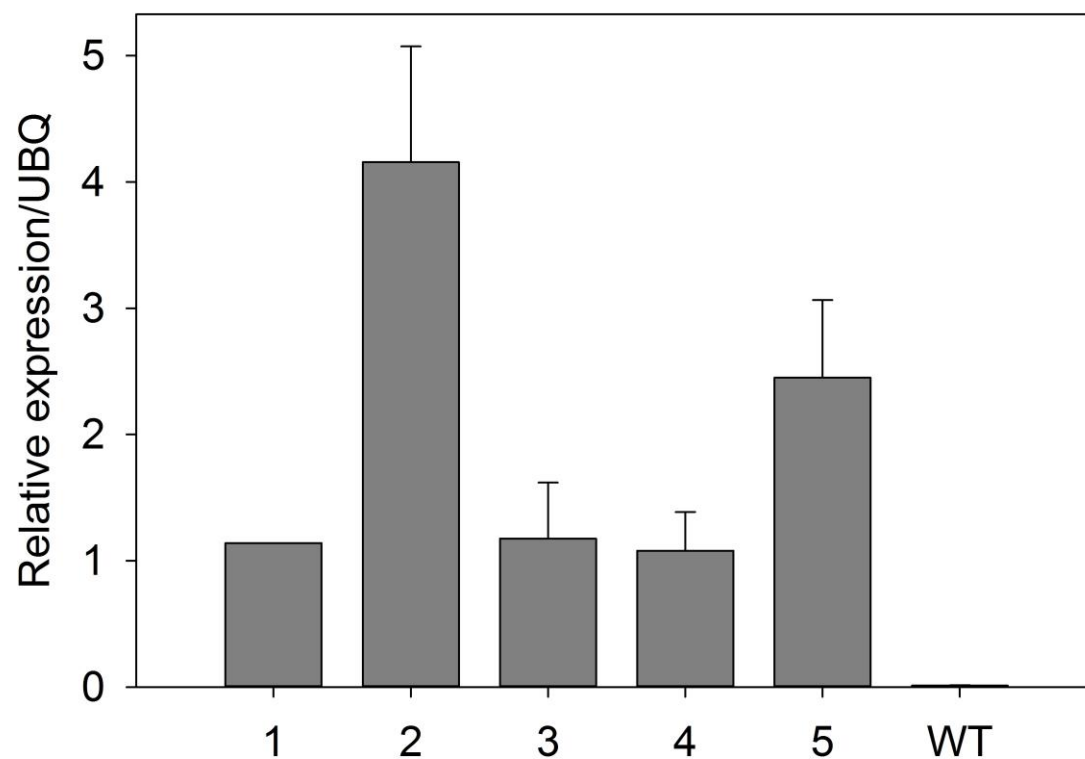


Figure 2-6 qRT-PCR analysis of *PvGA2ox9* overexpressing transgenic (1-5) and non-transgenic controls (WT) lines from RNA isolated from the top internode of each plant. The relative levels of transcripts were normalized to UBQ. Bars represent mean values of 3 biological replicates $\pm$ standard error.

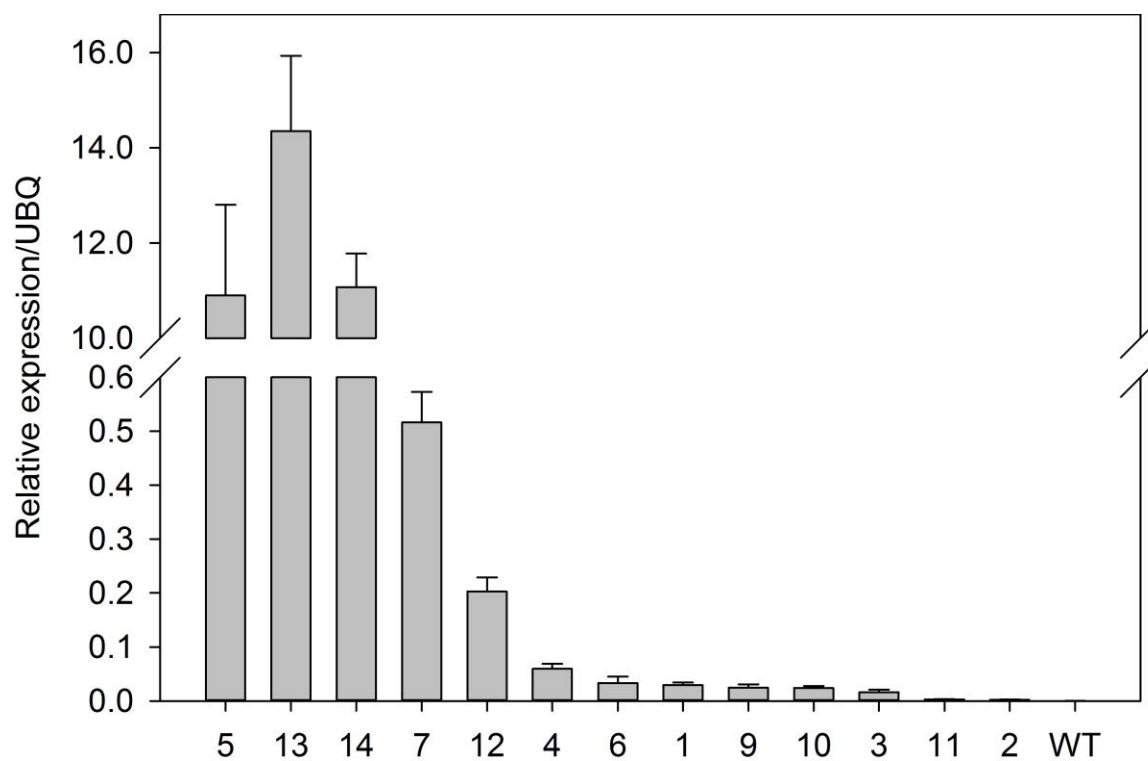


Figure 2-7 qRT-PCR analysis of *PvGA2ox5*-overexpressing transgenic (1-14) and non-transgenic (WT) lines from RNA isolated from the top internode of each plant. The relative levels of transcripts were normalized to UBQ. Bars represent mean values of 3 biological replicates $\pm$ standard error.

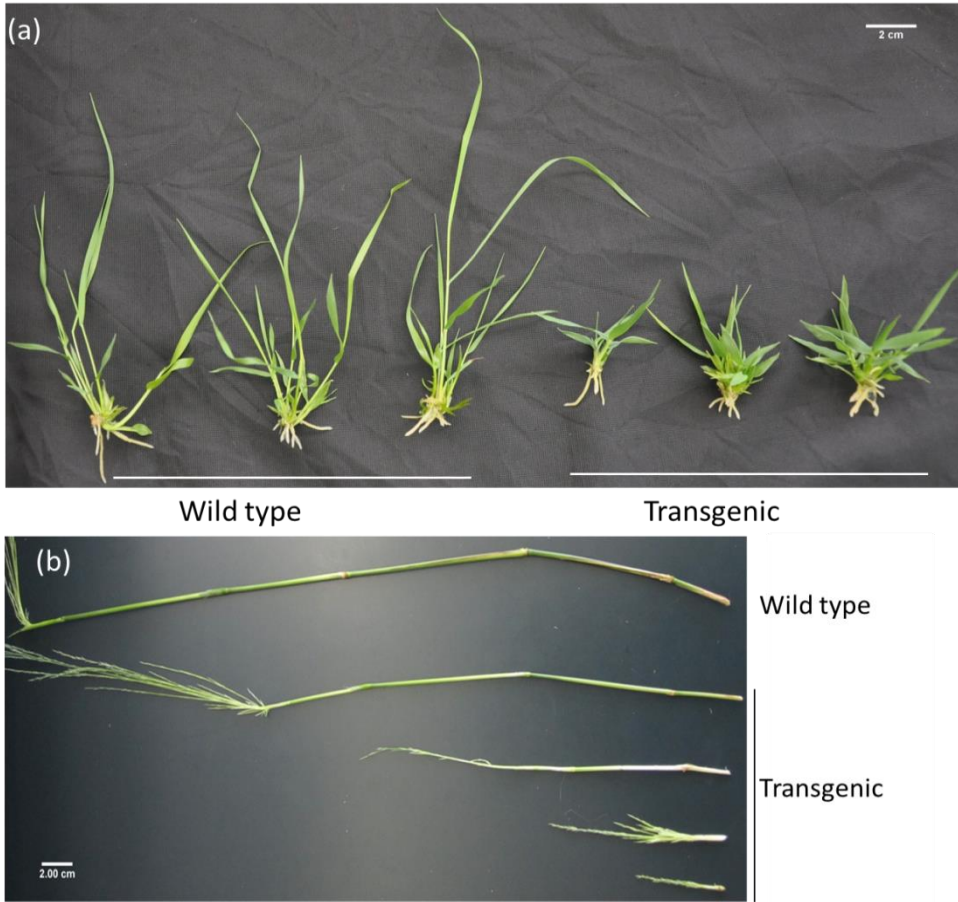


Figure 2-8 Morphology of plants overexpressing *PvGA2ox5* showing a stunted shoot growth but with less apparent reduction in root growth as compared to the non-transgenic control plant during early stages of development (a). Internode lengths are variably reduced in transgenic lines (b).

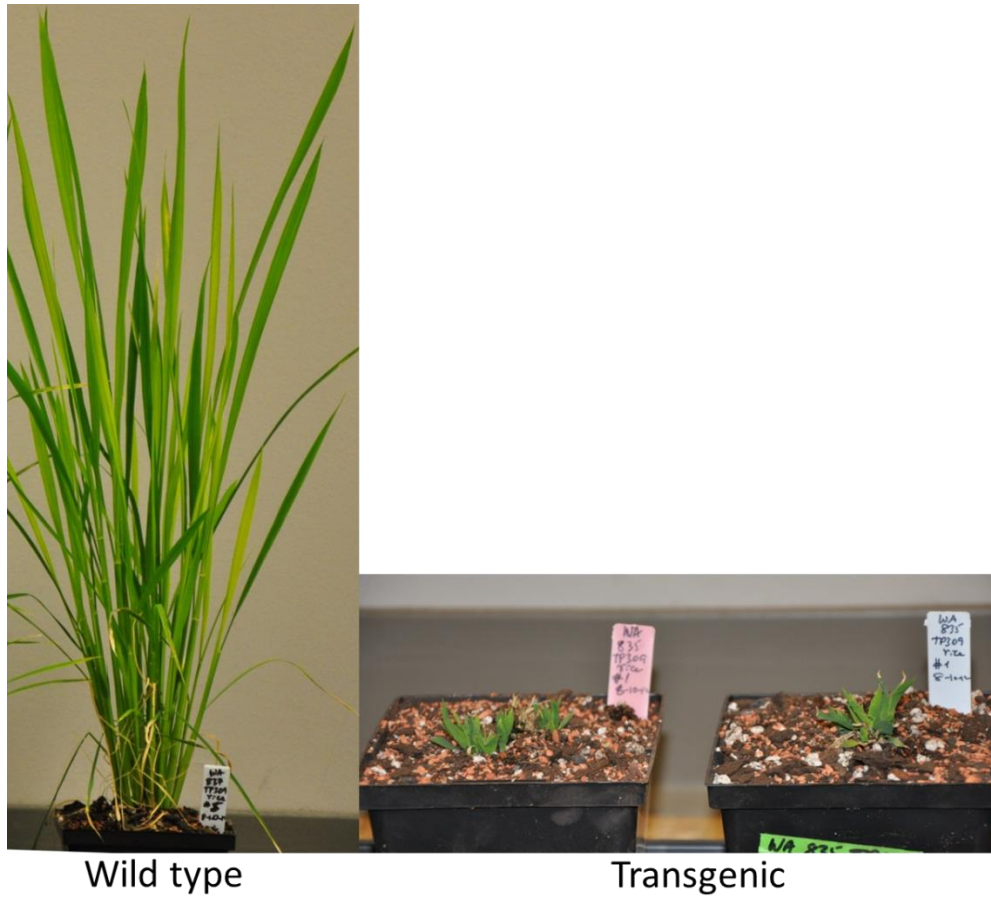


Figure 2-9 *PvGA2ox5* overexpressing rice plants showing extremely dwarf phenotypes as compared to the wild type control. Both the transgenic and non-transgenic wild type rice were regenerated from callus under same conditions and the plants were photographed two months after shoot regeneration.

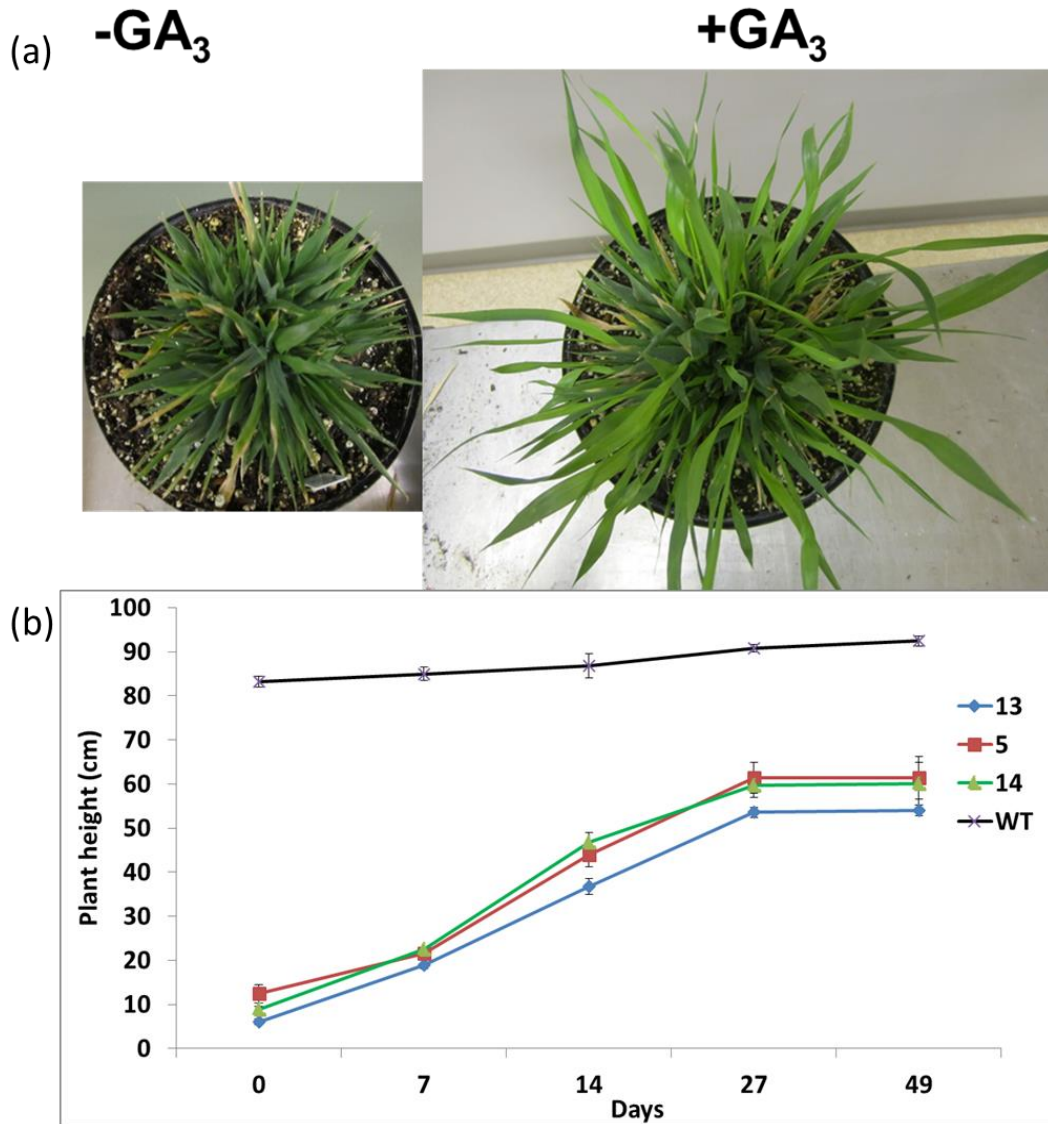


Figure 2-10 Foliar spray with 100 μ M GA₃ application restored normal growth in transgenic lines. Extremely dwarf transgenic plants without treatment with GA₃ (-) and after treatment with 100 μ M GA₃ (+) (a). Tiller heights of the transgenic and non-transgenic (WT) plants measured at 0, 7, 14, 27 and 49 days interval with the transgenic plants sprayed three times with 100 μ M GA₃ while the WT was sprayed only with water (b). The error bars represent the standard error ($n \geq 3$).

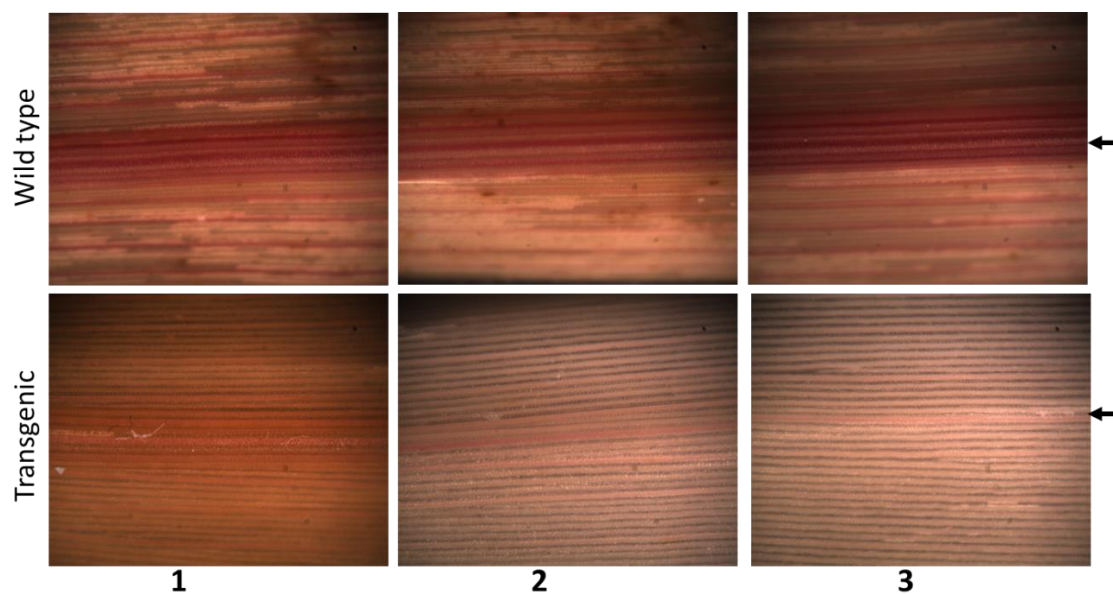


Figure 2-11 Histochemical detection of lignin in leaves of *PvGA2ox5* overexpressing and wild type lines in light microscopy. The phloroglucinol-HCl staining was done on leaves from three independent tillers indicated by numbers (1-3) as shown in the figure. The images were taken at 2x magnification. The arrows point to leaf midribs.

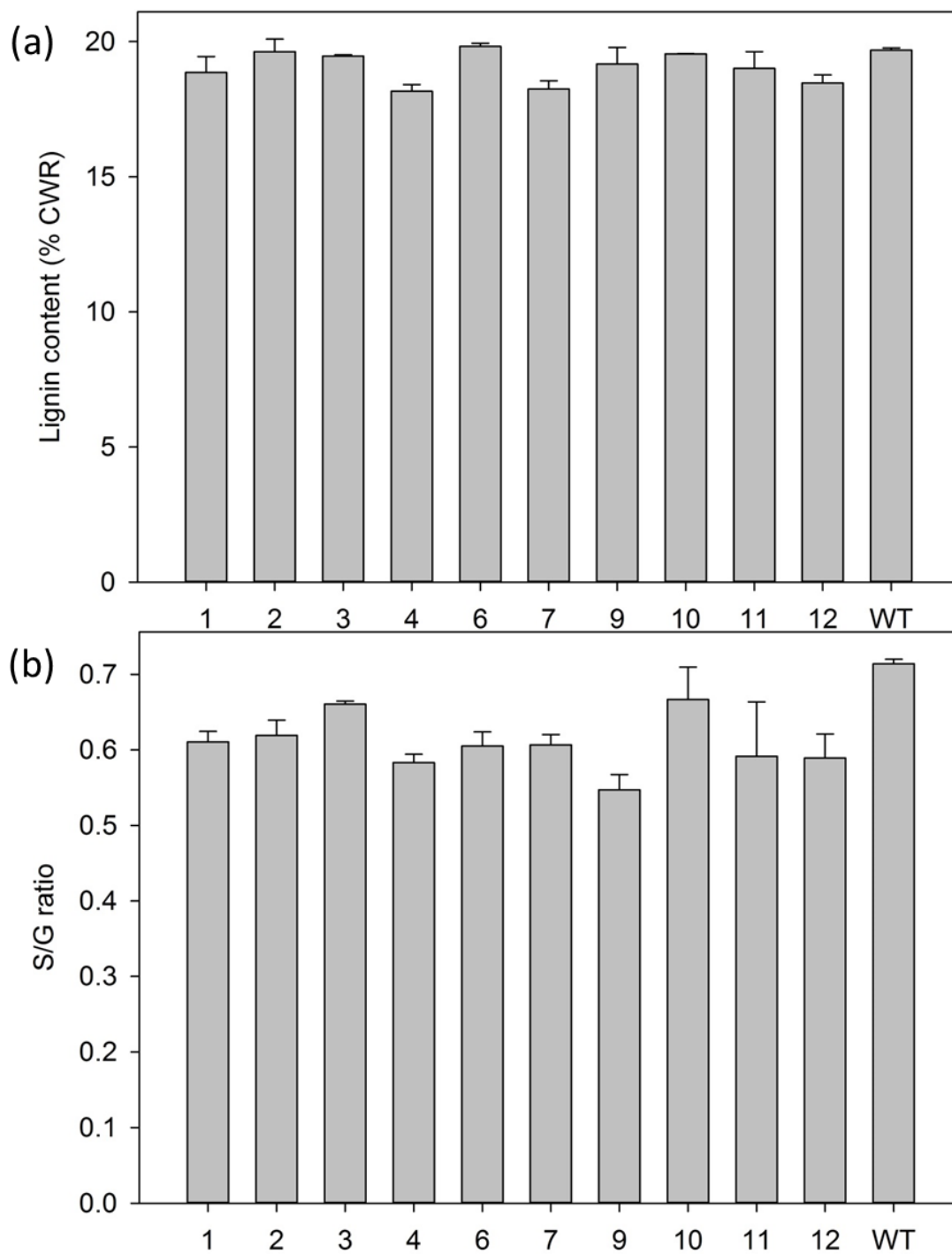


Figure 2-12 Lignin content (a) and S/G ratio (b) of transgenic and non-transgenic (WT) lines as determined via py-MBMS. Bars represent the average of the replicates $\pm$ standard error.

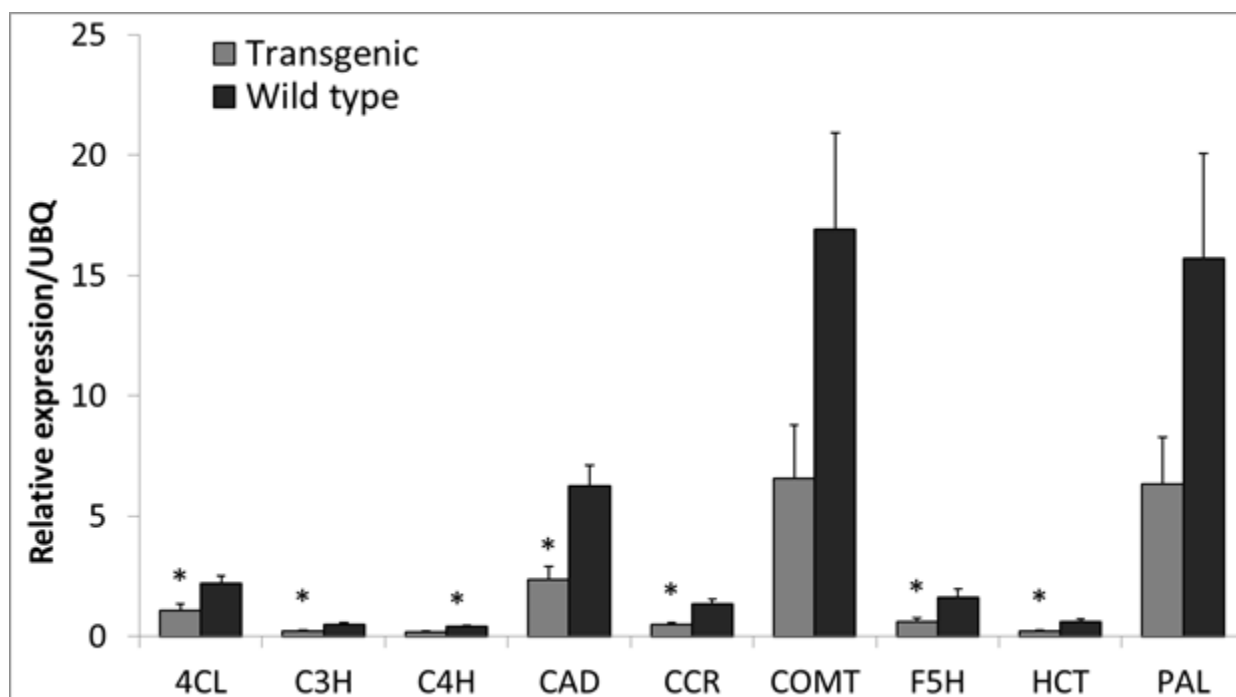


Figure 2-13 Relative expression of lignin biosynthetic genes in transgenic and non-transgenic lines as determined by qRT-PCR. The relative levels of transcripts were normalized to UBQ. Asterisks indicate significant differences from non-transgenic control plants at $P \leq 0.05$. *Pv4CL* (4-coumarate: CoA ligase); *PvC3H* (coumaroyl shikimate 3-hydroxylase); *PvC4H* (coumaroyl shikimate 4-hydroxylase); *PvCAD* (cinnamyl alcohol dehydrogenase); *PvCCR* (cinnamoyl CoA reductase); *PvCOMT* (caffeic acid 3-O-methyltransferase); *PvF5H* (ferulate 5-hydroxylase); *PvHCT* (hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase); *PvPAL* (phenylalanine ammonia-lyase).

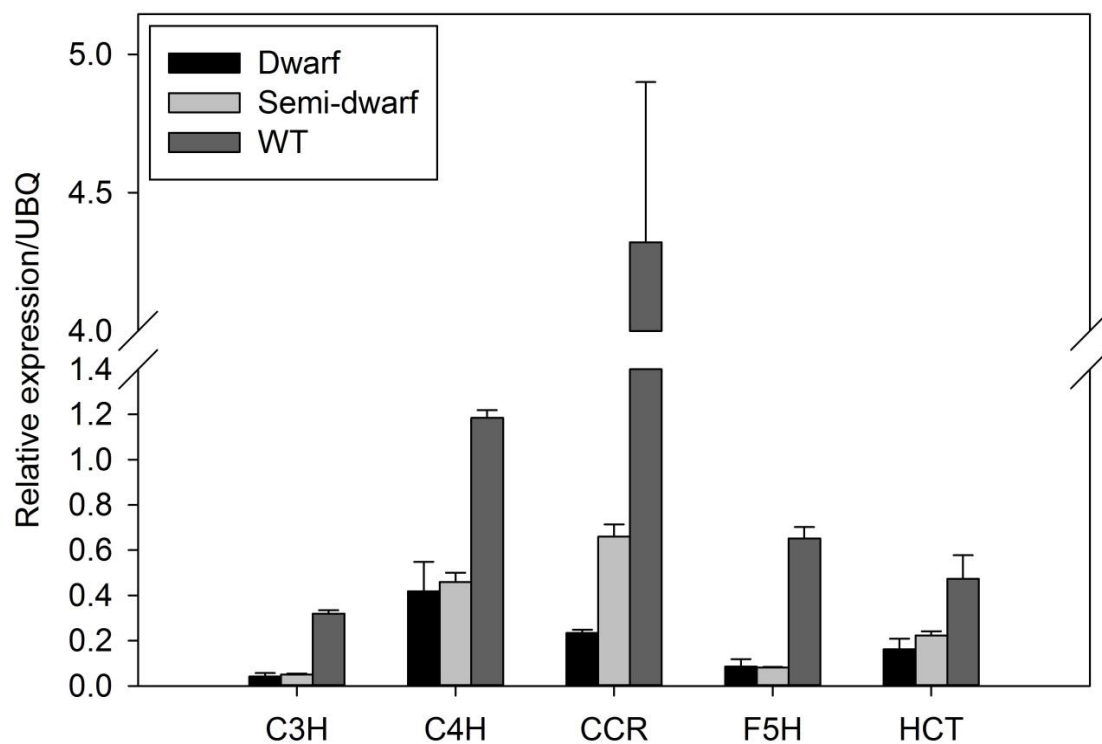


Figure 2-14 The relative expression of lignin biosynthetic genes in transgenic dwarf, semi-dwarf and non-transgenic (WT) lines as determined by qRT-PCR. The relative levels of transcripts were normalized to UBQ. Data points are mean values of 3 biological replicates $\pm$ standard error.

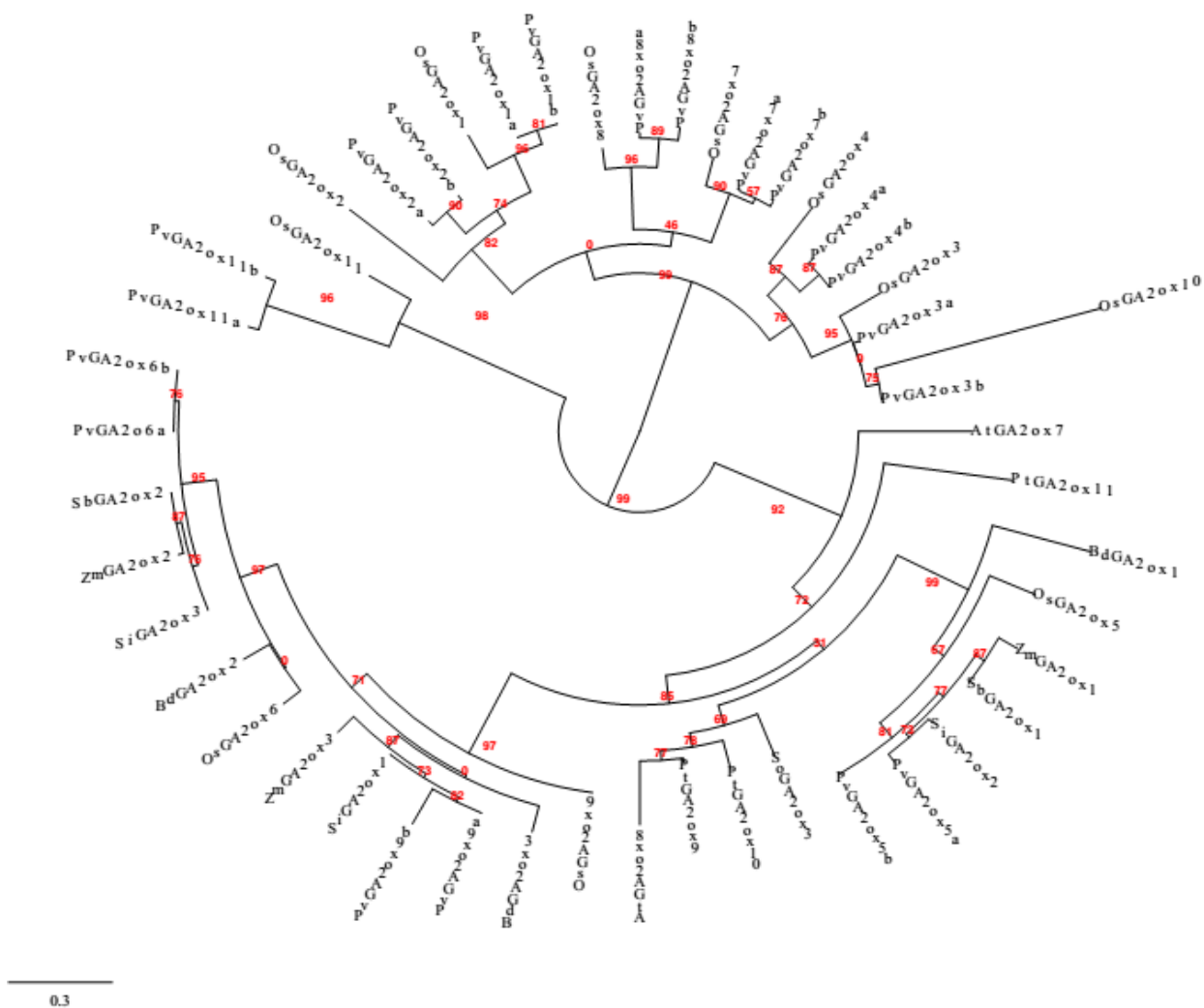


Figure 2-15 Cluster analysis of putative GA2ox proteins from monocots: switchgrass (*Panicum virgatum*), rice (*Oryza sativa*), sorghum (*Sorghum bicolor*), maize (*Zea mays*), foxtail millet (*Setaria italica*) and *Brachypodium distachyon*) and dicots: *Arabidopsis*, poplar (*Populus trichocarpa*) and spinach (*Spinacia oleracea*). The tree was constructed by the programs at phylogeny.fr (<http://www.phylogeny.fr>). Numbers on branches indicate bootstrap support values. The scale bar shows 0.3 amino acid substitutions per site. The GenBank accession numbers of the sequences used in this tree are listed in Table 2-5.

Table 2-1 List of primers used in this study.

Primer name	Primer sequence	Reference
Primers for genomic DNA PCR		
Hygro_F	TTGCATCTCCCGCCGTTTCACAG	
Hygro_R	CTGGGGCGTCGGTTTCCACTAT	
835_F	ACAGCGACTTCCTGACCATCCT	
835_R	CGATCATAGGCGTCTCGCATATCT	
Gene-specific primers for qRT-PCR		
<u>Primers for confirming the overexpression of <i>PvGA2ox5</i></u>		
F_835II	GCAACGCTTCCGGATCCAAG	
R_AcV5	ACCAGCCGCTCGCATCTTTC	
qRT-PCR primers for expression analysis of switchgrass C₂₀ GA2ox genes		
Primers for expression of <i>PvGA2ox5</i>		
PvGA2ox_F3	CCTACCACATCCCCTTGACC	
PvGA2ox_R3	GTGGACACGTCCTCGATCAC	
Primers for expression of <i>PvGA2ox6</i>		
35270F2	CAAGAGCGTGGAGCACAAGG	
35270R2	CGAAGGTGAAGGTCCTGTAGG	
Primers for expression of <i>PvGA2ox9</i>		
KanlCTG23388F2	CAGACGGAAGGTGCAGGAAGAC	
KanlCTG23388R2	AGGATGGGTTTGTGGGAAGGA	

Table 2-1 Continued.

Primer name	Unitranscript ID	Primer sequence	Reference
Primers for reference gene			
F_PvUBIQUITIN	AP13CTG25905	CAGCGAGGGCTCAATAATTCCA	Xu et al., 2011
R_PvUBIQUITIN		TCTGGCGGACTACAATATCCA	
Primers for the expression of lignin biosynthesis genes			
C4H1_1534F	AP13CTG28733	GGGCAGTTCAGCAACCAGAT	Shen et al., 2012
C4H1_1611R		CGCGTTTCCGGGACTCTAG	
PvCOMT_F461	KarlCTG02872	CAACCGCGTGTTCAACGA	Shen et al., 2012
PvCOMT_R534		CGGTGTAGAACTCGAGCAGCTT	
4CL1_1179_F	AP13CTG06049	CGAGCAGATCATGAAAGGTTACC	Shen et al., 2012
4CL1_1251_R		CAGCCAGCCGTCCTTGTC	
PvCCR1. 112_F	AP13ISTG52570	GCGTCGTGGCTCGTCAA	Shen et al., 2012
PvCCR1. 187_R		TCGGGTCATCTGGGTTCTT	
PvCAD_F116	KarlCTG19538	TCACATCAAGCATCCACCATCT	Shen et al., 2012
PvCAD_R184		GTTCTCGTGTCCGAGGTGTGT	
HCT_973_F	AP13CTG44530	GCAGAAGGAGCAGCAGTCATC	Shen et al., 2012
HCT_1035_R		CGAGCGGCAATAGTCGTTGT	
PAL_F1	KarlCTG00004	CATATAGTGTGCGTGCGTGTGT	
PAL_R1		CTGGCCCGCCAATCG	
C3H_F1	AP13ISTG41630	CGTGAACAATGGGATCAGGATAG	
C3H_R1		GCGGACACAACCATCTCAAATAC	
F5H_F1	AP13ISTG56842	CCCCGTGCACTGACGATCTAT	
F5H_R1		CCAAGCCAAGGGAAAACACAGTTA	

Table 2-2 Comparison of the deduced amino acid sequences among switchgrass C₂₀ PvGA2ox proteins. Numbers indicate % identity between the predicted proteins.

	GA2ox5a	GA2ox5b	GA2ox6a	GA2ox6b	GA2ox9a	GA2ox9b	GA2ox11a	GA2ox11b
GA2ox5a								
GA2ox5b	89							
GA2ox6a	45	44						
GA2ox6b	47	46	93					
GA2ox9a	45	45	66	67				
GA2ox9b	37	37	52	54	71			
GA2ox11a	32	29	35	35	33	25		
GA2ox11b	31	32	30	32	33	27	89	

Table 2-3 Morphology and biomass yields of transgenic lines overexpressing *PvGA2ox5* and non-transgenic control (WT) plants.

Lines	Tiller height (cm)	Tiller number	Internode length (cm)	Fresh weight (g)	Dry weight (g)	Fresh/dry weight ratio
1	66.7±1.5 ^{bc}	31.3±2.0 ^{ab}	7.1±0.2 ^{bc}	53.1±2.7 ^a	13.9±0.8 ^a	3.8±0.19 ^{bcdef}
2	78.0 ±4.6 ^{ab}	21.3±3.8 ^{bcde}	10.0±0.4 ^a	50.5±10.3 ^a	16.6±3.6 ^a	3.1±0.06 ^f
3	76.3±1.5 ^{ab}	25.3±0.9 ^{bcde}	8.7±0.6 ^{ab}	52.6±3.3 ^a	14.2±1.4 ^a	3.7±0.15 ^{bcdef}
4	68.7±1.2 ^{abc}	28.3±2.7 ^{abcd}	8.9±0.4 ^a	48.1±5.8 ^a	13.5±1.8 ^a	3.6±0.07 ^{cdef}
6	72.3±0.9 ^{ab}	25.7±0.9 ^{bcde}	8.9±0.3 ^a	51.1±2.9 ^a	14.5±1.1 ^a	3.5±0.07 ^{cdef}
7	59.7±2.2 ^{cd}	46.3±9.3 ^a	5.4±0.1 ^c	49.6±10.1 ^a	11.8±2.5 ^{ab}	4.2±0.09 ^{bcd}
9	76.0±1.2 ^{ab}	28.7±2.0 ^{abc}	9.8±0.5 ^a	57.1±2.0 ^a	17.5±0.9 ^a	3.3±0.05 ^{def}
10	69.0±2.6 ^{abc}	24.7±1.2 ^{bcde}	9.2±0.4 ^a	38.5±7.9 ^{ab}	13.2±2.7 ^a	2.9±0.07 ^{cdef}
11	75.7±5.2 ^{ab}	30.0±5.0 ^{abc}	9.1±0.4 ^a	38.0±0.8 ^{ab}	11.9±0.2 ^{ab}	3.2±0.05 ^{ef}
12	50.3±1.5 ^d	28.3±1.8 ^{abcd}	5.4±0.2 ^c	16.4±5.3 ^{bc}	4.0±1.4 ^{bc}	4.2±0.15 ^{bcde}
5	10.7±2.3 ^e	10.3±1.5 ^{de}	-	1.5±0.5 ^c	0.3±0.1 ^c	4.4±0.22 ^{bc}
13	7.0±0.6 ^e	12.7±2.7 ^{cde}	-	0.7±0.2 ^c	0.1±0.02 ^c	5.5±1.09 ^a
14	9.5±0.5 ^e	9.0±2.0 ^e	-	1.1±0.7 ^c	0.2±0.14 ^c	4.6±0.23 ^{ab}
WT	80.7±2.2 ^a	17.0±3.5 ^{bcde}	9.9±0.3 ^a	42.3±7.6 ^a	14.2±2.6 ^a	3.0±0.03 ^f

The tiller height was the average of 5 tallest tillers. The fresh biomass was measured from the aboveground plant biomass cut at similar stages of growth while the dry biomass was measured on fresh biomass dried at 42 °C for 5 days. The plants were grown in 12 L pots in growth chambers under the same conditions for about 9 month before the measurements were taken. Values represented by different letters are significantly different at $P \leq 0.05$ as tested by LSD method with SAS software (SAS Institute Inc.).

Table 2-4 Sugar release by enzymatic hydrolysis in transgenic and non-transgenic control (WT) lines.

Transgenic lines	Glucose release (g/g CWR)	Xylose release (g/g CWR)	Total release (g/g CWR)
1	0.239±0.011 ^a	0.172±0.004 ^{ab}	0.411±0.025 ^a
2	0.207±0.006 ^c	0.171±0.002 ^{abc}	0.378±0.014 ^a
3	0.236±0.006 ^{ab}	0.171±0.001 ^{ab}	0.408±0.008 ^a
4	0.232±0.007 ^{abc}	0.169±0.003 ^{abc}	0.401±0.013 ^a
6	0.238±0.021 ^a	0.181±0.011 ^a	0.418±0.057 ^a
7	0.239±0.003 ^a	0.153±0.002 ^c	0.393±0.004 ^a
9	0.219±0.003 ^{abc}	0.166±0.007 ^{abc}	0.385±0.017 ^a
10	0.234±0.004 ^{abc}	0.175±0.007 ^{ab}	0.409±0.020 ^a
11	0.227±0.014 ^{abc}	0.169±0.011 ^{abc}	0.396±0.042 ^a
12	0.241±0.003 ^a	0.161±0.002 ^{bc}	0.401±0.009 ^a
WT	0.209±0.007 ^{bc}	0.181±0.001 ^a	0.390±0.011 ^a

All data are means ± standard error (n=3). Values represented by different letters are significantly different at $P \leq 0.05$ as tested by LSD method with SAS software (SAS Institute Inc.). CWR, cell wall residues.

Table 2-5 List of the locus names and GenBank accession numbers of the sequences used in this study.

Gene name	Locus name/accession number	Species
PvGA2ox1a	Pavir.Ca01130	<i>Panicum virgatum</i>
PvGA2ox1b	Pavir.J13005	<i>Panicum virgatum</i>
PvGA2ox2a	Pavir.Ea01271	<i>Panicum virgatum</i>
PvGA2ox2b	Pavir.Eb02382	<i>Panicum virgatum</i>
PvGA2ox3a	Pavir.J27068	<i>Panicum virgatum</i>
PvGA2ox3b	Pavir.Eb03210	<i>Panicum virgatum</i>
PvGA2ox4a	Pavir.Ca01636	<i>Panicum virgatum</i>
PvGA2ox4b	Pavir.J11479	<i>Panicum virgatum</i>
PvGA2ox5a	Pavir.Ba04002	<i>Panicum virgatum</i>
PvGA2ox5b	Pavir.Bb00108	<i>Panicum virgatum</i>
PvGA2ox6a	Pavir.Ga01287	<i>Panicum virgatum</i>
PvGA2ox6b	Pavir.Gb01218	<i>Panicum virgatum</i>
PvGA2ox7a	Pavir.Ea00180	<i>Panicum virgatum</i>
PvGA2ox7b	Pavir.J04662	<i>Panicum virgatum</i>
PvGA2ox8a	Pavir.J34867	<i>Panicum virgatum</i>
PvGA2ox8b	Pavir.Ca00918	<i>Panicum virgatum</i>
PvGA2ox9a	KanlCTG23388 (Switchgrass Unitranscript ID)	<i>Panicum virgatum</i>
PvGA2ox9b	Pavir.J11300	<i>Panicum virgatum</i>
PvGA2ox11a	Pavir.Da02356	<i>Panicum virgatum</i>
PvGA2ox11b	Pavir.J38727	<i>Panicum virgatum</i>
OsGA2ox1	LOC_Os05g06670/NP_001054711	<i>Oryza sativa</i>
OsGA2ox2	LOC_Os01g22910/EAY73841	<i>Oryza sativa</i>
OsGA2ox3	LOC_Os01g55240/NP_001044292	<i>Oryza sativa</i>
OsGA2ox4	LOC_Os05g43880/EEE64351	<i>Oryza sativa</i>
OsGA2ox5	LOC_Os07g01340/NP_001058690	<i>Oryza sativa</i>
OsGA2ox6	LOC_Os04g44150/NP_001053341	<i>Oryza sativa</i>

Table 2-5 Continued.

Gene name	Locus name/accession number	Species
OsGA2ox7	LOC_Os01g11150/NP_001042364	<i>Oryza sativa</i>
OsGA2ox8	LOC_Os05g48700/NP_001056311	<i>Oryza sativa</i>
OsGA2ox9	LOC_Os02g41954/EEE57418	<i>Oryza sativa</i>
OsGA2ox10	LOC_Os05g11810/AAT01379	<i>Oryza sativa</i>
OsGA2ox11	LOC_Os04g33360/NP_001052718	<i>Oryza sativa</i>
AtGA2ox7	AT1G50960/NP_175509	<i>Arabidopsis thaliana</i>
AtGA2ox8	AT4G21200/NP_193852	<i>Arabidopsis thaliana</i>
SoGA2ox3	AY935713	<i>Spinacia oleracea</i>
PtGA2ox9	Potri.004G022800/XP_002304931	<i>Populus trichocarpa</i>
PtGA2ox10	Potri.011G026700	<i>Populus trichocarpa</i>
PtGA2ox11	Potri.011G134000/XP_002317574	<i>Populus trichocarpa</i>
BdGA2ox1	Bradi1g59570/XP_003561524	<i>Brachypodium distachyon</i>
BdGA2ox2	Bradi5g16040/XP_003581438	<i>Brachypodium distachyon</i>
BdGA2ox3	Bradi3g49390/XP_003575413	<i>Brachypodium distachyon</i>
ZmGA2ox1	GRMZM2G177104/DAA59321.1	<i>Zea mays</i>
ZmGA2ox2	GRMZM2G006964/ACN35669.1	<i>Zea mays</i>
ZmGA2ox3	GRMZM2G153359/AFW72428.1	<i>Zea mays</i>
SbGA2ox1	Sb02g000460/XP_002459201	<i>Sorghum biocolor</i>
SbGA2ox2	Sb06g022880/XP_002448199	<i>Sorghum biocolor</i>
SiGA2ox1	Si017671m.g/XP_004953177	<i>Setaria italica</i>
SiGA2ox2	Si020696m.g/XP_004987270	<i>Setaria italica</i>
SiGA2ox3	Si012266m.g/XP_004976262	<i>Setaria italica</i>

Chapter 3: Identification and overexpression of a Knotted1-like transcription factor in switchgrass (*Panicum virgatum* L.) for lignocellulosic feedstock improvement

(A version of this chapter has been submitted to Biotechnology for Biofuels, authored by Wegi A. Wuddineh, Mitra Mazarei, Jiyi Zhang, Geoffrey B. Turner, Robert W. Sykes, Stephen R. Decker, Mark F. Davis, Michael K. Udvardi, C. Neal Stewart, Jr.)

Wegi A. Wuddineh designed and conducted the experiments except lignin and sugar release assays and wrote the manuscript.

3.1 Abstract

High biomass production and wide adaptation has made switchgrass (*Panicum virgatum* L.) a leading candidate lignocellulosic bioenergy crop. One major limitation of this and other lignocellulosic feedstocks is the recalcitrance of complex carbohydrates to hydrolysis for conversion to biofuels. Lignin is the major contributor to recalcitrance as it limits the accessibility of cell wall carbohydrates to enzymatic breakdown into fermentable sugars. Therefore, genetic manipulation of the lignin biosynthesis pathway is one strategy to reduce recalcitrance. Here, we identified a switchgrass Knotted1 (PvKN1) transcription factor (TF) with the aim of genetically engineering switchgrass for reduced biomass recalcitrance for biofuel production. The endogenous *PvKN1* was mainly expressed in young inflorescences and stems. Ectopic overexpression of *PvKN1* in switchgrass altered growth phenotype particularly at early developmental stages as compared with the non-transgenic control. Transgenic lines had reduced expression of most lignin biosynthetic genes accompanied by a reduction in lignin content highlighting the involvement of *PvKN1* in the regulation of the lignin biosynthesis pathway. Moreover, the reduced expression of *Gibberellin 20-oxidase* (*GA20ox*) in tandem with increased expression of *Gibberellin 2-oxidase* (*GA2ox*) genes in transgenic *PvKN1* lines suggest that *PvKN1* may control the expression of some target genes via regulation of GA signalling. Furthermore, overexpression of *PvKN1* altered the expression of cellulose and hemicellulose biosynthetic genes and increased sugar release efficiency in transgenic lines. Our results demonstrated that switchgrass *PvKN1* is a putative ortholog of maize *KN1* that is linked to plant lignification, possibly via regulation of the GA-biosynthetic and catabolic pathway. *PvKN1* is also implicated in further control of cell wall biosynthetic pathways such as those controlling the synthesis of cellulose and hemicelluloses. Therefore, overexpression of *PvKN1* in bioenergy

feedstocks may provide one feasible strategy for reducing biomass recalcitrance and improving biomass characteristics for biofuel.

Keywords: switchgrass, Knotted1, KNOX transcription factor, biofuel

3.2 Introduction

Switchgrass, a C4 perennial prairie forage grass indigenous to North American grasslands, is a leading candidate as a lignocellulosic biofuel feedstock owing to its potential high biomass yield, resistance to stress conditions, high nutrient-use efficiency, fast growth and ability to thrive on marginal land unsuitable for row crops (Yuan et al., 2008). Controlled field studies to determine the net energy efficiency of switchgrass monoculture on marginal lands showed that it can produce as much as 540% more renewable energy than the non-renewable energy consumed for its production (Schmer et al., 2008), highlighting the potential of this feedstock as a renewable and clean energy source. One of the major obstacles for the development of this and other lignocellulosic biomass feedstocks for biofuels is biomass recalcitrance (resistance of the cellulose and hemicellulose in cell walls to breakdown into fermentable sugars). A current requirement is costly pre-treatment before saccharification and fermentation steps in the processing of lignocellulosic biomass (Ragauskas et al., 2006). Reducing recalcitrance to enhance the economic competitiveness of lignocellulosic-based biofuels is an overarching goal in bioenergy research (Lynd et al., 2008).

Lignin is a complex aromatic polymer composed of hydroxyphenyl, guaiacyl and syringyl monolignols. Transgenic approaches to reduce lignin, one of the major contributors to biomass recalcitrance. Mostly, a gene-by-gene strategy has been used to downregulate individual lignin biosynthetic genes, which has been effective to alter lignin (Fu et al., 2011a; Xu et al., 2011; Baxter et al., 2014; Saathoff et al., 2011; Fu et al., 2011b). In this regard, several enzymes responsible for monolignol biosynthesis have been identified and functionally characterized in switchgrass (Fu et al., 2011a; Xu et al., 2011; Saathoff et al., 2011; Fu et al., 2011b; Escamilla-Treviño et al., 2010; Shen et al., 2013). Genetically targeting individual lignin genes to block specific branches in the lignin biosynthesis pathway has been successful, but this strategy appears to have limitations as the build-up of low molecular weight phenolic compounds and other fermentation inhibitors may eventually reduce biofuel output (Tscharplinski et al., 2012). However, use of transcriptional repressors of the lignin biosynthesis pathway might be a more effective strategy, as shown by the recent work in which the switchgrass gene coding for MYB4 transcription factor (TF) was overexpressed (Shen et al., 2012; Shen et al., 2013).

Lignin biosynthesis is regulated by a multi-layered network of TFs including top-tier NAC (NAM, ATAF1/2 and CUC2) and second-tier MYB master regulators (Zhong & Ye, 2014; Handakumbura & Hazen, 2012). The NAC TFs, which activate the lignin biosynthesis pathway via activation of downstream MYB TFs have been well documented in eudicots such as *Arabidopsis* and *Populus* (Zhong & Ye, 2014; Zhao et al., 2010). Recent studies reported the identification of some rice and maize secondary wall NACs (SWN) that are orthologs of *Arabidopsis* SWNs (Zhong et al., 2011; Chai et al., 2015). Similarly, several second-tier MYBs, which control the activation of downstream MYB TFs specific to lignin biosynthesis pathway, have been characterized including *Arabidopsis* MYB46/83 and its homologs in *Populus* (PtrMYB2/3/20/21), pine (PtMYB4/8), *Eucalyptus* (EgMYB2), rice (OsMYB46) and maize (ZmMYB46) (Zhong & Ye, 2014; Zhong et al., 2011). The lignin-specific MYBs act either as a repressors or activators of lignin biosynthesis pathway genes. Several MYB TFs that are repressors of lignin biosynthesis pathway have been identified in dicots including *Arabidopsis* (MYB4/7/32) and *Eucalyptus* (MYB1) (Shen et al., 2012; Zhong & Ye, 2014; Fornale et al., 2010). MYB TFs that are activators of the lignin biosynthesis pathway were also identified in several dicots including *Arabidopsis* (MYB58/63 and MYB85), *Populus* (MYB26/28/90/152) and pine (MYB1) (Zhong & Ye, 2014; Zhou et al., 2009). However, with the exception of maize MYB31 (Fornale et al., 2010), wheat MYB4 (Ma et al., 2011) and a recently characterized switchgrass MYB4 TFs (Shen et al., 2013; Zhong & Ye, 2014), there is limited available information about transcriptional regulation of lignin biosynthetic program in monocots (Handakumbura & Hazen, 2012).

Knotted1(KN1)-like homeobox (KNOX) proteins are members of plant-specific three amino acid loop extension (TALE) family of TFs, which play a crucial role in the maintenance of meristem tissues and the regulation of various morphogenic processes throughout plant development (Hay & Tsiantis, 2010). These TFs are grouped into two major classes based on their homeodomain (HD) identity, intron positions and expression patterns (Hay & Tsiantis, 2010; Kerstetter et al., 1994; Mukherjee et al., 2009; Testone et al., 2012). Class I KNOX genes are mainly expressed in the shoot apical meristems (SAM) where they redundantly regulate stem cell maintenance, and in vascular cambium. Class II KNOX genes are diversely expressed and their functions are not as well characterized (Hake et al., 2004; Jackson et al., 1994; Du et al., 2009). Members of class I

KNOX genes from *Arabidopsis* (*KNAT1* also known as *BREVIPEDICELLUS/BP*), peach (*KNOPE1*), *Populus* (*ARBORKNOX2/ARK2*) and maize (*KN1*) (Testone et al., 2012; Du et al., 2009; Mele et al., 2003; Townsley et al., 2013), and a member of class II *KNOX* genes from *Arabidopsis* (*KNAT7*) (Li et al., 2012) have been shown to function as repressors of lignin biosynthetic genes. *KNAT1* overexpression in *Arabidopsis* has been shown to decrease lignin deposition as well as the expression of most lignin biosynthetic genes while binding the promoters of at least two lignin biosynthetic genes, *caffeic acid O-methyltransferase (COMT)* and *caffeoyl CoA 3-O-methyltransferase (CCoAOMT)* (Mele et al., 2003). Similarly, a *Populus* ortholog of *KNAT1*, *ARK2* has also been shown to negatively regulate several genes in the lignin biosynthesis pathway followed by marked reduction of the polymerized lignin in the stem (Du et al., 2009). Interestingly, *ARK2* overexpression in *Populus* was found to be associated with reduced expression of secondary cell wall cellulose synthase genes (*Ces*) including three xylem-specific (*PttCesA1*, *PttCesA3-1* and *PttCesA3-2*) and two phloem-specific (*PttCesA2* and *estExt_Genewise1_v1.C_LG_VI2188*) *CesA* genes (Du et al., 2009). Despite some information on class I *KNOX* gene regulation of cell wall biosynthesis pathway in eudicots, such reports are scant in monocots with the exception of maize.

A recent study showed that among three members of maize class I *KNOX* genes investigated for roles in the regulation of lignin biosynthesis, overexpression of *KN1* significantly reduced lignin deposition in maize and tobacco while altering the expression of only two of the four lignin biosynthetic genes analyzed in tobacco (Townsley et al., 2013). However, it was not reported whether similar changes in the expression of lignin biosynthetic genes were observed in maize. This gene has also been shown to regulate gibberellin (GA) signalling via the positive transcriptional regulation of maize *GA 2-oxidase (GA2ox1)* (codes for an enzyme that inactivates bioactive GA) (Bolduc & Hake, 2009). Overexpression of *KNAT1* has previously been shown to decrease the levels of *GA 20-oxidase (GA20ox1)* mRNA in *Arabidopsis* leaves (Hay et al., 2002) as did the overexpression of *ARK2* gene in hybrid aspen (*P. alba* x *P. tremula*) (Du et al., 2009). GA signalling plays vital roles in the regulation of various developmental and growth processes and its downregulation has been shown to alter plant architecture and lignin deposition (Thomas et al., 2005; Biemelt et al., 2004; Zhao et al., 2010). Our recent study demonstrated that overexpression of switchgrass GA catabolic gene *PvGA2ox5* caused reduced lignification and

enhanced sugar release efficiency (Wuddineh et al., 2015). It appears at least in model dicots that KNOX TFs may regulate the lignification process via modulation of the GA signalling pathways. However, to the best of our knowledge, there are no published reports on whether class I KNOX TFs from monocots regulate the lignification process and any mechanism thereof.

With the exception of maize KN1-induced alteration of plant lignification, the potential of KNOX TFs to alter various growth and biomass characteristics of bioenergy crops is largely untapped. Investigation of the effect of overexpression of *KN1* genes in monocots on the lignin biosynthesis pathway and the potential effects on downstream plant growth parameters is important. Therefore, the objective of this study was to identify and characterize the switchgrass class I KNOX gene (*PvKN1*) and investigate how it regulates various plant morphological and developmental processes, including lignin deposition, and impacts cell wall chemistry and recalcitrance in switchgrass.

3.3 Materials and Methods

3.3.1 Plants and growth conditions

Transgenic and non-transgenic control plants were grown under the same conditions (16 h day/8 h night light at 24°C, 390 $\mu\text{E m}^{-2} \text{s}^{-1}$) in growth chambers and watered three times per week, including weekly nutrient supplements with 100 mg/L Peter's 20-20-20 fertilizer. For quantification of growth parameters, each transgenic and non-transgenic line was propagated from a single tiller to yield three clonal replicates each (Hardin et al., 2013). Leaf blades, leaf sheath, internode sections, and panicle samples pooled from the top two internodes and young roots, were collected from tillers at the R1 developmental stage (More et al., 1991; Shen et al., 2009) to assay transcript abundance. Each sample was snap-frozen in liquid nitrogen and pulverized with mortar and pestle in liquid nitrogen. The pulverized samples were used for RNA extraction as described below.

3.3.2 Gene identification, vector construction and transgenic plant production

TBLASTN was used to identify the homologous KNOX gene sequences from switchgrass EST databases (Zhang et al., 2013) and draft genome (*Panicum virgatum* v1.1 DOE-JGI) at phytozome (<http://www.phytozome.net/panicumvirgatum>) using the amino acid sequences of

ZmKN1 (gb/AAP21616.1) and KNAT1 (At4g08150) as heterologous probes. Subsequently, the most closely related genes (*PvKN1a* and *PvKN1b*) were identified for cloning based on the cluster analysis and multiple sequence alignment (MSA) analysis. Overexpression cassettes were constructed by isolating target gene ORFs from switchgrass cDNAs of the ST1 clonal genotype of ‘Alamo’ switchgrass using individual gene-specific primers flanking the ORF of each gene and subsequently cloning each into pCR8 entry vector for sequence confirmation. The list of primer pairs used for cloning is shown in Table 3-1. Sequence-confirmed ORFs were then sub-cloned into pANIC-10A expression vector by GATEWAY recombination (Mann et al., 2012) to place each gene of interest under the control of the maize ubiquitin 1 (*ZmUbi1*) promoter. Embryogenic callus derived from Alamo switchgrass ST1 or SA37 genotypes (King et al., 2014) was transformed with the expression vector construct using *Agrobacterium*-mediated transformation (Burriss et al., 2009). Selection for positive transformants was carried out for approximately two months on 30- 50 mg/L hygromycin followed by regeneration of OFP reporter-positive callus sections on regeneration medium (Li & Qu, 2011) containing 400 mg/L timentin. Regenerated plants were rooted on MS medium (Murashige & Skoog, 1962) containing 250 mg/L cefotaxime (Grewal et al., 2006) and the transgenic lines were screened based on the presence of the insert and expression of the transgene. The non-transgenic control ST1 and SA37 lines were generated in tissue culture in parallel with the transgenic lines. Rice transformation was performed using callus derived from mature seeds of rice variety TP309 as described before (Nishimura et al., 2006).

3.3.3 RNA extraction and qRT-PCR

RNA extraction and analysis of transgene transcripts were performed as previously described (Wuddineh et al., 2015). Briefly, total RNA was extracted from the shoot-tips of transgenic and non-transgenic lines at E4 developmental stage using Tri-Reagent (Molecular Research Center, Cincinnati, OH). Each 3 µg of the purified RNA sample was treated with DNase-I (Promega, Madison, WI) to eliminate any potential genomic DNA contaminants. High-Capacity cDNA Reverse Transcription kit (Applied Biosystems, Foster City, CA) was used for first-strand cDNA synthesis using the DNase-treated RNA. qRT-PCR analysis was conducted using Power SYBR Green PCR master mix (Applied Biosystems) according to the manufacturer's protocol. All the experiments were conducted in triplicates. Gene-specific forward primers and an AcV5 tag-

spanning reverse primer was used for qRT-PCR of the transgenic lines. The list of all primer pairs used for qRT-PCR is shown in Table 3-1. The relative expression was analyzed using the C_T method using ubiquitin (UBQ) (Switchgrass Unitranscript ID: AP13CTG25905) as a reference gene (Xu et al., 2011; Shen et al., 2009). No amplification products were observed with all the primer pairs when using only the RNA samples or water instead of cDNA.

3.3.4 Phloroglucinol staining

Qualitative lignin analysis was performed as previously described (Wuddineh et al., 2015). Briefly, leaf samples at R1 developmental stage were collected and cleared in a 2:1 solution of ethanol and glacial acetic acid for 5 days (Bart et al., 2010). The 3rd leaves from the base of the stem were used in this experiment as these were the most uniform-looking in terms of maturity. Subsequently, the cleared leaf sample was immersed in 1% phloroglucinol (in 2:1 ethanol/HCl) overnight for staining. Photographs were taken at 1x magnification using infinity X32 digital camera mounted on Fisher Scientific Stereomaster microscope (Pittsburgh, PA).

3.3.5 Determination of lignin content and composition by py-MBMS

Quantitative lignin analysis and S/G lignin monomer ratio were determined using tillers collected at R1 developmental stage and air dried for 3 weeks at room temperature, followed by milling to 1 mm (20 mesh) particle size. Lignin content and composition were determined at NREL using high-throughput py-MBMS on extractives- and starch-free samples (Baxter et al., 2014; Sykes et al., 2009).

3.3.6 Determination of sugar release

For enzymatic hydrolysis, tiller samples were collected at R1 developmental stage and air-dried for 3 weeks at room temperature before grinding to 1 mm (20 mesh) particle size. Subsequently, sugar release efficiency was determined using high-throughput sugar release assays on extractives- and starch-free samples (Selig et al., 2010; Decker et al., 2012). Briefly, the milled samples were prepared in a 96-well plate along with biomass-only controls, sugar standards, enzyme-only, and blank wells. The pretreatment was performed in a steam chamber at 180 °C for 40 minutes. The enzyme cocktails used for hydrolysis comprises Novozymes Cellic CTec2 cellulase (Novozymes North America, Franklinton, NC, USA) at 70 mg protein/g initial biomass

and Novo188 β -glucosidase (Novozymes) at 2.5 mg protein/g initial biomass. The mixture was incubated for 72 hours at 54 °C and glucose and xylose release were determined by colorimetric assays. The total sugar release was defined as the sum of glucose and xylose released. Each sample for sugar release assays were run in triplicate.

3.3.7 Data analysis

Analysis of variance (ANOVA) with Fisher's least significant difference (LSD) procedure was used to analyze the differences between treatment means while PROC TTEST procedure was used to examine the statistical difference between the expression of target genes in transgenic vs non-transgenic lines using SAS version 9.3 (SAS Institute Inc., Cary, NC).

3.4 Results

3.4.1 KNOX family of TFs in switchgrass and the identification of PvKN1

A scan of the publicly available switchgrass genome for the characteristic Meinox domain (KNOX1 and KNOX2) and HD revealed a total of 10 genes belonging to the KNOX family of TFs. The tetraploid switchgrass genome of Alamo AP13 ($2n=4x=36$) contains at least two sub-genomic gene variants representing the 'A' and 'B' subgenomes (Figure 3-1). Phylogenetic and sequence analysis showed 7 *KNOX* genes belonging to class I family of KNOX TFs while the remaining 3 were grouped into the class II family of KNOX TFs (Figures 3-1 and 3-2). Analysis of the amino acid sequences of the encoded proteins indicated that all share KNOX1, KNOX2, ELK and HD conserved domains found in the maize KN1, *Arabidopsis* KNAT1 and orthologous TFs in eudicots (peach and *Populus*). Moreover, higher similarity within members of each class (>44% in class I and >60% in class II) than between the two classes of *KNOX* genes ($\leq 37\%$) was observed (Table 3-2). Analysis of gene structure among the two classes of *KNOX* genes in comparison to the homologs from *Arabidopsis* (*KNAT1* and *KNAT7*) and maize (*KN1*) showed a highly conserved pattern in each class, typically with 5 exons and 4 introns in the genomic DNA with variable length of introns among the genes (Figure 3-3). The exception to this pattern was *PvKN9*, a member of class II *KNOX* genes family, which had an additional intron inserted in the last exon and, therefore, had a structure of 6 exons and 5 introns.

Based on amino acid sequence analysis, sub-genomic variants belonging to class I family of KNOX TFs designated as PvKN1a, PvKN1b, PvKN2a and PvKN2b showed the highest identity to the maize KN1 and *Arabidopsis* KNAT1 when compared with the other switchgrass class I family members of KNOX TFs (Figure 3-2 and Table 3-3). However, based on cluster analysis of the KNOX proteins, PvKN1a and PvKN1b showed more similarity to the group containing the previously characterized KNOX TFs from *Arabidopsis* (KNAT1/BP), maize (ZmKN1), rice (OSH1), wheat (WKNOX1), *Populus* (ARK2) and peach (PpKNOPE1) while PvKN2a and PvKN2b clustered more closely with maize rough sheath 1 (ZmRS1), rice homeobox (OsHOS3) and wheat rough sheath 1 (WRS1) TFs (Figures 3-1 and 3-2). Moreover, the two gene variants of *PvKN1* (*PvKN1a* and *PvKN1b*) with 95% identity in the predicted amino acid sequences shared over 88% identity with *ZmKN1* as opposed to 54% for *PvKN2a* or *PvKN2b* (Table 3-3). More importantly, comparison of the HD domain, essential for binding target sequences, showed over 98% identity between the PvKN1 and ZmKN1 proteins versus 89% in the case of PvKN2 (Table 3.3). Moreover, PvKN1 displayed the highest conservation at the glycine, serine and glutamic acid-rich motif (also known as GSE) (the stretch of amino acids between the KNOX2 and ELK domains) compared with PvKN2 (Table 3.3).

3.4.2 Expression patterns of *PvKN1* variants

Quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis showed detectable levels of expression for both *PvKN1* gene variants in stems, leaves, leaf sheaths and inflorescences at reproductive stage 1 (R1), the same stage when sugar release, lignin content and growth parameters were analyzed (Figure 3-4). The relative transcript abundance of *PvKN1a* appears to be higher than that of *PvKN1b* in stems, leaf sheath and young inflorescences. Moreover, *PvKN1a* expression was the highest in young inflorescence tissues followed by the stem with the lowest expression observed in roots and leaves (Figure 3-4). The expression of *PvKN1b* was not significantly different among tissues assayed (Figure 3-4).

3.4.3 Constitutive overexpression of *PvKN1* in switchgrass

The switchgrass homologs of maize *KN1* gene, *PvKN1a* and *PvKN1b* were cloned from cDNA and constitutively overexpressed in switchgrass under the control of the maize ubiquitin 1 promoter. In addition, the same *PvKN1a* construct was overexpressed in rice to examine its

effect on plant development. Only transgenic calli expressing the OFP marker gene were regenerated (Figure 3-5). Five independent transgenic switchgrass lines overexpressing *PvKN1a* and eight *PvKN1b*-overexpressing lines were recovered.

3.4.4 Altered growth phenotypes in transgenic plants

The transgenic lines overexpressing the two *PvKN1* gene variants had similar phenotypes, which suggest that each variant may have similar functions upon overexpression (Figure 3-6). Therefore, subsequent work focussed on detailed analysis of *PvKN1a* overexpressing lines. We observed unusually narrowed leaf blades with sheath-like tissues and curled leaves just subsequent to plant regeneration. In addition, the *PvKN1a* overexpression lines had severely inhibited shoot and root elongation as compared to the non-transgenic controls at similar developmental stage (Figure 3-7a & b). Many of the transgenic lines recovered were dwarfed and some were not fully regenerable. However, several transgenic lines that were regenerated were able to develop into mature plants displaying less severe phenotypes after establishment in the rooting media (Figure 3-7c); some of these lines eventually grew normally (Figure 3-7d). The successfully established transgenic lines were confirmed by genomic PCR using transgene and hygromycin-resistance gene specific primers, as well as visualization of OFP in transgenic plants compared to the non-transgenic control lines (Figure 3-5a-c). Moreover, there were two- to nine-fold increases in the *PvKN1* transcript levels in transgenics relative to control lines (Figure 3-8). There were no statistically significant differences in tiller height, tiller number or biomass yield between transgenic lines and the non-transgenic control (Table 3-4).

Overexpression of *PvKN1* in a different switchgrass genotype (SA37) background showed abnormal phenotypes at later stages of growth compared to the genotype (ST1) used in this study. These transgenic lines had relatively shorter internodes and twisted or bent culms with altered leaf structure including curly leaves, fused leaf sheath-blade boundary and missing ligules (Figure 3-9a-c). The relative expression of the *PvKN1* transcript was >50-fold in transgenic compared to the non-transgenic control lines (Figure 3-9d).

When *PvKN1* was overexpressed in rice, similar aberrant phenotypes were observed with clumps of vegetative shoots developed from the transgenic callus after incubation on the shoot

regeneration medium. In general, shoot elongation was severely reduced in rice even after longer incubation on the growth medium while root growth was apparently unaffected (Figure 3-10).

3.4.5 Overexpression of *PvKN1* in switchgrass and subsequent changes in the expression of putative target genes in transgenic plants

To test whether overexpression of *PvKN1* in switchgrass could affect the expression of putative target genes, qRT-PCR analysis was conducted. *PvKN1*-overexpressing switchgrass displayed significantly reduced expression of lignin biosynthetic genes including *PvC4H*, *PvCAD* and *PvCCR* compared with the control (Figure 3-11a). The expression of putative cellulose and hemicellulose biosynthetic genes were also evaluated in transgenic and non-transgenic lines. The results showed that the relative transcript level of putative primary cell wall cellulose biosynthetic gene *PvCESA1* was significantly reduced, while the expression of other putative primary cell wall biosynthetic genes, including *PvCESA3* and that of three secondary cell wall associated putative *PvCESA* genes *PvCESA4*, *PvCESA7* and *PvCESA8*, were unchanged in transgenic compared with the control lines. A significant increase in the expression level of the putative hemicellulose biosynthetic gene, *PvCSLD1* was also observed in transgenic compared to the control lines (Figure 3-11b). The expression of putative GA signalling pathway genes was also evaluated in transgenic and non-transgenic lines. The results showed that the expression of putative *PvGA20ox* gene (*PvGA20ox1a*) was significantly reduced in transgenic compared to the non-transgenic control lines. In contrast, the relative expression level of one (*PvGA20ox6*) of the three C₂₀ *PvGA20ox* genes (responsible for catabolism of the 20-carbon precursors of bioactive GA) evaluated in this study showed a significant 6-fold increase in transgenic switchgrass compared with non-transgenic controls (Figure 3-11c).

3.4.6 Changes in lignin content and composition in transgenic switchgrass

Histochemical staining of leaves with phloroglucinol-HCl showed reduction in lignin staining in transgenic lines relative to the non-transgenic control (Figure 3-12). Moreover, quantitative analysis of total lignin content in whole tillers by pyrolysis- molecular beam mass spectrometry (py-MBMS) from the R1 developmental stage showed a slight reduction in lignin content by up to 8% in the transgenic lines as compared to the control (Figure 3-13a). The syringyl/guaiacyl

(S/G) lignin monomer ratio in R1 tillers of transgenic lines was not different from that in the non-transgenic control (Figure 3-13b).

3.4.7 Effect of overexpression of PvKN1 on sugar release efficiency in switchgrass

Sugar release in R1 whole tillers was significantly increased in transgenic line 4 that has the highest expression of the transgene, with 15% more glucose and 12% more xylose released as compared to the non-transgenic control. The total sugar release (glucose and xylose combined) in this line was increased by up to 13% as compared to the control (Figure 3-14).

3.5 Discussion

In this study, we identified 10 putative KNOX TF genes in the switchgrass genome, which include 7 members of the class I and 3 of the class II families (Figure 3-1 and Table 3-3). So far, a total of 13 *KNOX* genes in maize, 12 in rice, 15 in *Populus*, 8 in *Arabidopsis*, 10 in peach, 5 in *Selaginella moellendorffii*, 5 in moss and 1 in *Chlamydomonas reinhardtii* have been reported (Hay & Tsiantis, 2010; Mukherjee et al., 2009; Testone et al., 2012). Thus, it appears that tetraploid switchgrass could harbor more *KNOX* genes, which may be revealed as the more complete genome sequences are made available. In general, the identified switchgrass *KNOX* genes showed high similarity to homologs in other plant species in terms of gene structure, amino acid sequence identity of deduced proteins, with shared specific domains (Figures 3-1, 3-2 & 3-3, Table 3.3). All members share the characteristic domains found in the maize KN1 including Meinox (essential for suppressing target gene expression and homodimerization), ELK (nuclear localization signal) and HD (homeodomain) (Figure 3-2) (Kerstetter et al., 1994; Nagasaki et al., 2001). Specifically, PvKN1 from class I family KNOX TFs had the highest amino acid sequence identity (>90%) with the well-characterized maize KN1 but even more identity (>98%) in the HD domain (Figure 3-2, Table 3-3). These remarkable structural resemblances between the *KNOX* genes among both monocots and eudicots perhaps reflect functional conservation among homologs, which is supported by functional studies (Testone et al., 2012; Du et al., 2009; Mele et al., 2003; Townsley et al., 2013).

Most Class I *KNOX* genes are highly expressed in meristematic tissue in accordance with their proposed function in the maintenance of meristem identity (Kerstetter et al., 1994; Tioni et al.,

2003; Testone et al., 2008; Chuck et al., 1996; Kerstetter et al., 1997). Similarly, *PvKN1* was also expressed mainly in stem and inflorescence tissues containing meristems, in line with the importance of *KNOX* homologs in the regulation of apical or intercalary meristem development in young tissues of other species (Kerstetter et al., 1994; Kerstetter et al., 1997). It is interesting to note that *PvKN1* transcripts, albeit at low levels, were detected in differentiated lateral organs, including leaves and leaf sheaths, in contrast to the situation in maize, rice, *Arabidopsis* and *Populus* where expression of *KNOX* genes is scantily detected in such tissues (Kerstetter et al., 1994; Du et al., 2009; Chuck et al., 1996; Matsuoka et al., 1993). However, similar expression pattern of *KNOX* genes in differentiated tissues has been reported in barley, sunflower and tomato (Tioni et al., 2003; Muller et al., 2001; Hareven et al., 1996). *KNOX* gene expression in tomato leaves was associated with compound leaf development (Hareven et al., 1996). One possible reason for the lack of *KNOX* gene expression phenotype (ectopic SAM formation) in differentiated switchgrass tissues might be the low transcript level, which might be below the required threshold level to induce the phenotype. Moreover, the involvement of various cell and species-specific factors such as posttranscriptional regulators in the determination of developmental responses to the expression of *KN1*-like genes might also contribute to this deviation as previously suggested (Hay & Tsiantis, 2010; Tioni et al., 2003). Whether or not *PvKN1* has a role in non-meristematic tissues remains to be seen.

Ectopic expression of *KNOX* genes have been shown to modify plant growth and development in both monocots and dicots (Hay & Tsiantis, 2010; Du et al., 2009; Matsuoka et al., 1993; Sentoku et al., 2000). Similarly, ectopic overexpression of *PvKN1* in switchgrass dramatically altered growth phenotype, including plant architecture, leaf structure and shoot elongation in transgenic plants especially, at early developmental stages (Figure 3-7b and 3-6b). There was a gradation of phenotype severity in different transgenic lines. In the worst-case, growth was arrested at the early seedling stage (Figure 3-7a), whereas other lines eventually attained normal shoots and roots after prolonged incubation on the rooting medium (Figure 3-7c) even though *PvKN1* transcript level in these lines was up to 9 times higher than the control. In contrast, higher levels of *PvKN1* transcripts (upwards of 50-fold increase) in transgenic lines in the SA37 switchgrass genotype background displayed aberrant phenotypes even at later stages of development (Figure 3-9). The apparent phenotypic differences in these transgenic plants might be attributed to the

differences in transcript abundance as well as the genotypic differences in response to *PvKN1* overexpression as previously reported for maize (Kerstetter et al., 1997; Foster et al., 1999). The dependence of phenotype severity on *KNOX* transcript level has been reported in rice and *Arabidopsis* (Hay et al., 2003; Sentoku et al., 2000).

Interestingly, ectopic expression of *PvKN1* failed to induce ectopic meristem development in lateral organs, in contrast to other plants with simple leaves such as maize *KN1* dominant gain-of-function mutants (Vollbrecht et al., 1991), rice overexpressing *OSH1* (Matsuoka et al., 1993) or *Populus* overexpressing *ARK1* (Groover et al., 2006). However, differences in the effect of *KNOX* gene overexpression may be related to levels or patterns of ectopic overexpression. Consistent with this idea, no ectopic meristems or shoot phenotypes were observed in eudicots such as tobacco that overexpressed the rice *OSH1* gene (Kano-Murakami et al., 1993), and *Populus* that overexpressed *ARK2* genes (Du et al., 2009), or in rice overexpressing *OSH1* under a weak CaMV 35S promoter (Sentoku et al., 2000). On the other hand, *PvKN1* overexpressing plants clearly exhibited altered distal-proximal patterns of leaf growth where the development of distal leaf blades was inhibited, which, in turn gave rise to the development of proximal leaf sheath-like tissues as previously reported in rice and maize expressing *KNOX* genes (Figure 3-7a, 3-6a) (Sentoku et al., 2000; Vollbrecht et al., 1991). This observation is in accordance with the maturation schedule hypothesis, which proposes that leaf primordia progress through a series of developmental stages along the proximal to distal axis of the leaf to form distinct tissue types, namely sheath, ligule, auricle and leaf blade, and ectopic expression of *KNOX* genes causes developmental delays impeding cells from normal progression through subsequent phases of maturation schedule (Freeling, 1992).

Another intriguing observation was the twisted stem and altered leaf phenotypes in SA37 switchgrass plants overexpressing *PvKN1*, which resembled the phenotypes observed in dominant mutants of maize *KNAT4* (*Gnarley1*) (Foster et al., 1999). This specific phenotype was not observed in rice and maize overexpressing the homologs of *PvKN1* (Sentoku et al., 2000; Vollbrecht et al., 1991). The diverse effects of ectopic *KNOX* gene expression in lateral organs may reflect the differential competence of tissues in addition to the transcript level in dictating tissue response to *KNOX* gene action. Two major factors have been suggested to modulate the competence of leaf tissues in *Arabidopsis* and tomato i.e. *KNOX*-independent regulation of

common target genes, and the regulation of tissue differentiation (Hay & Tsiantis, 2010). The interaction between different KNOX proteins and BEL1-like homeodomain (BELL) or OVATE proteins might, in turn, be involved in the regulation of these factors (Bolduc et al., 2014). In general, such differences in the phenotypes among homologous gene expression could also indicate divergence in the gene regulatory mechanisms of KNOX genes among plant species as radial adaptation.

Rice with *PvKN1* overexpression produced clumps of multiple shoots, but with no aberrant effect on root development. These phenotypes were consistent with the previous findings in rice where class I KNOX genes such as *OSH1* and *OSH3* were overexpressed (Nagasaki et al., 2001; Sentoku et al., 2000), further supporting the homology between these genes. The difference between switchgrass and rice in response to the expression of *PvKN1* might be attributed to differential *cis*-regulation of the gene by different regulatory elements, as suggested by Hay and Tsiantis. Taken together, these results hinted that the mechanism behind *PvKN1*-induced regulation of genes and downstream phenotypes is complex and may differ from species to species.

PvKN1 overexpression resulted in slight reduction in lignin deposition in the leaves accompanied by reduced expression of lignin biosynthetic genes including *PvC4H*, *PvCAD* and *PvCCR* (Figure 3-11a). This observation is generally in line with the previous reports in *Arabidopsis*, peach and *Populus* (Testone et al., 2012; Du et al., 2009; Mele et al., 2003). Interestingly, *PvKN1* overexpression appeared to downregulate the expression of a primary cell wall cellulose biosynthetic gene, *PvCESA1*, while upregulated the expression of hemicellulose biosynthetic gene (*PvCSLD1*) (Figure 3-11b) indicating that *PvKN1* may be involved in the regulation of cell wall biosynthetic genes. Unlike overexpression of *ARK2* in *Populus*, which downregulated the expression of three secondary cell wall cellulose biosynthetic genes (Du et al., 2009), no significant alteration of expression of these genes was observed in *PvKN1*-overexpressing transgenic switchgrass (Figure 3-11b). This difference might be attributed to the biological differences between the two species or the differences between the tissues analyzed. However, the regulatory mechanism of cell wall biosynthesis involving KN1 TFs is still less clear.

It was recently suggested that the maize *KN1* gene regulates the expression of genes responsible for the biosynthesis of lignin, and perhaps other cell wall components indirectly via the regulation of 'executive' genes (Bolduc et al., 2012). However, the direct regulation of a few lignin biosynthetic genes including *COMT* and *CCoAOMT* in *Arabidopsis* (Mele et al., 2003), and *CCoAOMT* in peach (Testone et al., 2012) has also been demonstrated. The GA-signalling pathway has been reported as one of the direct targets of *KN1* through which other plant morphological and biochemical characteristics are regulated (Bolduc et al., 2012). In this study, we showed that overexpression of *PvKN1* downregulated the expression of putative *PvGA20ox* (the rate-limiting enzyme in the GA biosynthesis pathway) while upregulated the expression of *C₂₀ PvGA2ox* genes (Figure 3-11c) suggesting that *PvKN1* might limit shoot elongation and lignification at early growth stages via regulation of GA biosynthetic and catabolic genes and GA signalling.

The link between *PvKN1* and the regulation of cellulose biosynthetic genes remains to be investigated (Figure 3-11b). One possible scenario is that *PvKN1* regulates the expression of cellulose biosynthetic genes through modulation of brassinosteroid (BR) signalling, which has been shown to positively regulate the expression of cellulose biosynthetic genes (Xie et al., 2011). This idea is supported by the recent report that maize *KN1* bound BR biosynthetic and catabolic genes (Bolduc et al., 2012) while *OSH1* regulates BR-signalling in rice (Tsuda et al., 2014).

The prospect of drastically decreasing lignin by ectopic expression of *KNOX* genes motivated our study in switchgrass. The challenge here, as with the case of transgenic switchgrass in which *MYB4* (Shen et al., 2012) and *miRNA156* (Fu et al., 2012) gene expression was changed, is to empirically determine the optimal TF transgene expression level that decreases lignin and possibly provides favorable plant architecture changes, without reducing shoot growth. Clearly, if *KN1* expression is too high, plants are dwarfed and abnormal. Indeed, relatively low *PvKN1* overexpression resulted in slightly reduced lignin deposition/biomass recalcitrance and improved sugar release efficiency without significantly affecting various growth parameters such as tiller height, tiller number, fresh and dry biomass weight (Figure 3-12, 3-14, Table 3-4). Therefore, the observed increase in sugar release in transgenic switchgrass with reduced lignin content highlight the potential biotechnological application of *PvKN1* for the improvement of biomass

characteristics for bioenergy feedstocks as well as forage grasses. Previous studies have suggested that expression of *KNOX* genes may enhance cytokinin levels (Frugis et al., 2001; Yanai et al., 2005; Jasinski et al., 2005) while cytokinin accumulation was shown to increase drought tolerance via coordinated regulation of carbon and nitrogen assimilation (Reguera et al., 2013; Peleg et al., 2011). Whether *PvKN1* expression plays similar roles should be the subject of future investigation as such traits may enhance the value of using transgenic lines as improved bioenergy feedstocks.

In summary, we identified a gene coding for class I KNOX TF in switchgrass, *PvKN1*, as a putative ortholog of maize *KN1*. Our results demonstrated that *PvKN1* may facilitate shoot elongation and lignification via transcriptional regulation of the GA-biosynthesis pathway. Moreover, we showed that *PvKN1* may also be involved in the regulation of the biosynthetic genes of other cell wall polymers (cellulose and hemicellulose) and hence play an important role in cell wall biosynthesis. *PvKN1*-overexpressing lines displaying normal growth phenotypes but with reduced recalcitrance to enzymatic saccharification could potentially be utilized for the improvement of lignocellulosic bioenergy feedstocks for biofuels or for the improvement of forage grass. *PvKN1* could also be potentially used in gene-stacking studies along with genes imparting novel traits, such as higher biomass yield, biotic or abiotic stress resistance to develop more improved bioenergy feedstock.

3.6 Acknowledgements

We thank Angela Ziebell, Erica Gjersing, Crissa Doeppke, and Melvin Tucker for their assistance with the cell wall characterization and Susan Holladay for her assistance with data entry into LIMS. We thank Wayne Parrot for providing the switchgrass SA37 clone. This work was supported by funding from the BioEnergy Science Center. The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. We also thank Tennessee Agricultural Experiment Station for providing partial financial support for WAW. We also thank JGI-DOE for the prepublication access to the switchgrass genome.

References

- Ambavaram MM, Krishnan A, Trijatmiko KR, and Pereira A. 2011. Coordinated activation of cellulose and repression of lignin biosynthesis pathways in rice. *Plant Physiology* 155:916-931.
- Bart RS, Chern M, Vega-Sanchez ME, Canlas P, Ronald PC. 2010. Rice Snl6, a cinnamoyl-CoA reductase-like gene family member, is required for NH1-mediated immunity to *Xanthomonas oryzae* pv. *oryzae*. *PLOS Genetics* 6:e1001123.
- Baxter HL, Mazarei M, Labbe N, Kline LM, Cheng Q, Windham MT, Mann DG, Fu C, Ziebell A, Sykes RW, Rodriguez Jr. M, Davis MF, Mielenz JR, Dixon RA, Wang ZY, Stewart CN Jr. 2014. Two-year field analysis of reduced recalcitrance transgenic switchgrass. *Plant Biotechnology Journal* 12: 914-924.
- Biemelt S, Tschiersch H, Sonnewald U. 2004. Impact of altered gibberellin metabolism on biomass accumulation, lignin biosynthesis, and photosynthesis in transgenic tobacco plants. *Plant Physiology* 135:254-265.
- Bolduc N, Hake S. 2009. The maize transcription factor KNOTTED1 directly regulates the gibberellin catabolism gene *GA2ox1*. *Plant Cell* 21:1647-1658.
- Bolduc N, Tyers RG, Freeling M, Hake S. 2014. Unequal redundancy in maize *knotted1* homeobox genes. *Plant Physiology* 164:229-238.
- Bolduc N, Yilmaz A, Mejia-Guerra MK, Morohashi K, O'Connor D, Grotewold E, Hake S. 2012. Unraveling the KNOTTED1 regulatory network in maize meristems. *Genes & Development* 26:1685-1690.
- Burris J, Mann DJ, Joyce B, Stewart CN Jr. 2009. An improved tissue culture system for embryogenic callus production and plant regeneration in switchgrass (*Panicum virgatum* L.). *BioEnergy Research* 2:267-274.

- Chai M, Bellizzi M, Wan C, Cui Z, Li Y, and Wang G-L. 2015. The NAC transcription factor OsSWN1 regulates secondary cell wall development in *Oryza sativa*. *Journal of Plant Biology* 58:44-51.
- Chuck G, Lincoln C, Hake S. 1996. KNAT1 induces lobed leaves with ectopic meristems when overexpressed in *Arabidopsis*. *Plant Cell* 8:1277-1289.
- Decker SR, Carlile M, Selig MJ, Doeppke C, Davis M, Sykes RW, Turner G, Ziebell A. 2012. Reducing the effect of variable starch levels in biomass recalcitrance screening. In: Himmel M (ed.), *Biomass conversion: methods and protocols*, Springer, New York, 181-195.
- Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research* 36:W465-469.
- Du J, Mansfield SD, Groover AT. 2009. The *Populus* homeobox gene *ARBORKNOX2* regulates cell differentiation during secondary growth. *Plant Journal* 60:1000-1014.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792-1797.
- Escamilla-Trevino LL, Shen H, Uppalapati SR, Ray T, Tang Y, Hernandez T, Yin Y, Xu Y, Dixon RA. 2010. Switchgrass (*Panicum virgatum*) possesses a divergent family of cinnamoyl CoA reductases with distinct biochemical properties. *New Phytologist* 185:143-155.
- Fornale S, Shi X, Chai C, Encina A, Irar S, Capellades M, Fuguet E, Torres JL, Rovira P, Puigdomenech P, Rigau J, Grotewold E, Gray J, Caparros-Ruiz D. 2010. ZmMYB31 directly represses maize lignin genes and redirects the phenylpropanoid metabolic flux. *Plant Journal* 64:633-644.
- Foster T, Yamaguchi J, Wong BC, Veit B, Hake S. 1999. Gnarley1 is a dominant mutation in the *knox4* homeobox gene affecting cell shape and identity. *Plant Cell* 11:1239-1252.

Freeling M. 1992. A conceptual framework for maize leaf development. *Developmental Biology* 153:44-58.

Frugis G, Giannino D, Mele G, Nicolodi C, Chiappetta A, Bitonti MB, Innocenti AM, Dewitte W, Van Onckelen H, Mariotti D. 2001. Overexpression of *KNAT1* in lettuce shifts leaf determinate growth to a shoot-like indeterminate growth associated with an accumulation of isopentenyl-type cytokinins. *Plant Physiology* 126:1370-1380.

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. *Proceedings of the National Academy of Science USA* 108:3803-3808.

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 Biotechnology Journal* 10:443-452.

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

Grewal D, Gill R, Gosal SS. 2006. Influence of antibiotic cefotaxime on somatic embryogenesis and plant regeneration in indica rice. *Biotechnology Journal* 1:1158-1162.

Groover AT, Mansfield SD, DiFazio SP, Dupper G, Fontana JR, Millar R, Wang Y. 2006. The *Populus* homeobox gene *ARBORKNOX1* reveals overlapping mechanisms regulating the shoot apical meristem and the vascular cambium. *Plant Molecular Biology* 61:917-932.

Guo AY, Zhu QH, Chen X, Luo JC: GSDS: a gene structure display server. *Yi Chuan* 2007, 29:1023-1026.

Hake S, Smith HM, Holtan H, Magnani E, Mele G, Ramirez J. 2004. The role of *KNOX* genes in plant development. *Annual Review of Cell and Developmental Biology* 20:125-151.

- Handakumbura PP, Hazen SP. 2012. Transcriptional regulation of grass secondary cell wall biosynthesis: playing catch-up with *Arabidopsis thaliana*. *Frontiers in Plant Science* 3:74.
- Hardin CF, Fu C, Hisano H, Xiao X, Shen H, Stewart CN Jr, Parrott W, Dixon RA, Wang ZY. 2013. Standardization of switchgrass sample collection for cell wall and biomass trait analysis. *BioEnergy Research* 6:1-8.
- Hareven D, Gutfinger T, Parnis A, Eshed Y, Lifschitz E. 1996. The making of a compound leaf: genetic manipulation of leaf architecture in tomato. *Cell* 84:735-744.
- Hay A, Jackson D, Ori N, Hake S. 2003. Analysis of the competence to respond to KNOTTED1 activity in *Arabidopsis* leaves using a steroid induction system. *Plant Physiology* 131:1671-1680.
- Hay A, Kaur H, Phillips A, Hedden P, Hake S, Tsiantis M. 2002. The gibberellin pathway mediates KNOTTED1-type homeobox function in plants with different body plans. *Current Biology* 12:1557-1565.
- Hay A, Tsiantis M. 2010. *KNOX* genes: versatile regulators of plant development and diversity. *Development* 137:3153-3165.
- Jackson D, Veit B, Hake S. 1994. Expression of maize *KNOTTED1* related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot. *Development* 120:405-413.
- Jasinski S, Piazza P, Craft J, Hay A, Woolley L, Rieu I, Phillips A, Hedden P, Tsiantis M. 2005. KNOX action in *Arabidopsis* is mediated by coordinate regulation of cytokinin and gibberellin activities. *Current Biology* 15:1560-1565.
- Kano-Murakami Y, Yanai T, Tagiri A, Matsuoka M. 1993. A rice homeotic gene, *OSH1*, causes unusual phenotypes in transgenic tobacco. *FEBS Letters* 334:365-368.
- Kerstetter R, Vollbrecht E, Lowe B, Veit B, Yamaguchi J, Hake S. 1994. Sequence analysis and expression patterns divide the maize *knotted1-like* homeobox genes into two classes. *Plant Cell* 6:1877-1887.

- Kerstetter RA, Laudencia-Chingcuanco D, Smith LG, Hake S. 1997. Loss-of-function mutations in the maize homeobox gene, *knotted1*, are defective in shoot meristem maintenance. *Development* 124:3045-3054.
- King ZR, Bray AL, Lafayette PR, Parrott WA. 2014. Biolistic transformation of elite genotypes of switchgrass (*Panicum virgatum* L.). *Plant Cell Reports* 33:313-322.
- Li E, Bhargava A, Qiang W, Friedmann MC, Forneris N, Savidge RA, Johnson LA, Mansfield SD, Ellis BE, Douglas CJ. 2012. The Class II KNOX gene *KNAT7* negatively regulates secondary wall formation in *Arabidopsis* and is functionally conserved in *Populus*. *New Phytologist* 194:102-115.
- Li R, Qu R. 2011. High throughput Agrobacterium-mediated switchgrass transformation. *Biomass & Bioenergy* 35:1046-1054.
- Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, Himmel M, Keller M, McMillan JD, Sheehan J, Wyman CE. 2008. How biotech can transform biofuels. *Nature Biotechnology* 26:169-172.
- Ma QH, Wang C, Zhu HH. 2011. *TaMYB4* cloned from wheat regulates lignin biosynthesis through negatively controlling the transcripts of both *cinnamyl alcohol dehydrogenase* and *cinnamoyl-CoA reductase* genes. *Biochimie* 93:1179-1186.
- Mann DG, Lafayette PR, Abercrombie LL, King ZR, Mazarei M, Halter MC, Poovaiah CR, Baxter H, Shen H, Dixon RA, Parrott WA, Stewart CN Jr. 2012. Gateway-compatible vectors for high-throughput gene functional analysis in switchgrass (*Panicum virgatum* L.) and other monocot species. *Plant Biotechnology Journal* 10:226-236.
- Matsuoka M, Ichikawa H, Saito A, Tada Y, Fujimura T, Kano-Murakami Y. 1993. Expression of a rice homeobox gene causes altered morphology of transgenic plants. *Plant Cell* 5:1039-1048.
- Mele G, Ori N, Sato Y, Hake S. 2003. The knotted1-like homeobox gene *BREVIPEDICELLUS* regulates cell differentiation by modulating metabolic pathways. *Genes Development* 17:2088-2093.

- Moore KJ, Moser LE, Vogel KP, Waller SS, Johnson BE, Pedersen JF. 1991. Describing and quantifying growth stages of perennial forage grasses. *Agronomy Journal* 83:1073-1077.
- Mukherjee K, Brocchieri L, Burglin TR. 2009. A comprehensive classification and evolutionary analysis of plant homeobox genes. *Molecular Biology and Evolution* 26:2775-2794.
- Muller J, Wang Y, Franzen R, Santi L, Salamini F, Rohde W. 2001. In vitro interactions between barley TALE homeodomain proteins suggest a role for protein-protein associations in the regulation of KNOX gene function. *Plant Journal* 27:13-23.
- Murashige T, Skoog F. 1962. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia Plantarum* 15:473-497.
- Nagasaki H, Sakamoto T, Sato Y, Matsuoka M. 2001. Functional analysis of the conserved domains of a rice KNOX homeodomain protein, OSH15. *Plant Cell* 13:2085-2098.
- Nishimura A, Aichi I, Matsuoka M. 2006. A protocol for Agrobacterium-mediated transformation in rice. *Nature Protocols* 1:2796-2802.
- Peleg Z, Reguera M, Tumimbang E, Walia H, Blumwald E. 2011. Cytokinin-mediated source/sink modifications improve drought tolerance and increase grain yield in rice under water-stress. *Plant Biotechnology Journal* 9:747-758.
- Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ, Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T. 2006. The path forward for biofuels and biomaterials. *Science* 311:484-489.
- Reguera M, Peleg Z, Abdel-Tawab YM, Tumimbang EB, Delatorre CA, Blumwald E. 2013. Stress-induced cytokinin synthesis increases drought tolerance through the coordinated regulation of carbon and nitrogen assimilation in rice. *Plant Physiology* 163:1609-1622.
- 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* 6:804-807.

Schaplinski TJ, Standaert RF, Engle NL, Martin MZ, Sangha AK, Parks JM, Smith JC, Samuel R, Jiang N, Pu Y, Ragauskas AJ, Hamilton CY, Fu C, Wang Z-Y, Davison BH, Dixon RA and Mielenz JR. 2012. Down-regulation of the *caffeic acid O-methyltransferase* gene in switchgrass reveals a novel monolignol analog. *Biotechnology for Biofuels* 5:71.

Schmer MR, Vogel KP, Mitchell RB, Perrin RK. 2008. Net energy of cellulosic ethanol from switchgrass. *Proceedings of the National Academy of Science USA* 105:464-469.

Selig MJ, Tucker MP, Sykes RW, Reichel KL, Brunecky R, Himmel ME, Davis MF, Decker SR. 2010. Biomass recalcitrance screening by integrated high throughput hydrothermal pretreatment and enzymatic saccharification. *Industrial Biotechnology* 6:104-111.

Sentoku N, Sato Y, Matsuoka M. 2000. Overexpression of rice *OSH* genes induces ectopic shoots on leaf sheaths of transgenic rice plants. *Developmental Biology* 220:358-364.

Shen H, Fu C, Xiao X, Ray T, Tang Y, Wang Z, Chen F. 2009. Developmental control of lignification in stems of lowland switchgrass variety Alamo and the effects on saccharification efficiency. *BioEnergy Research* 2:233-245.

Shen H, He X, Poovaiah CR, Wuddineh WA, Ma J, Mann DG, Wang H, Jackson L, Tang Y, Stewart CN Jr., Chen F, Dixon RA. 2012. Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks. *New Phytologist* 193:121-136.

Shen H, Mazarei M, Hisano H, Escamilla-Trevino L, Fu C, Pu Y, Rudis MR, Tang Y, Xiao X, Jackson L, Li G, Hernandez T, Chen F, Ragauskas AJ, Stewart CN Jr., Wang ZY, Dixon RA. 2013. A genomics approach to deciphering lignin biosynthesis in switchgrass. *Plant Cell* 25:4342-4361.

Shen H, Poovaiah CR, Ziebell A, Tschaplinski TJ, Pattathil S, Gjersing E, Engle NL, Katahira R, Pu Y, Sykes R, Chen F, Ragauskas AJ, Mielenz JR, Hahn MG, Davis M, Stewart CN Jr., Dixon RA. 2013. Enhanced characteristics of genetically modified switchgrass (*Panicum virgatum* L.) for high biofuel production. *Biotechnology for Biofuels* 6:71.

Sykes RW, Yung M, Novaes E, Kirst M, Peter G, Davis M. 2009. High-throughput screening of plant cell-wall composition using pyrolysis molecular beam mass spectroscopy. *Biofuels* Springer, 169-183.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 30:2725-2729.

Testone G, Bruno L, Condello E, Chiappetta A, Bruno A, Mele G, Tartarini A, Spano L, Innocenti AM, Mariotti D, Bitonti MB, Giannino D. 2008. Peach [*Prunus persica* (L.) Batsch] *KNOPE1*, a class 1 KNOX orthologue to *Arabidopsis* BREVIPEDICELLUS/KNAT1, is misexpressed during hyperplasia of leaf curl disease. *Journal of Experimental Botany* 59:389-402.

Testone G, Condello E, Verde I, Nicolodi C, Caboni E, Dettori MT, Vendramin E, Bruno L, Bitonti MB, Mele G, Giannino D. 2012. The peach (*Prunus persica* L. Batsch) genome harbours 10 KNOX genes, which are differentially expressed in stem development, and the class 1 KNOPE1 regulates elongation and lignification during primary growth. *Journal of Experimental Botany* 63:5417-5435.

Thomas SG, Rieu I, Steber CM. 2005. Gibberellin metabolism and signalling. *Vitamines & Hormones* 72:289-338.

Tioni MF, Gonzalez DH, Chan RL. 2003. Knotted1-like genes are strongly expressed in differentiated cell types in sunflower. *Journal of Experimental Botany* 54:681-690.

Townsley BT, Sinha NR, Kang J. 2013. *KNOX1* genes regulate lignin deposition and composition in monocots and dicots. *Frontiers in Plant Science* 4:121.

Tschaplinski TJ, Standaert RF, Engle NL, Martin MZ, Sangha AK, Parks JM, Smith JC, Samuel R, Jiang N, Pu Y, Ragauskas AJ, Hamilton XY, Fu C, Wang ZY, Davison BH, Dixon RA, Mielenz JR. 2012. Down-regulation of the *caffeic acid O-methyltransferase* gene in switchgrass reveals a novel monolignol analog. *Biotechnology for Biofuels* 5:71.

- Tsuda K, Kurata N, Ohyanagi H, Hake S. 2014. Genome-wide study of KNOX regulatory network reveals brassinosteroid catabolic genes important for shoot meristem function in rice. *Plant Cell* 26:3488-3500.
- Vollbrecht E, Veit B, Sinha N, Hake S. 1991. The developmental gene *Knotted-1* is a member of a maize homeobox gene family. *Nature* 350:241-243.
- Wuddineh WA, Mazarei M, Zhang J, Poovaiah CR, Mann DGJ, Ziebell A, Sykes R, Davis M, Udvardi M, and Stewart CN Jr. 2015. Identification and overexpression of *gibberellin 2-oxidase* (GA2ox) in switchgrass (*Panicum virgatum* L.) for improved plant architecture and reduced biomass recalcitrance. *Plant Biotechnology Journal* 13:636-647.
- Xie L, Yang C, Wang X. 2011. Brassinosteroids can regulate cellulose biosynthesis by controlling the expression of *CESA* genes in *Arabidopsis*. *Journal of Experimental Botany* 62:4495-4506.
- Xu B, Escamilla-Trevino LL, Sathitsuksanoh N, Shen Z, Shen H, Zhang YH, Dixon RA, Zhao B. 2011. Silencing of *4-coumarate:coenzyme A ligase* in switchgrass leads to reduced lignin content and improved fermentable sugar yields for biofuel production. *New Phytologist* 192:611-625.
- Yanai O, Shani E, Dolezal K, Tarkowski P, Sablowski R, Sandberg G, Samach A, Ori N. 2005. *Arabidopsis* KNOXI proteins activate cytokinin biosynthesis. *Current Biology* 15:1566-1571.
- Yuan JS, Tiller KH, Al-Ahmad H, Stewart NR, Stewart CN Jr. 2008. Plants to power: bioenergy to fuel the future. *Trends in Plant Science* 13:421-429.
- 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 Journal*, 74:160-173.

Zhao Q, Gallego-Giraldo L, Wang H, Zeng Y, Ding SY, Chen F, Dixon RA. 2010. A NAC transcription factor orchestrates multiple features of cell wall development in *Medicago truncatula*. *Plant Journal* 63:100-114.

Zhao XY, Zhu DF, Zhou B, Peng WS, Lin JZ, Huang XQ, He RQ, Zhuo YH, Peng D, Tang DY, Li MF, Liu XM. 2010. Overexpression of the *AtGA2ox8* gene decreases the biomass accumulation and lignification in rapeseed (*Brassica napus* L.). *Journal of Zhejiang University Science B* 11:471-481.

Zhong R, Lee C, Mccarthy RL, Reeves CK, Jones EG, and Ye Z-H. 2011. Transcriptional activation of secondary wall biosynthesis by rice and maize NAC and MYB transcription factors. *Plant & Cell Physiology* 52:1856-1871.

Zhong R, Ye ZH. 2014. Complexity of the transcriptional network controlling secondary wall biosynthesis. *Plant Science* 229:193-207.

Zhou J, Lee C, Zhong R, and Ye ZH. 2009. MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in *Arabidopsis*. *Plant Cell* 21:248-266.

Appendix

Figures and tables:

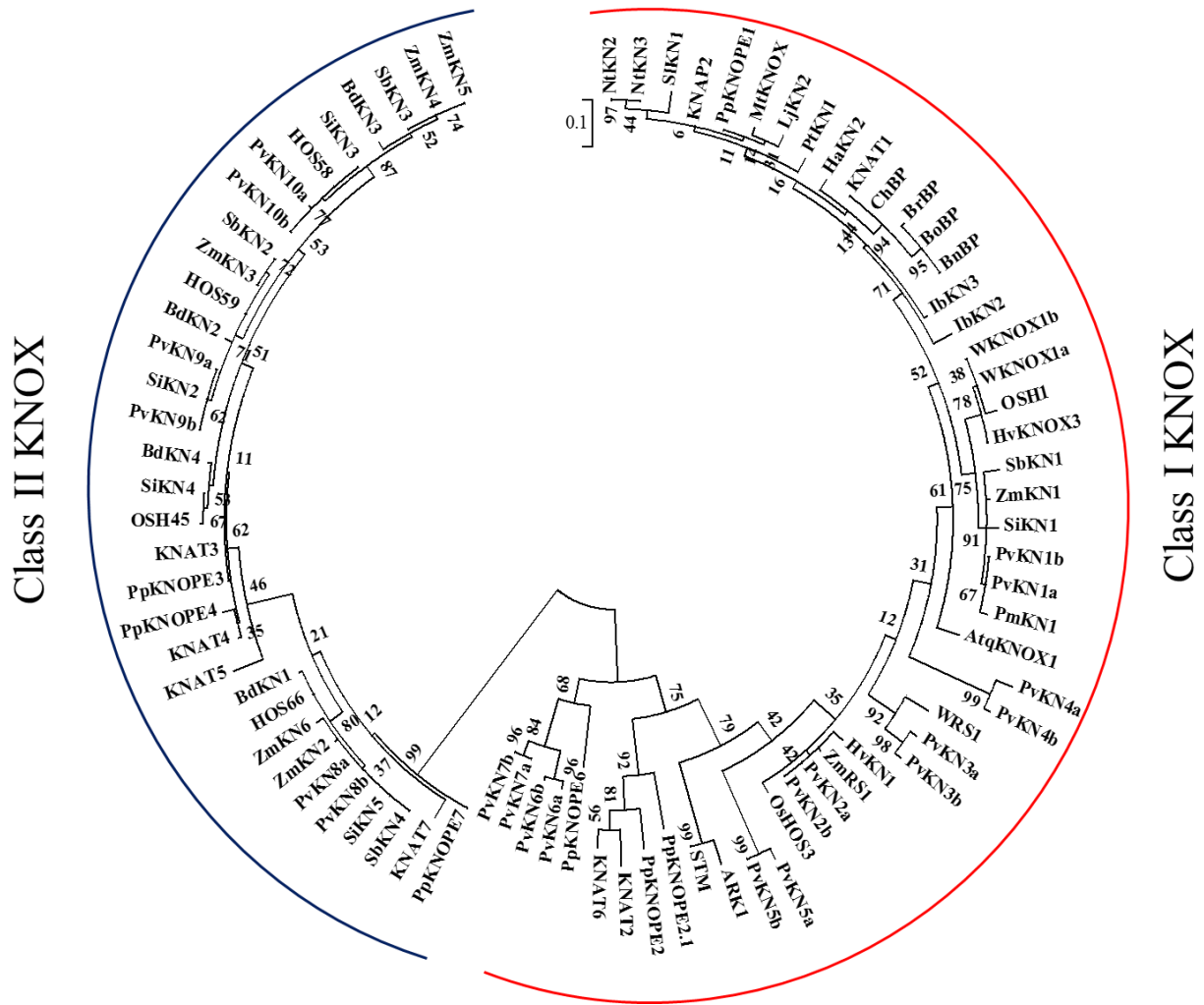


Figure 3-1 Maximum likelihood tree of KNOX family of transcription factors (TFs) for the deduced amino acid sequences from switchgrass along with already-characterized proteins from both monocots and dicots. The sequences were aligned using MUSCLE program (Edgar, 2004) and the alignment was curated by Gblocks at the phylogeny.fr (Dereeper et al., 2008) to eliminate poorly aligned positions and divergent regions. The tree was constructed by maximum likelihood procedure using MEGA6.0 program (Tamura et al., 2013). Analysis using 1000 bootstrap replicates was performed. The numbers on branches indicate bootstrap support values. The scale bar shows 0.1 amino acid substitutions per site. Names of the species and the locus name or GenBank accession numbers of the sequences used in this tree are listed in Table 3-5.

Figure 3-2 Multiple amino acid sequence alignment of the C-termini of class I: PvKN1a, PvKN2b, PvKN6b, ZmKN1, KNAT1 and SHOOT MERISTEMLESS (STM) and class II (PvKN8b, PvKN9a, PvKN10a and KNAT7) family of KNOX TFs. The conserved domains are underlined in red; the strictly conserved amino acid residues are indicated in bold. The amino acid residues specific to class I KNOX TFs are highlighted in dark grey while those specific to class II are highlighted in light grey. The MSA was constructed using the amino acid sequences of respective genes by MUSCLE program (Edgar, 2004).

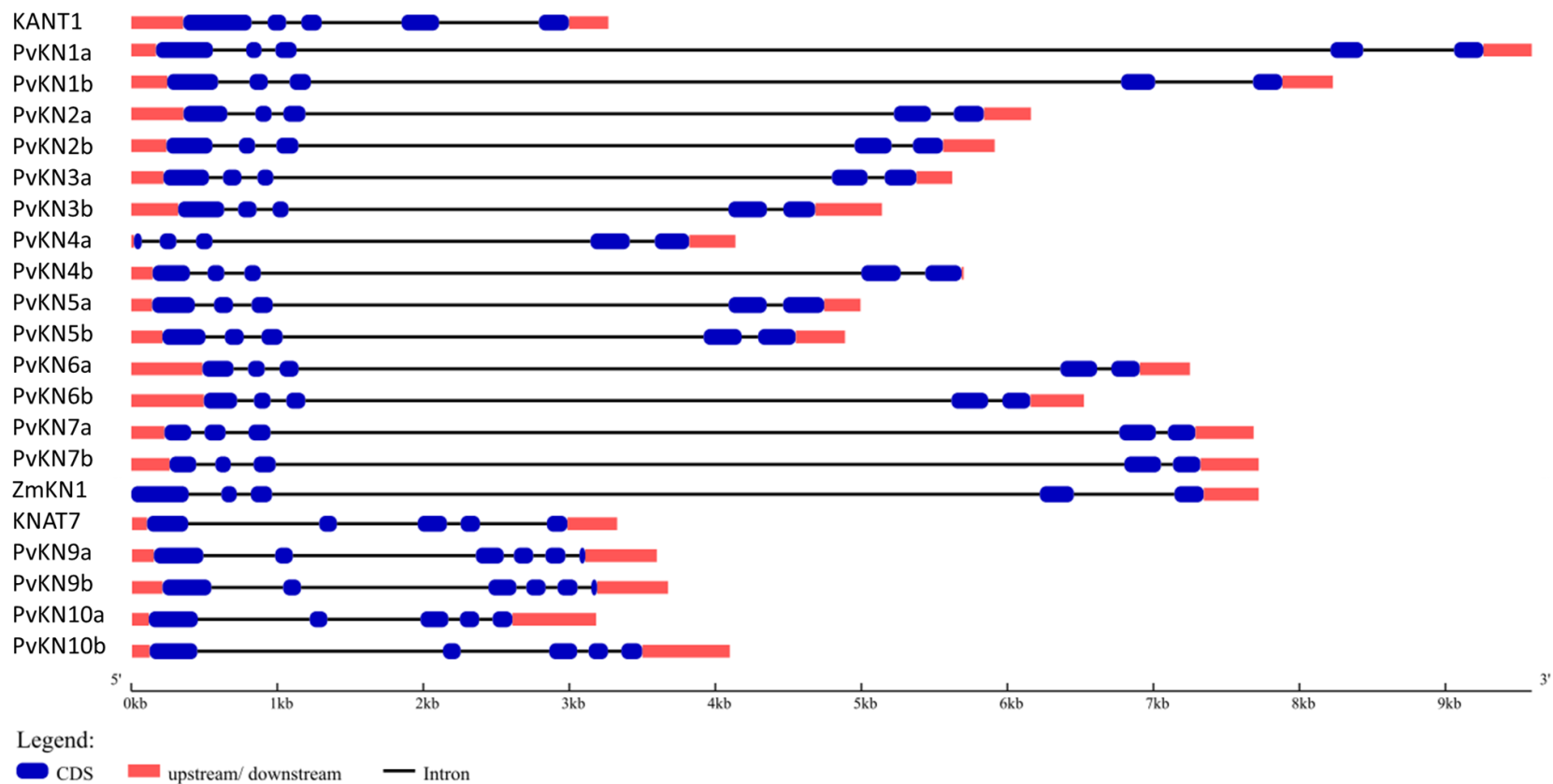


Figure 3-3 Gene structures of class I and II KNOX transcription factor family coding genes from *Arabidopsis*, switchgrass (*Panicum virgatum*) and maize (*Zea mays*). The coding DNA sequences (CDS) and the untranslated regions (UTR) are shown by filled blue and light red boxes, respectively. The introns are shown by thick black lines. The gene features were visualized by the gene structure display server (Guo et al., 2007).

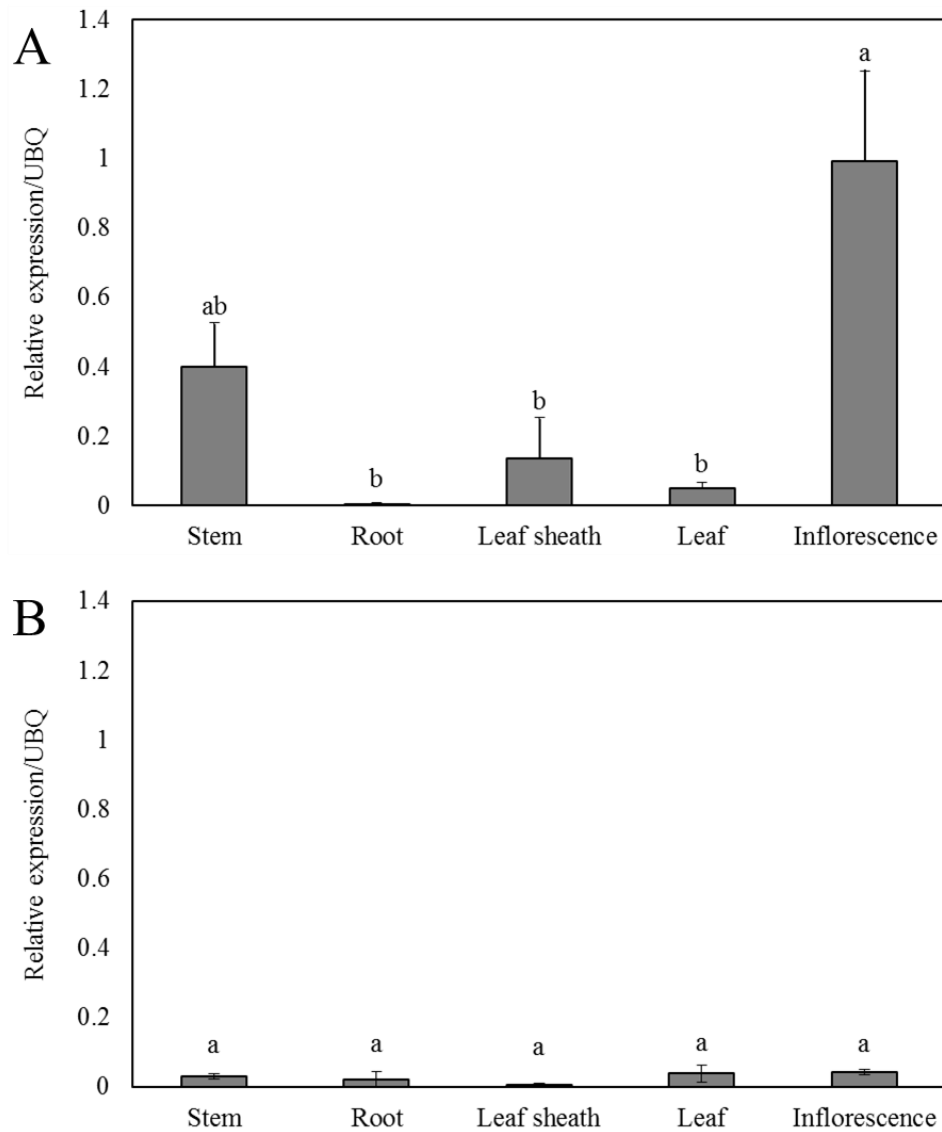


Figure 3-4 Expression patterns of *PvKN1a* (A) and *PvKN1b* (B) in different plant tissues as determined by qRT-PCR. Plant samples for RNA extraction used in the qRT-PCR experiments were collected at R1 (reproductive stage 1) developmental stage. The dissociation curve for the qRT-PCR products showed that the primers were gene-specific. The relative levels of transcripts were normalized to ubiquitin (UBQ). Bars represent mean values of 3 replicates $\pm$ standard error. Bars represented by different letters are significantly different at $P \leq 0.05$ as tested by LSD method with SAS software (SAS Institute Inc.).

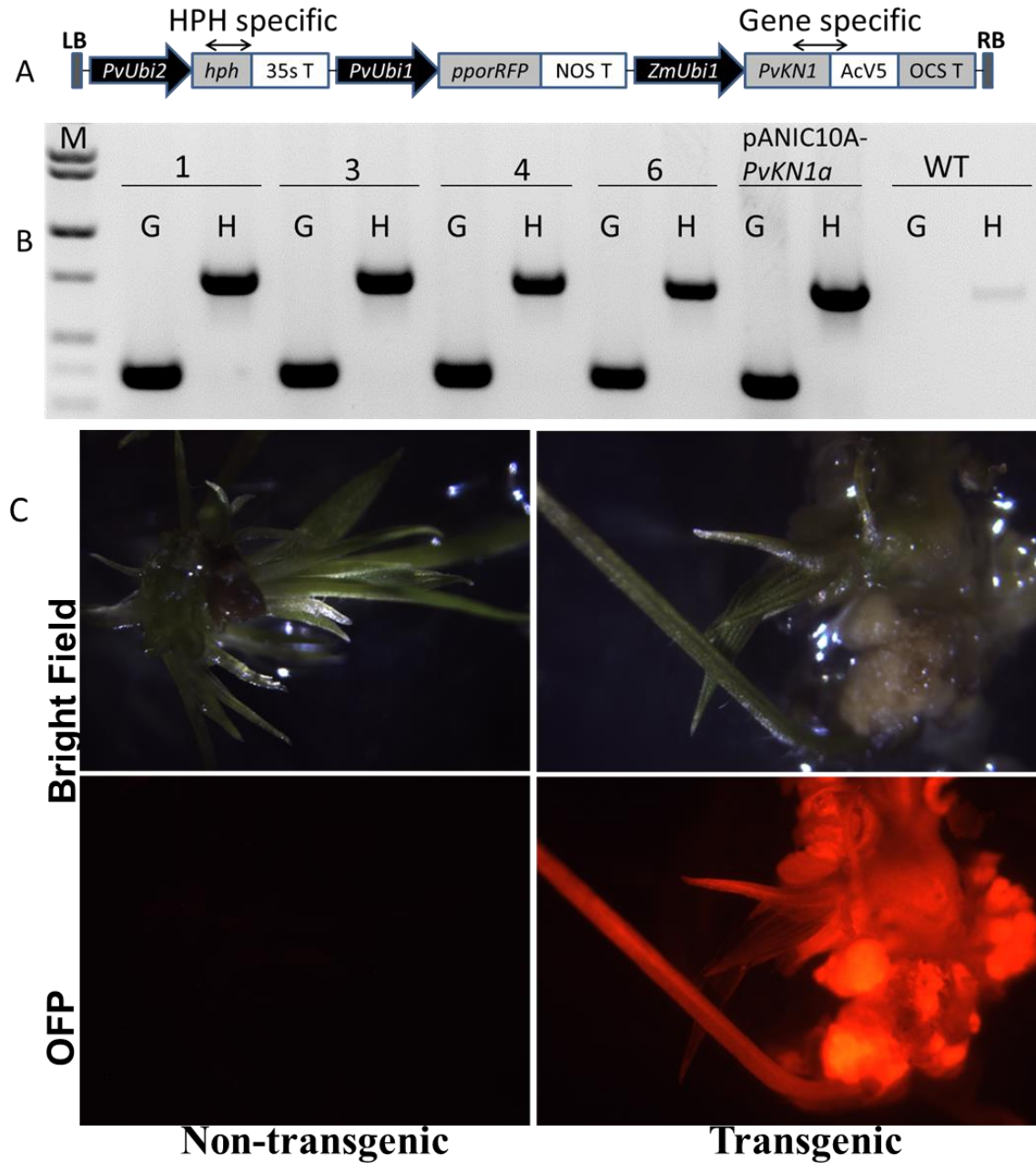


Figure 3-5 Molecular characterization of transgenic switchgrass plants overexpressing the *PvKN1a* gene. A) The T-DNA from pANIC10A vector used for overexpression of *PvKN1a* in transgenic switchgrass. B) Genomic PCR for assaying transgene (G) insertion and the hygromycin-resistance (H) genes in putative transgenic lines. C) Orange fluorescence protein (*ppORRFP*; OFP) visualization in transgenic plants compared to the non-transgenic control.

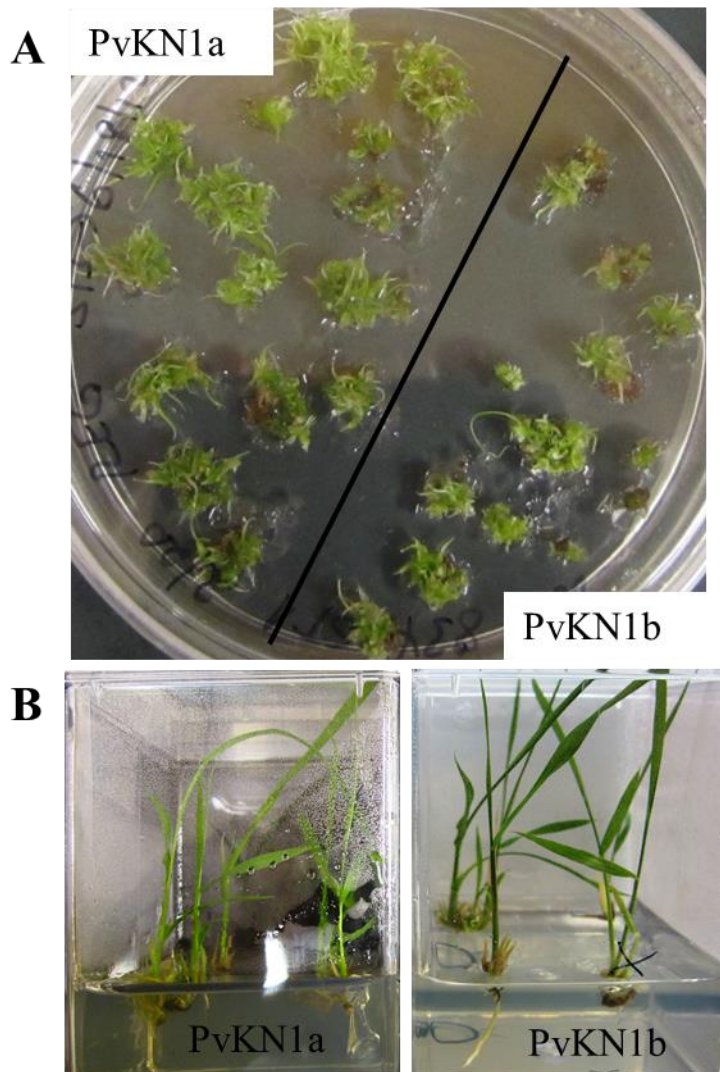


Figure 3-6 Comparison of morphological phenotypes in transgenic switchgrass overexpressing *PvKN1a* and *PvKN1b*. Transformants overexpressing *PvKN1a* and *PvKN1b* after shoot regeneration (A) and 9-weeks after transfer to rooting media.

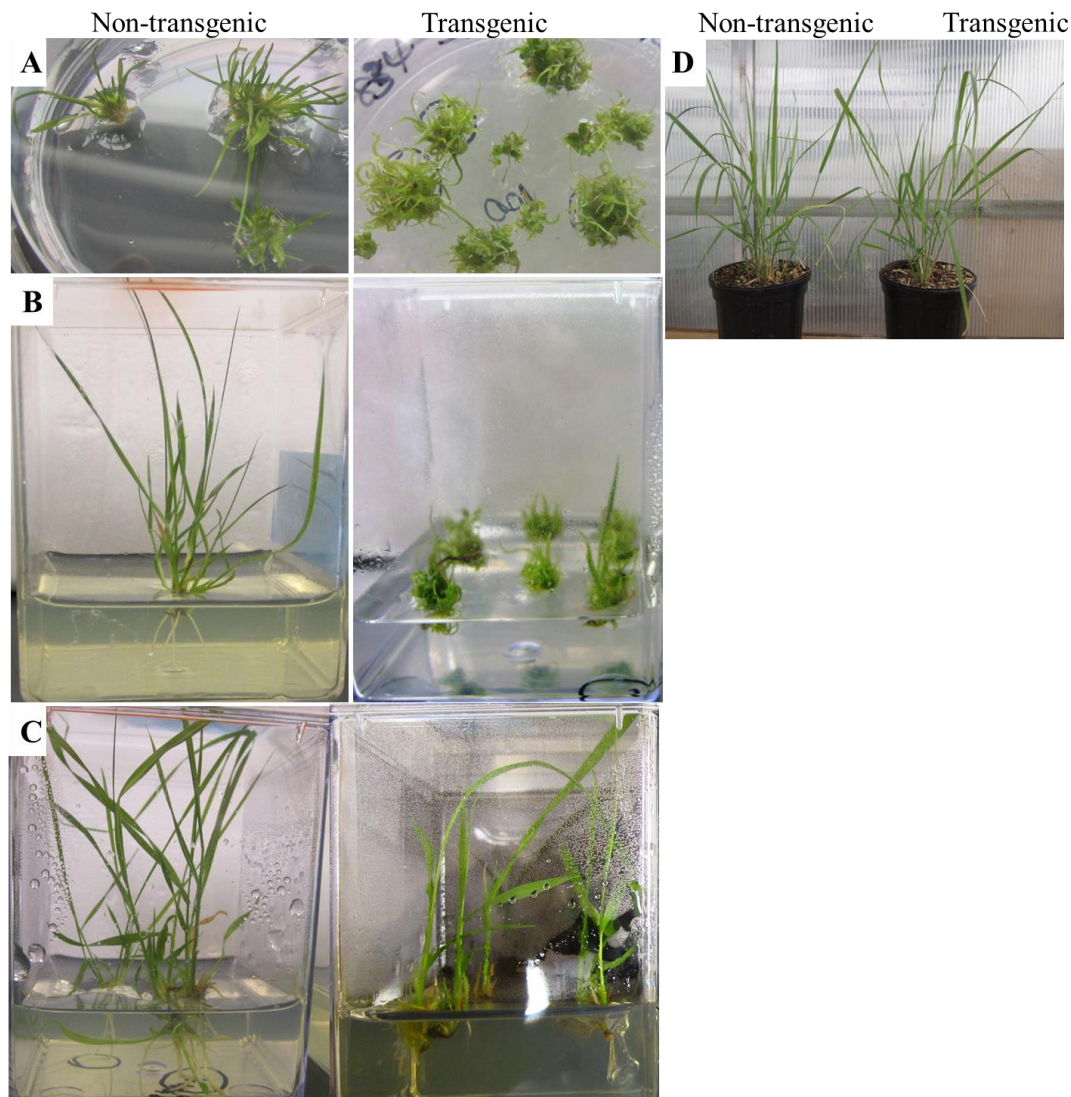


Figure 3-7 Shoot development and leaf morphology of transgenic *PvKN1a*-overexpressing switchgrass (ST1) plants. A) Transgenic plants overexpressing *PvKN1a* exhibited reduced shoot elongation and curly leaves compared to the normal phenotype in non-transgenic lines. B) Transgenic plants with substantially reduced growth compared to the non-transgenic lines at 5 weeks after incubation on the rooting media. C) Transgenic plants with elongated shoots compared to the non-transgenic lines at 9 weeks after incubation on the rooting media. D) Morphological phenotypes in 3-mo-old transgenic lines overexpressing *PvKN1a* compared to non-transgenic lines.

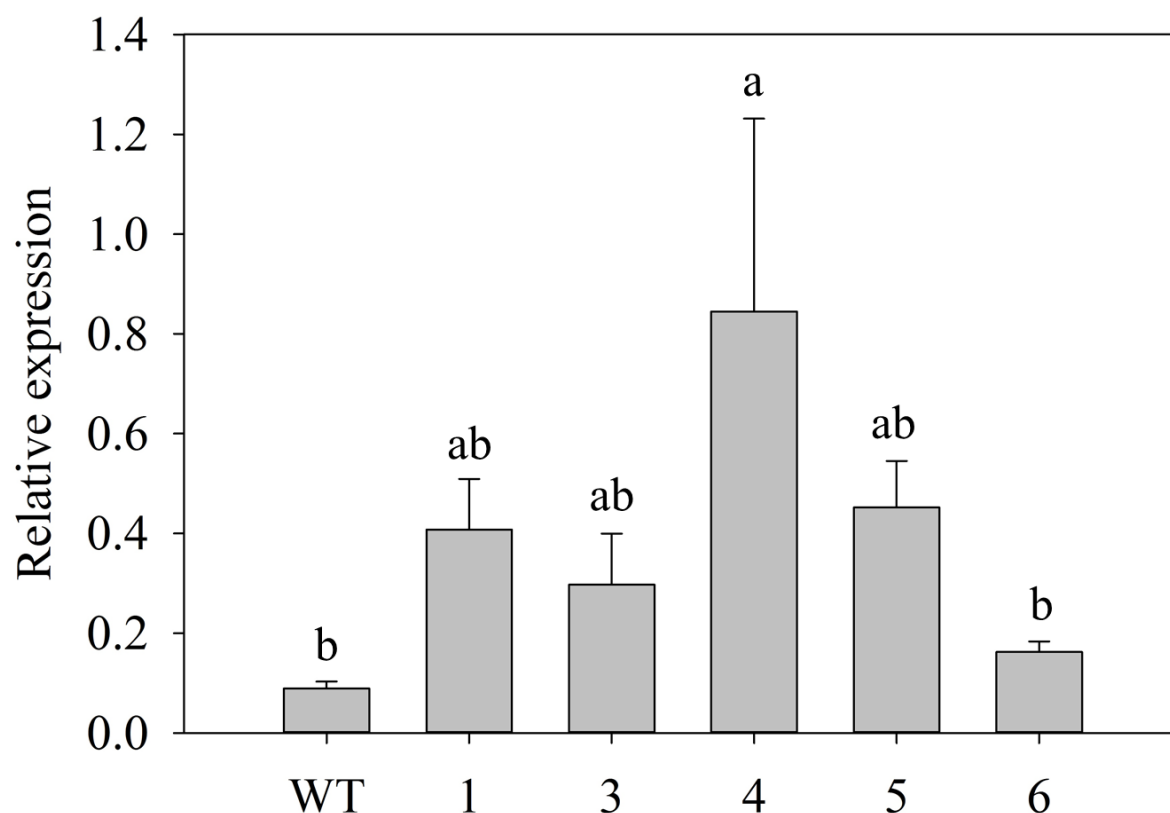


Figure 3-8 Relative transcript levels of the *PvKN1a* in transgenic and non-transgenic (WT) switchgrass (ST1) plants. The expression analysis was done using RNA derived from the shoot tips at E4 developmental stage. The dissociation curve for the qRT-PCR products showed that the primers were gene-specific. The relative levels of transcripts were normalized to ubiquitin (UBQ). Bars represent mean values of 3 replicates $\pm$ standard error. Values represented by different letters are significantly different at $P \leq 0.05$ as tested by LSD grouping with SAS software (SAS Institute Inc.).

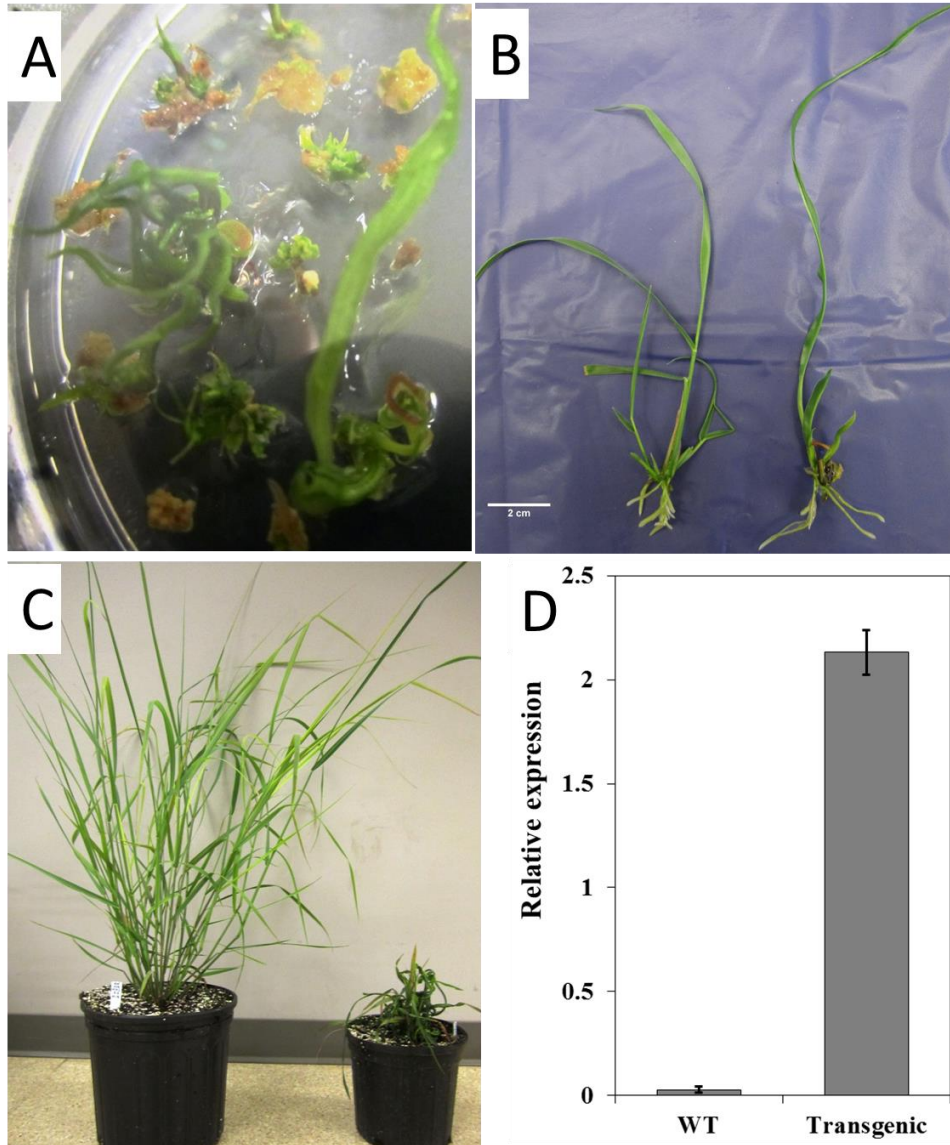


Figure 3-9 Transgenic switchgrass (SA37) plants overexpressing *PvKN1a* showing abnormal phenotypes throughout various stages of plant development. A) Transformants at early developmental stage. B) Rooted seedlings of non-transgenic (left) and transgenic (right) lines. C) Mature non-transgenic (left) and transgenic (right) lines 4 months after transferred to soil. D) Relative transcript levels of *PvKN1* in transgenic and non-transgenic control (WT) lines. The relative levels of transcripts were normalized to a ubiquitin (UBQ) housekeeping gene. Bars represent mean values of 3 replicates $\pm$ standard error.

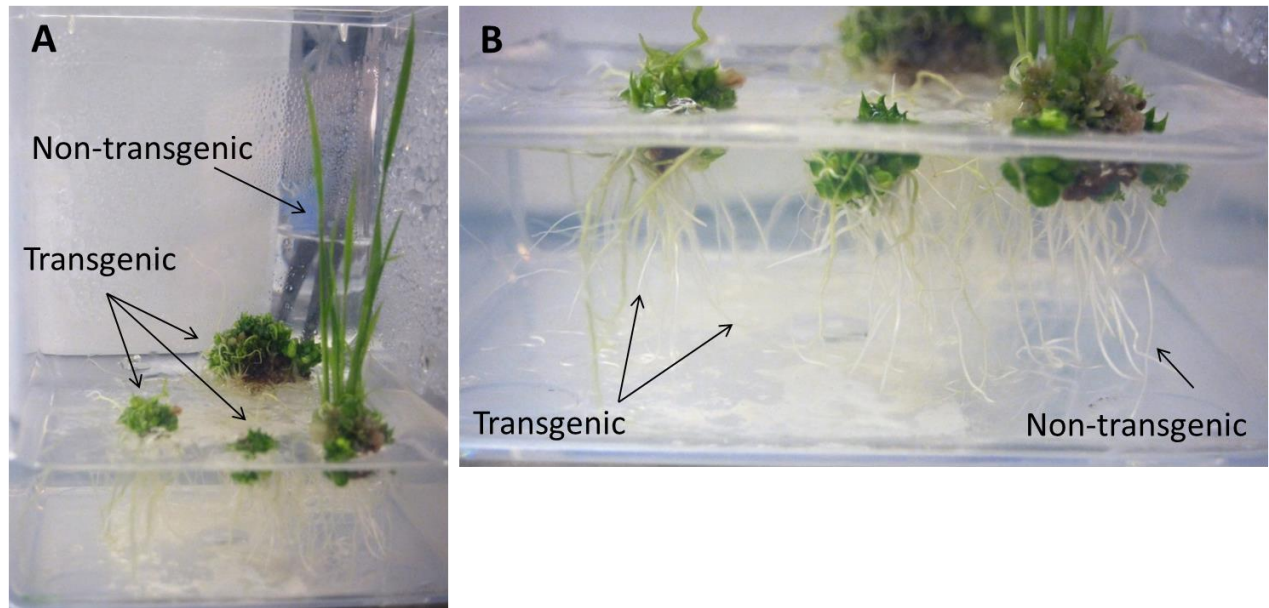


Figure 3-10 Overexpression of *PvKN1* in rice altered shoot but not root development. (A). Transgenic rice showing abnormal shoot development compared to the normal phenotype in the non-transgenic plants. (B). Transgenic rice showing normal root development as well as the non-transgenic plants.

Figure 3-11 The relative expression of putative target genes of *PvKN1* in transgenic switchgrass (line 4) overexpressing *PvKN1* vs the non-transgenic (WT) plants as determined by qRT-PCR. A) Lignin biosynthetic genes. B) Cell wall biosynthetic genes. C) Gibberellin (GA) catabolic and biosynthetic genes. The relative levels of transcripts were normalized to ubiquitin (UBQ). Asterisks indicate significant differences from non-transgenic control plants at $P \leq 0.05$ as determined by PROC TTEST procedure using SAS software (SAS Institute Inc.). Bars represent mean values of 3 replicates $\pm$ standard error. The lignin and GA biosynthetic and catabolic genes were as described before in Wuddineh et al. (2015). The cellulose and hemicellulose biosynthetic genes were labelled according to the naming from the closest rice or *Arabidopsis* homologs used in Ambavaram et al. (2011) (Figure 3-15). (*Pv*)*4CL* (4-coumarate: CoA ligase); (*Pv*)*C3H* (coumaroyl shikimate 3-hydroxylase); (*Pv*)*C4H* (coumaroyl shikimate 4-hydroxylase); (*Pv*)*CAD* (cinnamyl alcohol dehydrogenase); (*Pv*)*CCR1* (cinnamoyl CoA reductase1); (*Pv*)*COMT* (caffeic acid 3-O-methyltransferase); (*Pv*)*F5H* (ferulate 5-hydroxylase); (*Pv*)*PAL* (phenylalanine ammonia-lyase); (*Pv*)*LAC1* (laccase1); (*Pv*)*CESA* (cellulose synthase); (*Pv*)*CSL* (cellulose synthase-like).

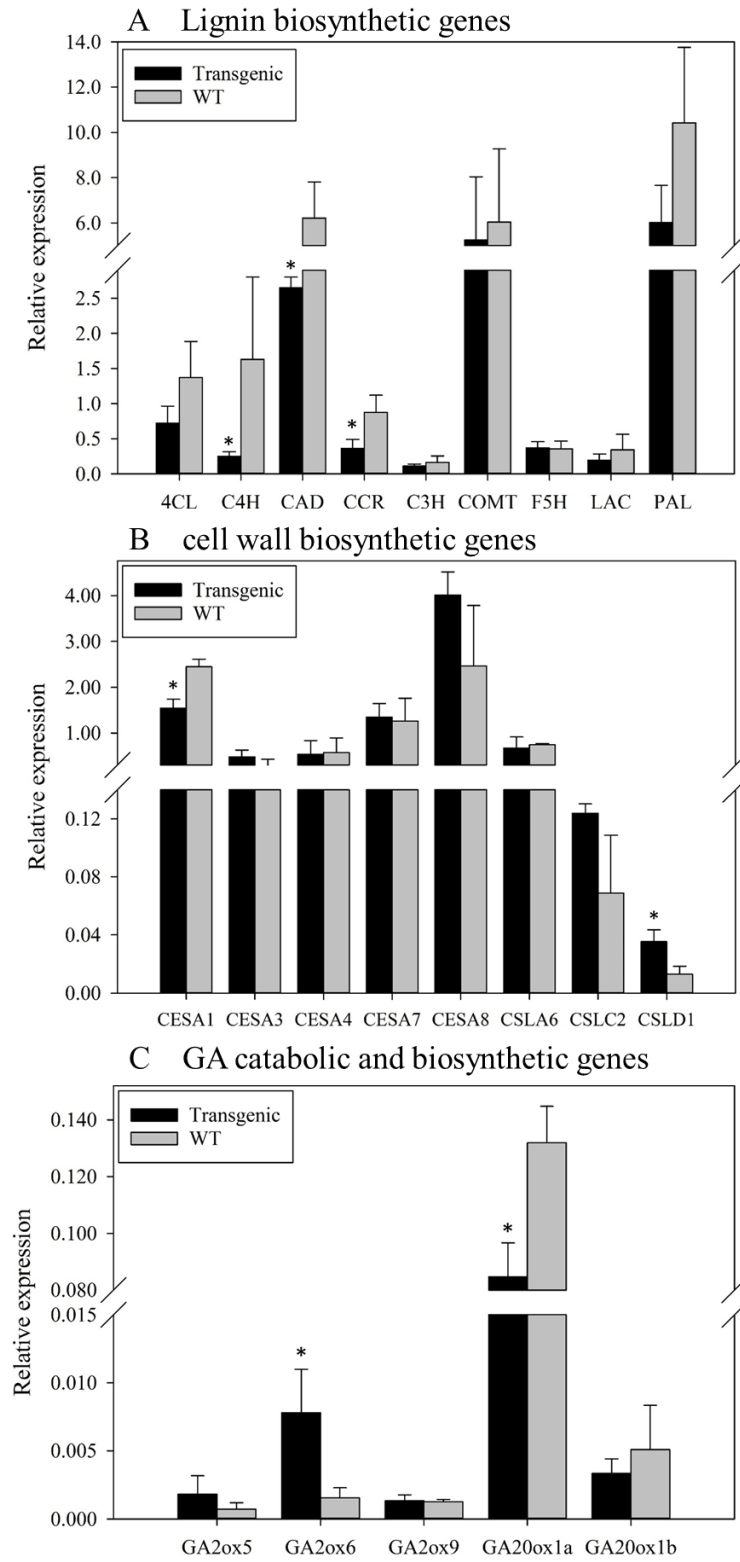


Figure 3.11 Continued.

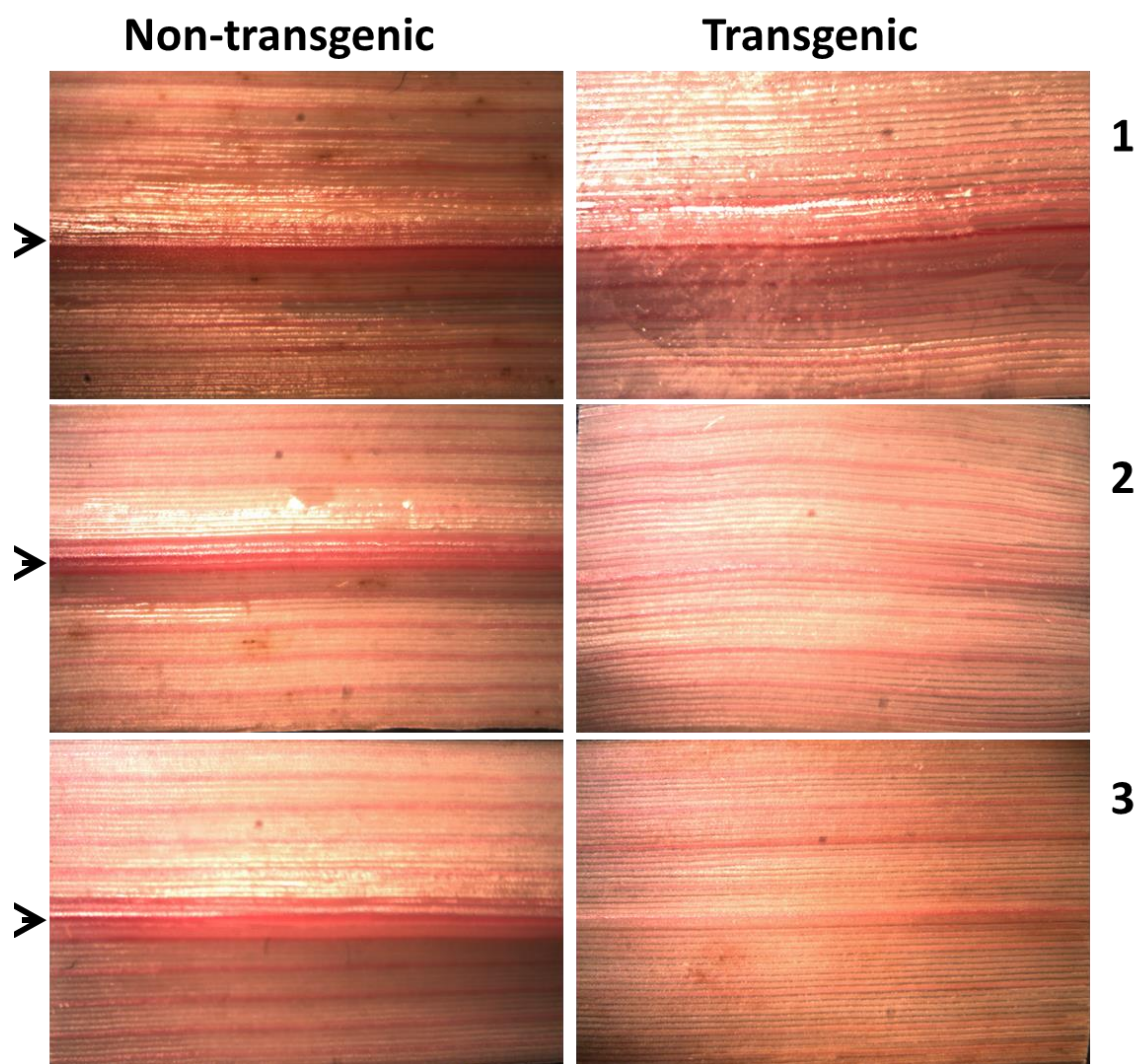


Figure 3-12 Histochemical detection of lignin in leaves of *PvKN1* overexpressing and non-transgenic control switchgrass lines using light microscopy. The phloroglucinol-HCl staining was done on leaves from three independent tillers indicated by numbers (1-3). The images were taken at 1x magnification. The arrows indicate midribs of leaves.

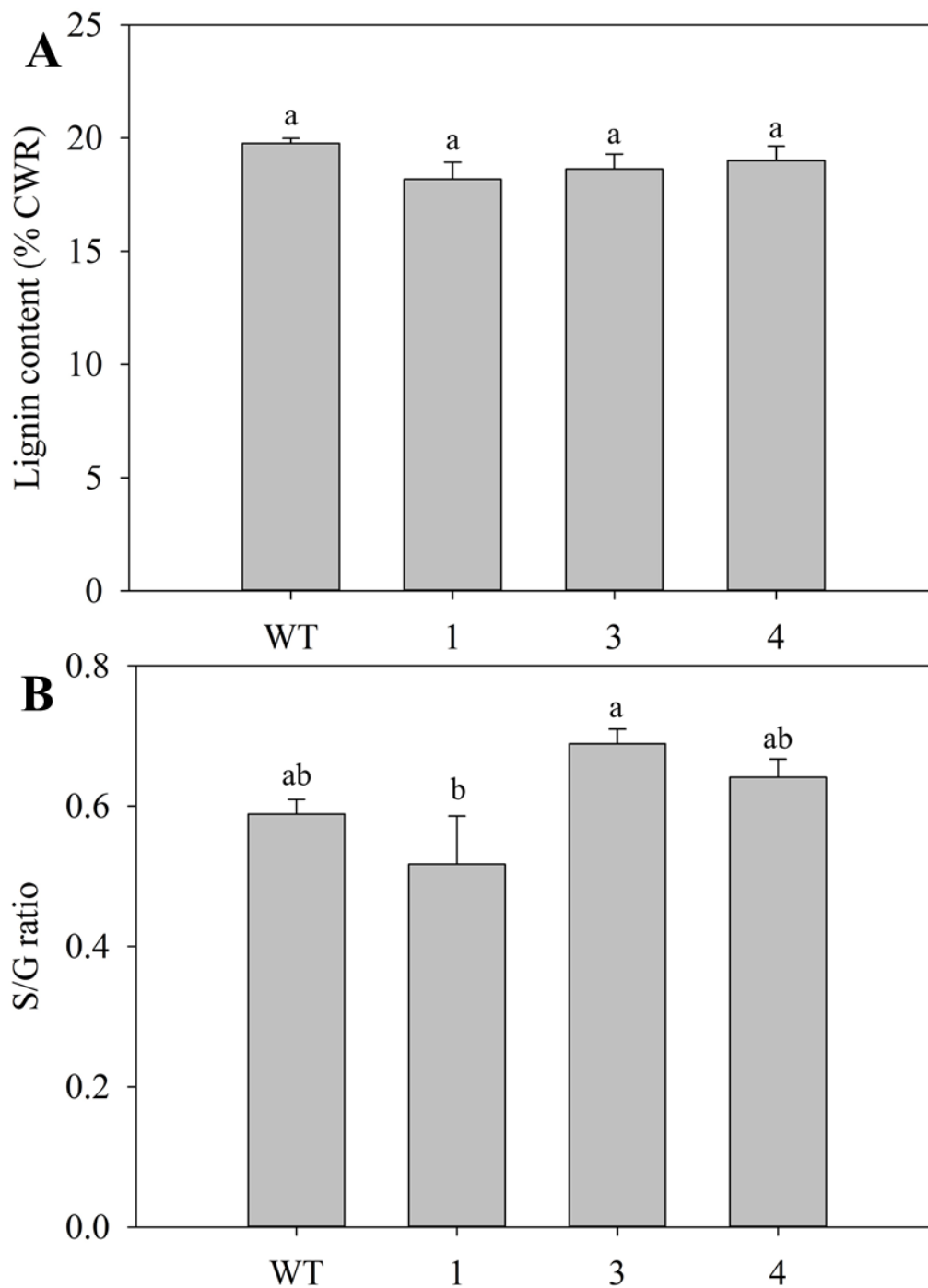


Figure 3-13 Lignin content (A) and S/G ratio (B) of transgenic and non-transgenic (WT) switchgrass lines as determined via py-MBMS. All data represent the average of the replicates $\pm$ standard deviation. Bars represented by different letters are significantly different at $P \leq 0.05$ as tested by LSD method with SAS software (SAS Institute Inc.). CWR, cell wall residues.

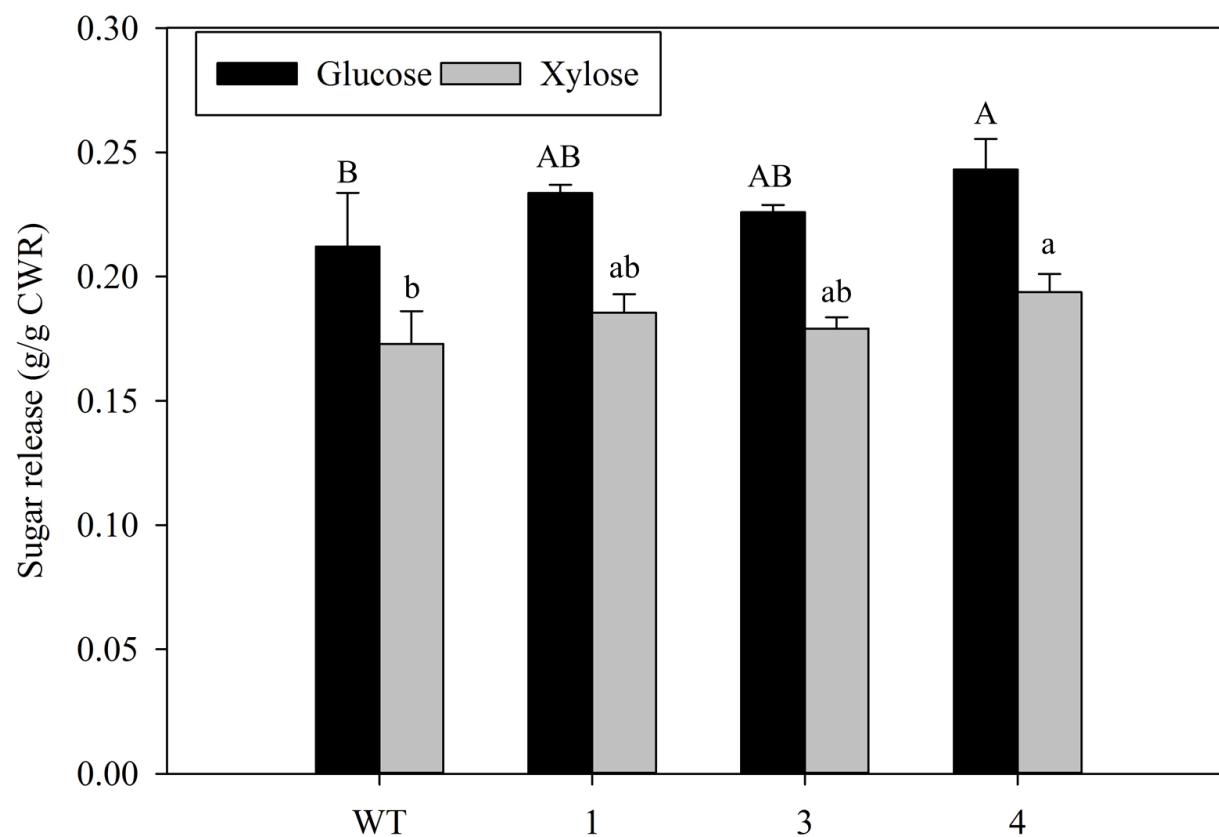


Figure 3-14 Sugar release by enzymatic hydrolysis in transgenic and non-transgenic control (WT) switchgrass lines. All data are means $\pm$ standard deviation (n=3). Values represented by different letters are significantly different at $P \leq 0.05$ as tested by LSD method with SAS software (SAS Institute Inc.). CWR, cell wall residues.

Figure 3-15 Dendrogram showing the relatedness of the PvKN1 target genes [PvLAC1 (A), GA20ox1 (B) and cellulose and hemicellulose biosynthetic genes (C)] indicated in Figure 3-11 with the known homologs in monocots and *Arabidopsis*. Cluster analysis was performed using the deduced amino acid sequences from switchgrass along with already-characterized proteins from monocots and *Arabidopsis*. The sequences were aligned using MUSCLE program (Edgar, 2004). The tree was constructed by maximum likelihood procedure using MEGA6.0 program (Tamura et al., 2013). Analysis using 1000 bootstrap replicates was performed. The numbers on branches indicate bootstrap support values. The scale bar shows 0.1 or 0.2 amino acid substitutions per site. The untranscript/locus IDs of switchgrass genes are listed in Table 3-5. Locus IDs of the genes from barley, rice, maize and *Arabidopsis* are: HvGA20ox1 (AAT49058.1), ZmGA20ox1 (NP_001241783.1), OsGA20ox1 (LOC_Os03g63970.1), OsGA20ox1-B (LOC_Os07g07420.1), OsLAC1 (LOC_Os01g62480.1), OsLAC22 (LOC_Os11g48060.1), OsLAC23 (LOC_Os12g01730.1), OsCSLA1 (LOC_Os02g09930.1), OsCESA1 (LOC_Os05g08370.1), OsCESA3 (LOC_Os07g24190.1), OsCESA7 (Os10g32980.1), OsCESA8 (LOC_Os07g10770.1), OsCESA4 (Os01g54620.1), OsCSLD2 (LOC_Os06g02180.1), OsCSLC10 (LOC_Os07g03260.1), AtGA20ox1 (AT4G25420), AtGA20ox2 (AT5G51810), AtGA20ox3 (AT5G07200), AtGA20ox4 (AT1G60980), AtCESA1 (AT4G32410), AtCESA2 (AT4G39350.1), AtCESA3 (AT5G05170), AtCESA7 (AT5G17420.1), AtCESA8 (AT4G18780), AtCSLA2 (AT5G22740), AtCSLC12 (AT4G07960), AtCSLD3 (AT3G03050), AtLAC1 (AT1G18140), AtLAC2 (AT2G29130), AtLAC3 (AT2G30210), AtLAC4 (AT2G38080), AtLAC5 (AT2G40370), AtLAC6 (AT2G46570), AtLAC7 (AT3G09220), AtLAC8 (AT5G01040), AtLAC10 (AT5G01190), AtLAC11 (AT5G03260), AtLAC12 (AT5G05390), AtLAC13 (AT5G07130), AtLAC14 (AT5G09360), AtLAC15 (AT5G48100), AtLAC16 (AT5G58910), AtLAC17 (AT5G60020).

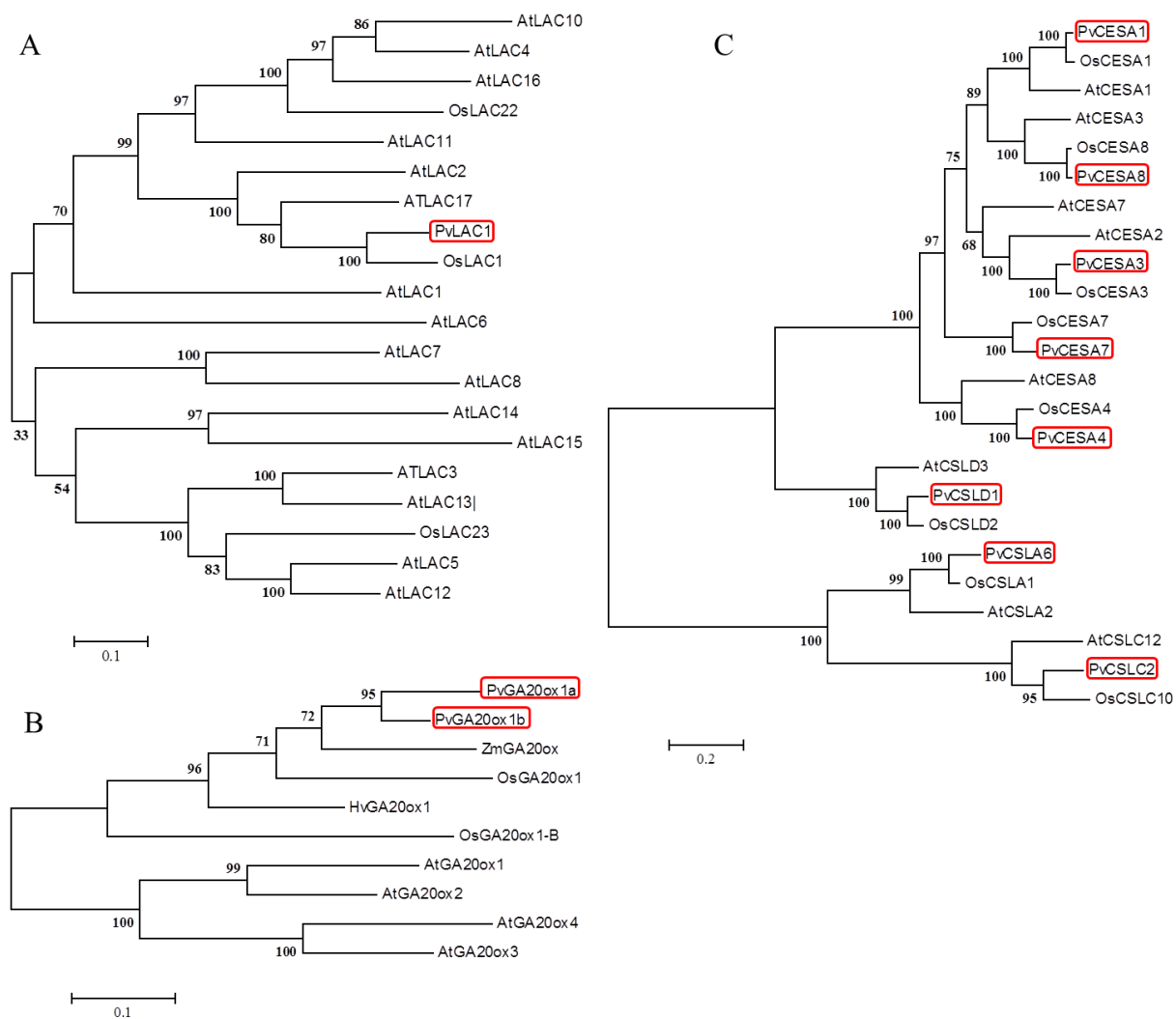


Figure 3-15 Continued.

Table 3-1 List of primers used in this study.

Primer name	Primer sequence
Primers for Cloning <i>PvKN1a</i> cDNA	
KNF1	ATGGAGGAGATCACCCACC
KNR12	GCCGAGCCGGTACAGCC
Primers for Cloning <i>PvKN1b</i> cDNA	
834 OE-F1	ATGGAGGAGATCACCCACCAGTA
834 OE-R1	ATAGCCGAGCCGGTACAGCCCCGC
Primers for genomic DNA PCR	
Hygro_F	TTGCATCTCCCGCCGTTACAG
Hygro_R	CTGGGGCGTCGGTTTCCACTAT
834_F	TGGGATCTGCACTACAAATGGCCT
834_R/common	ACAGCGACTTCCTGACCATCCT
Gene-specific primers for qRT-PCR	
Primers for confirming the overexpression of <i>PvKN1</i>	
F_834III	GGACGGGCACTTCATCAACG
R_AcV5	ACCAGCCGCTCGCATCTTTC
Gene specific primers <i>PvKN1a</i>:	
PvKN1_F2	ATCTGAGGAAATGCACCACCTG
PvKN1_R2	ATGAAGTGCCCGTCCATGTAGA
Gene specific primers for <i>PvKN1b</i>	
AP13CTG11002F2	CCTGAAATTGATGCACATGGTGTTG
AP13CTG11002R2	AGTGCAGATCCCACCAGCTAAG

Table 3-1 Continued.

Primer name	Unitranscript ID	Primer sequence	Reference
Primers for reference gene			
F_PvUBIQUITIN	AP13CTG25905	CAGCGAGGGCTCAATAATTCCA	Xu et al., 2011
R_PvUBIQUITIN		TCTGGCGGACTACAATATCCA	
Primers for analyzing the expression of lignin biosynthetic genes			
C4H1_1534F	AP13CTG28733	GGGCAGTTCAGCAACCAGAT	Shen et al., 2012
C4H1_1611R		CGCGTTTCCGGGACTCTAG	
PvCOMT_F461	KanlCTG02872	CAACCGCGTGTTCAACGA	Shen et al., 2012
PvCOMT_R534		CGGTGTAGAACTCGAGCAGCTT	
4CL1_1179_F	AP13CTG06049	CGAGCAGATCATGAAAGGTTACC	Shen et al., 2012
4CL1_1251_R		CAGCCAGCCGTCCTTGTC	
PvCCR1. 112_F	AP13ISTG52570	GCGTCGTGGCTCGTCAA	Shen et al., 2012
PvCCR1. 187_R		TCGGGTCATCTGGGTTTCCT	
PvCAD_F116	KanlCTG19538	TCACATCAAGCATCCACCATCT	Shen et al., 2012
PvCAD_R184		GTTCTCGTGTCCGAGGTGTGT	
PAL_F1	KanlCTG00004	CATATAGTGTGCGTGCGTGTGT	Wuddineh et al., 2015
PAL_R1		CTGGCCCGCCAATCG	
C3H_F1	AP13ISTG41630	CGTGAACAATGGGATCAGGATAG	Wuddineh et al., 2015
C3H_R1		GCGGACACAACCATCTCAAATAC	
F5H_F1	AP13ISTG56842	CCCCGTGCACTGACGATCTAT	Wuddineh et al., 2015
F5H_R1		CCAAGCCAAGGGAAAACACAGTTA	
Laccase_F2	AP13CTG11594	AGCATCCTGGGCATCGAGAG	Figure 3- 15A
Laccase_R2		GTCGTAGTTGCCGTACCCTTG	
Primers for cellulose and hemicellulose synthetic genes			
CESA1_F1	AP13CTG06092	GCATCCAGGGTCCAGTTTATGTG	Figure 3- 15C
CESA1_R1		CCAGATCGGCTTCGGTCAATAC	
CESA3_F1	AP13CTG00607	GCATTTCCCTCCTCGTCGTC	Figure 3- 15C
CESA3_R1		ACTTGCTCCTCTCGCTCTAGC	

Table 3-1 Continued.

Primer name	Unitranscript ID	Primer sequence	Reference
CESA4_F1	AP13CTG01684	GAATGCTCTGGTCCGAGTGTC	Figure 3-15C
CESA4_R1		TCTGCCAACAGTAGGGTCCATC	
CESA7_F1	AP13CTG19403	CTTCTCGCTCGTCTGGGTTAG	Figure 3-15C
CESA7_R1		AGCTCGATCAATTCAGCACTCG	
CESA8_F1	AP13CTG00048_1	CGTGAGAAGCGTCCTGGATTC	Figure 3-15C
CESA8_R1		CCTGAGAGCCTTGCTGTTGTTG	
CSLA6_F1	AP13CTG14018	TGCACATTACGGAGCTTGGTG	Figure 3-15C
CSLA6_R1		TGGTCCCTCCCATACGCAAG	
CSLC2_F1	AP13CTG06284	GTAGAAGCAGCCAAGGCACTG	Figure 3-15C
CSLC2_R1		GAGTCGCTGTACGCCTCTTG	
CSLD1_F1	AP13CTG08033	CACAGCTACCACGTCCACATC	Figure 3-15C
CSLD1_R1		CGATGACCTTGTCCATGAGGTG	
Primers for gibberellin biosynthetic and catabolic genes			
46534F1	Pavirv00046534	ACACCGACAGCGACTTCCTC	Wuddineh et al., 2015
46534R1		GGTCTCCGACGTTGACGATG	
35270F2	Pavirv00035270m	CAAGAGCGTGGAGCACAAGG	Wuddineh et al., 2015
35270R2		CGAAGGTGAAGGTCCTGTAGG	
KanlCTG23388F1	KanlCTG23388	AGCGTGGAGCACAAGGTGATG	Wuddineh et al., 2015
KanlCTG23388R1		TCTTCCTGCACCTTCCGTCTG	
GA20-oxF1	AP13ISTG69826	CTGCTGCTCTGATCTGCTCCTG	Figure 3-15B
GA20-oxR1		GCAGGCAGGCATACATCGTCT	
GA20-oxF3	AP13ISTG41447	GCTCTCGCTGGAGATCATGGAG	Figure 3-15B
GA20-oxR3		GGAGTCGTTCCCCTCGAAGAAG	

Table 3-2 Percent identity matrix of the switchgrass KNOX transcription factors analyzed in this study.

	PvKN1a	PvKN1b	PvKN2a	PvKN2b	PvKN3a	PvKN3b	PvKN4a	PvKN4b	PvKN5a	PvKN5b	PvKN6a	PvKN6b	PvKN7a	PvKN7b	PvKN8a	PvKN8b	PvKN9a	PvKN9b	PvKN10a
PvKN1b	96.24																		
PvKN2a	57.01	58.10																	
PvKN2b	56.48	57.86	96.44																
PvKN3a	49.85	51.40	72.67	72.29															
PvKN3b	49.85	51.55	70.93	71.20	93.73														
PvKN4a	54.81	54.13	59.59	59.59	58.02	57.38													
PvKN4b	47.71	48.03	52.49	51.97	52.43	51.77	94.64												
PvKN5a	49.38	50.78	51.44	50.95	46.45	47.12	49.19	44.30											
PvKN5b	50.46	51.08	51.27	50.78	47.28	47.94	49.60	44.67	90.45										
PvKN6a	47.12	47.84	49.82	49.12	46.07	46.43	47.64	44.96	45.05	45.92									
PvKN6b	47.35	48.06	48.60	47.92	45.26	45.26	47.64	44.29	44.44	45.30	97.66								
PvKN7a	48.70	48.34	48.92	48.38	48.69	49.06	48.70	45.90	47.89	47.72	83.86	84.21							
PvKN7b	48.70	48.34	49.28	48.38	48.69	49.06	48.70	46.27	47.54	47.37	84.21	84.56	99.66						
PvKN8a	29.96	29.63	30.19	30.08	28.52	28.63	31.98	29.39	26.52	27.50	30.45	30.83	32.06	32.06					
PvKN8b	30.37	30.04	30.22	30.11	28.20	28.30	31.98	29.06	26.60	27.56	30.48	30.86	32.45	32.45	98.68				
PvKN9a	32.84	31.87	31.97	30.77	30.11	29.74	32.03	31.25	34.04	34.15	33.96	33.33	33.08	32.71	60.69	60.82			
PvKN9b	32.35	31.62	31.25	30.29	29.63	29.26	31.60	30.29	33.92	33.68	33.46	32.85	32.71	32.33	60.07	60.20	99.36		
PvKN10a	31.37	30.63	32.96	31.48	31.32	31.32	34.53	33.58	32.03	31.80	36.57	37.04	34.60	34.98	61.43	61.15	89.49	88.96	
PvKN10b	32.46	31.37	33.08	31.84	31.18	31.30	34.22	32.95	33.09	32.86	36.23	36.33	34.87	34.48	61.90	61.62	90.17	89.60	98.67

Amino acid sequences aligned with Muscle program (Edgar, 2004) and percent identity was calculated as the number of identical positions, including gaps divided by the length of the alignment.

Table 3-3 Features of switchgrass KNOX proteins compared to the maize ZmKN1 protein.

Gene	Amino acid identity with ZmKN1 (%)			
	HD domain ^b	GSE ^c	MEINOX ^d	Entire sequence
Class I^a				
PvKN1a	98.4	100	92.8	90.7
PvKN1b	98.4	94.5	89.7	90.5
PvKN2a	88.7	25.8	62.5	54.0
PvKN2b	88.7	25.8	62.5	55.2
PvKN3a	82.3	12.5	56.7	47.1
PvKN3b	82.3	24.0	57.7	47.6
PvKN4a	80.7	13.8	55.7	53.5
PvKN4b	80.7	13.8	58.9	47.1
PvKN5a	82.3	30.3	50.5	47.8
PvKN5b	83.9	36.4	54.6	49.5
PvKN6a	74.2	16.7	51.6	48.8
PvKN6b	74.2	16.7	51.6	48.6
PvKN7a	75.8	8.1	48.4	49.6
PvKN7b	75.8	8.1	48.4	50.0
Class II^a				
PvKN8a	61.3	13.5	25.8	29.8
PvKN8b	61.3	13.5	25.8	29.1
PvKN9a	56.5	10.8	34.0	32.3
PvKN9b	56.5	10.8	33.0	32.2
PvKN10a	56.5	10.8	35.1	32.4
PvKN10b	56.5	10.8	35.1	32.1

- Class I and II are phylogenetic classes of KNOX transcription factor family based on sequence similarity in the HD region, expression pattern and intron positions.
- Homeodomain (HD) is a 60-63 amino acid sequence helix- turn -helix DNA-binding motif that characterizes KNOX genes family.
- GSE is a relatively smaller less conserved domain found between the MEINOX and ELK domains
- MEINOX is a conserved N-terminal domain shared between the MEIS genes in animals and KNOX genes in plants both belonging to a subclass of TALE (three amino acid loop extension) family of genes.

Table 3-4 Morphology and biomass yields of transgenic switchgrass lines overexpressing *PvKN1* and non-transgenic control (WT) plants.

Lines	Tiller height (cm)	Tiller number	Fresh weight (g)	Dry weight (g)
1	50.7±1.6 ^a	7.5±0.5 ^a	24.8±6.5 ^a	7.5±1.2 ^a
3	40.5±1.7 ^b	12.7±4.5 ^a	24.3±11.8 ^a	9.3±4.3 ^a
4	45.0±3.4 ^{ab}	10.3±4.2 ^a	21.9±3.2 ^a	8.4±0.4 ^a
WT	45.3±2.5 ^{ab}	9.7±1.3 ^a	27.7±9.3 ^a	10.7±1.3 ^a

Tiller height estimates were determined for each plant by taking the mean of the 5 tallest tillers within each biological replicate. The fresh and dry biomass measurements were taken on 4 month old aboveground plant biomass harvested at similar growth stages at the 4 month stage. Values are means of 3 biological replicates ± standard errors (n=3). Values represented by different letters are significantly different at $P \leq 0.05$ as determined by Tukey grouping using SAS software (SAS Institute Inc.).

Table 3-5 List of the locus names and/or GenBank accession numbers of the sequences used in this study.

Gene name	Locus name/accession number	Species
PvKN1a	Pavir.Ca01130	<i>Panicum virgatum</i>
PvKN1b	Pavir.J13005	<i>Panicum virgatum</i>
PvKN2a	Pavir.Ea01271	<i>Panicum virgatum</i>
PvKN2b	Pavir.Eb02382	<i>Panicum virgatum</i>
PvKN3a	Pavir.J27068	<i>Panicum virgatum</i>
PvKN3b	Pavir.Eb03210	<i>Panicum virgatum</i>
PvKN4a	Pavir.Ca01636	<i>Panicum virgatum</i>
PvKN4b	Pavir.J11479	<i>Panicum virgatum</i>
PvKN5a	Pavir.Ba04002	<i>Panicum virgatum</i>
PvKN5b	Pavir.Bb00108	<i>Panicum virgatum</i>
PvKN6a	Pavir.Ga01287	<i>Panicum virgatum</i>
PvKN6b	Pavir.Gb01218	<i>Panicum virgatum</i>
PvKN7a	Pavir.Ea00180	<i>Panicum virgatum</i>
PvKN7b	Pavir.J04662	<i>Panicum virgatum</i>
PvKN8a	Pavir.J34867	<i>Panicum virgatum</i>
PvKN8b	Pavir.Ca00918	<i>Panicum virgatum</i>
PvKN9a	KanlCTG23388 (Switchgrass unitranscript ID)	<i>Panicum virgatum</i>
PvKN9b	Pavir.J11300	<i>Panicum virgatum</i>
PvKN10a	Pavir.Da02356	<i>Panicum virgatum</i>
PvKN10b	Pavir.J38727	<i>Panicum virgatum</i>
WKNOX1b	AF224499_1	<i>Triticum aestivum</i>
WRS1	BAH03543	<i>Triticum aestivum</i>
WKNOX1a	AF224498_1	<i>Triticum aestivum</i>
HvKNOX3	BAK07974	<i>Hordeum vulgare</i>
HvKN1	AAQ11882	<i>Hordeum vulgare</i>
SbKN1	ABC71525	<i>Sorghum bicolor</i>
SbKN2	Sb10g025440	<i>Sorghum bicolor</i>

Table 3-5 Continued.

Gene name	Locus name/accession number	Species
SbKN3	XP_002451671	<i>Sorghum bicolor</i>
SbKN4	Sobic.001G526200	<i>Sorghum bicolor</i>
OsHOS3	BAA77817	<i>Oryza sativa Japonica Group</i>
OSH1	ABF98653/ LOC_Os03g51690	<i>Oryza sativa Japonica Group</i>
OSH45	KNOSD_ORYSJ/LOC_Os08g19650	<i>Oryza sativa Japonica Group</i>
HOS66	KNOS3_ORYSJ/LOC_Os03g03164	<i>Oryza sativa Japonica Group</i>
HOS58	KNOS2_ORYSJ/LOC_Os02g08544	<i>Oryza sativa Japonica Group</i>
HOS59	KNOSB_ORYSJ/LOC_Os06g43860	<i>Oryza sativa Japonica Group</i>
ZmRS1	NP_001149651	<i>Zea mays</i>
ZmKN3	XP_008659775	<i>Zea mays</i>
ZmKN1	NP_001105436	<i>Zea mays</i>
ZmKN2	XP_008660553	<i>Zea mays</i>
ZmKN4	AFW66300	<i>Zea mays</i>
ZmKN6	NP_001150419	<i>Zea mays</i>
ZmKN5	AFW70462	<i>Zea mays</i>
SiKN1	ABC71528	<i>Setaria italica</i>
SiKN5	XP_004985913	<i>Setaria italica</i>
SiKN3	XP_004951667	<i>Setaria italica</i>
SiKN4	XP_004973169	<i>Setaria italica</i>
SiKN2	XP_004965737	<i>Setaria italica</i>
BdKN1	XP_003558936	<i>Brachypodium distachyon</i>
BdKN2	XP_003563314	<i>Brachypodium distachyon</i>
BdKN4	XP_003573667	<i>Brachypodium distachyon</i>
BdKN3	XP_003570788	<i>Brachypodium distachyon</i>
AtqKNOX1	ADN43388	<i>Agave tequilana</i>
PmKN1	ABC71526	<i>Panicum miliaceum</i>
KNAT1	AT4G08150/NP_192555	<i>Arabidopsis thaliana</i>

Table 3-5 Continued.

Gene name	Locus name/accession number	Species
KNAT2	NP_177208/ AT1G70510	<i>Arabidopsis thaliana</i>
KNAT3	AT5G25220/ NP_001031938	<i>Arabidopsis thaliana</i>
KNAT4	AT5G11060/ NP_196667	<i>Arabidopsis thaliana</i>
KNAT5	NP_194932/ AT4G32040	<i>Arabidopsis thaliana</i>
KNAT6	NP_850951/ AT1G23380	<i>Arabidopsis thaliana</i>
KNAT7	NP_564805/ AT1G62990	<i>Arabidopsis thaliana</i>
STM	NP_176426/AT1G62360	<i>Arabidopsis thaliana</i>
PpKNOPE1	ABD52723	<i>Prunus persica</i>
PpKNOPE2	ABO28750	<i>Prunus persica</i>
PpKNOPE2.1	JQ038131	<i>Prunus persica</i>
PpKNOPE3	ACJ71731	<i>Prunus persica</i>
PpKNOPE4	ABO26062	<i>Prunus persica</i>
PpKNOPE6	ADC35598.1	<i>Prunus persica</i>
PpKNOPE7	JQ038132	<i>Prunus persica</i>
ChBP	ABG66654	<i>Cardamine hirsuta</i>
BrBP	ACS28249	<i>Brassica rapa</i>
BoBP	ACS28250	<i>Brassica oleracea</i>
BnBP	ACR83812	<i>Brassica napus</i>
MtKNOX	ABO33479	<i>Medicago truncatula</i>
KNAP2	KNAP2_MALDO	<i>Malus domestica</i>
IbKN3	BAF93480	<i>Ipomoea batatas</i>
IbKN2	BAF93479	<i>Ipomoea batatas</i>
SlKN1	NP_001233807	<i>Solanum lycopersicum</i>
NtKN2	AAQ11889	<i>Nicotiana tabacum</i>
NtKN3	AAQ11890	<i>Nicotiana tabacum</i>
HaKN2	AAM28232	<i>Helianthus annuus</i>
LjKN2	AAX21346	<i>Lotus japonicus</i>
ARK1	AAV28488	<i>Populus tremula</i> x <i>Populus alba</i>
PtKN1	XP_002301134	<i>Populus trichocarpa</i>

Chapter 4: Genome-wide analysis of switchgrass AP2/ERF transcription factor superfamily, and overexpression of *PvERF001* for improvement of biomass characteristics for biofuel.

(A version of this chapter was published in *Frontiers in Bioengineering and Biotechnology* (3:101. doi: 10.3389/fbioe.2015.00101) with authors, Wegi A. Wuddineh, Mitra Mazarei, Geoffrey B. Turner, Robert W. Sykes, Stephen R. Decker, Mark F. Davis, C. Neal Stewart, Jr.)

Wegi A Wuddineh designed and conducted the experiments except lignin and sugar release assays and wrote the manuscript.

4.1 Abstract

Switchgrass (*Panicum virgatum*) is a leading candidate lignocellulosic bioenergy crop due to its high biomass yield, versatility as well as high potential for marginal lands. Cost-efficient biofuel production from lignocellulosic biomass requires not only reducing biomass recalcitrance to make the cell wall carbohydrates more accessible to enzymatic breakdown into fermentable sugars, but also enhancements in yield potential under intense stress factors prevalent in such environments. The APETALA2/ ethylene response factor (AP2/ERF) superfamily of TFs play essential roles in the regulation of various growth and developmental programs including stress responses. Some of these TFs in other plant species have also been implicated to play a role in the regulation of cell wall biosynthesis. Here, we identified a total of 207 AP2/ERF TFs in the available switchgrass genome, which were grouped into 4 families and subfamilies comprising 25 AP2, 121 ERF, 55 DREB (dehydration responsive element binding)-, and 5 RAV (related to API3/VP) genes, as well as a singleton gene not fitting any of the above families. Detailed analysis of these gene families is presented including the gene structure, distribution of conserved motifs and clustering of putative homologs. *In silico* functional analysis revealed that many genes in these families might be associated in the regulation of responses to environmental stimuli via transcriptional regulation of the response genes. Moreover, these genes had diverse endogenous expression patterns in switchgrass during seed germination, vegetative growth, flower development and seed formation. Interestingly, several members of ERF and DREB subfamilies were found to be highly expressed in plant tissues where active lignification and secondary cell wall formation takes place. These results provide vital resources for mining genes that impart tolerance to environmental stress as well as reduced recalcitrance to help in the genetic improvement of switchgrass. Overexpression of one of the ERF subfamily gene

(*PvERF001*) in switchgrass was associated with increased biomass yield and sugar release efficiency in transgenic lines, exemplifying the potential of these TFs in the development of lignocellulosic feedstocks with improved biomass characteristics for biofuel.

Keywords: AP2, ethylene response factors, stress response, transcription factors, biofuel, *PvERF001*, overexpression, sugar release

4.2 Introduction

Switchgrass (*Panicum virgatum*) is an outcrossing perennial C4 grass known for its vigorous growth and wide adaptability, and hence, is being developed as a candidate lignocellulosic biofuel feedstock (Yuan et al., 2008). The feasibility of commercial production of liquid transportation biofuel from switchgrass biomass is hampered by biomass recalcitrance (the resistance of cell wall to enzymatic breakdown into simple sugars). Lignin is considered to be a primary contributor to biomass recalcitrance as it hinders the accessibility of cell wall carbohydrates to hydrolytic enzymes. Substantial progress has been made in engineering the switchgrass lignin biosynthesis pathway to reduce lignin content and/or modify its composition (Fu et al., 2011a; Fu et al., 2011b; Shen et al., 2012; Shen et al., 2013a; Shen et al., 2013b; Baxter et al., 2014; Baxter et al., 2015). The downregulation of individual genes in the lignin biosynthesis pathway has been effective to reduce lignin, but can result in the production of metabolites that can impede downstream fermentation processes (Tschaplinski et al., 2012). Alternatively, overexpression of transcription factors (TFs) such as switchgrass *MYB4* has been shown to circumvent this inhibitory effect while leading to significantly reduced biomass recalcitrance and improved ethanol production (Shen et al., 2012; Shen et al., 2013b; Baxter et al., 2015).

The master regulators of gene clusters with altered expression, could, in turn, endow such traits as increased biomass yield, tiller number, improved germination/plant establishment or root growth as well as tolerance to environmental stresses (Xu et al., 2011; Licausi et al., 2013; Ambavaram et al., 2014). Therefore, identification of TFs with such putative roles would provide a dynamic approach to developing better biofuel feedstocks that could thrive under adverse environmental conditions. The availability of switchgrass ESTs (Zhang et al., 2013) and draft genome sequences produced by Joint Genome Institute (JGI), Department of Energy, USA, provides a vital resource for the discovery of relevant target genes that could be utilized in the genetic improvement of perennial grasses, which could be used as dedicated bioenergy feedstocks. However, compared to dicots such as *Arabidopsis*, relatively little is known about the key regulatory mechanisms in monocots that control lignification and cell wall formation; this is especially true of switchgrass. Likewise, we also have depauperate knowledge about stress responses and defense against pests in these species.

APETALA2/ethylene responsive factor (AP2/ERF) is a large group of regulatory protein families in plants that are characterized by the presence of one or two conserved AP2 DNA binding domains. AP2/ERF TFs are involved in the transcriptional regulation of various growth and developmental processes and responses to environmental stressors. The AP2 domain is a stretch of 60-70 conserved amino acid sequences that is essential for the activity of AP2/ERF TFs (Jofuku et al., 2005). It has been demonstrated that the AP2 domain binds the cis-acting elements including the GCC box motif (Ohme-Takagi and Shinshi, 1995), the dehydration responsive element (DRE)/C-repeat element (CRT) (Sun et al., 2008) and/or TTG motif (Wang et al., 2015) present in the promoter regions of target genes thereby regulating their expression. The AP2/ERF superfamily can be divided into three major families namely, ERF, AP2, and RAV (related to API3/VP) (Licausi et al., 2013). The ERF family is further subdivided into two subfamilies, ERF and dehydration responsive element binding proteins (DREB) based on similarities in amino acid residues in the AP2 domain. The DREB subfamily in *Arabidopsis* and rice has been further classified into 4 distinct groups while ERF subfamily was clustered into 8 groups in *Arabidopsis* and 11 groups in rice based on analysis of gene structure and conserved motifs (Nakano et al., 2006). The AP2 family comprises two groups of proteins differing in the number of AP2 domain in their amino acid sequences. The majority of proteins in this group are characterized by the presence of two AP2 domains, but a few members of this group has only a single AP2 domain that is more similar to the AP2 domains in the double domain groups. RAV proteins, on the other hand, are a small family TFs characterized by the presence of B3 DNA binding domain besides a single AP2 domain. Genome-wide analysis of AP2/ERF TFs has been extensively studied in many dicots including *Arabidopsis* (Nakano et al., 2006), *Populus* (Zhuang et al., 2008; Vahala et al., 2013), Chinese cabbage (Liu et al., 2013), grapevine (Licausi et al., 2010), peach (Zhang et al., 2012) and castor bean (Xu et al., 2013). However, with the exception of rice (Nakano et al., 2006; Rashid et al., 2012), and foxtail millet (Lata et al., 2014), little information is available on the AP2/ERF TF families in monocots.

Numerous genes coding for AP2/ERF superfamily TFs have been identified and functionally characterized in various plant species (Xu et al., 2011; Licausi et al., 2013). The DREB subfamily proteins have been extensively studied with regard to tolerance to abiotic stress such as freezing (Jaglo-Ottosen et al., 1998; Ito et al., 2006; Fang et al., 2015), drought (Hong and

Kim, 2005; Oh et al., 2009; Fang et al., 2015), heat (Qin et al., 2007) and salinity (Hong and Kim, 2005; Bouaziz et al., 2013). Moreover, it has been reported that DREB genes play roles in the regulation of ABA-mediated gene expression in response to osmotic stress during germination and early vegetative growth stage (Fujita et al., 2011). ERF TFs, on the other hand, have been shown to participate in the regulation of defense responses against various biotic stress (Guo et al., 2004; Dong et al., 2015) and/or tolerance to environmental stressors such as drought (Aharoni et al., 2004; Zhang et al., 2010b), osmotic stress (Zhang et al., 2010a), salinity (Guo et al., 2004), hypoxia (Hattori et al., 2009) and freezing (Zhang and Huang, 2010). Moreover, AP2/ERF TFs in aspen (*PtaERF1*) and *Arabidopsis* (*AtERF004* and *AtERF038*) have been suggested to be associated with the regulation of cell wall biosynthesis in some tissues (Van Raemdonck et al., 2005; Lasserre et al., 2008; Ambavaram et al., 2011). The functions of AP2 family TFs, on the other hand, have been associated with plant organ-specific regulation of growth and developmental programs (Elliott et al., 1996; Jofuku et al., 2005; Horstman et al., 2014). Genes in the RAV TF family have been shown to play a role in the regulation of gene expression in response to phytohormones such as ethylene and brassinosteroid as well as in response to biotic and abiotic stress (Mittal et al., 2014). Therefore, AP2/ERF TF superfamily may hold tremendous potential for the improvement of bioenergy feedstocks, such as switchgrass, that is intended to be grown on marginal lands that could impose undue environmental stress.

In this study, we report the identification of 207 AP2/ERF TF genes in the switchgrass genome. Cluster analysis of the identified proteins, distribution of conserved motifs, analysis of their gene structure and expression profiling are presented. We highlight the potential application of these data to identify putative target genes that might be exploited to improve bioenergy feedstocks. To that end, we cloned one of the ERF subfamily genes, which was subsequently overexpressed in switchgrass to improve biomass productivity and sugar release efficiency.

4.3 Materials and methods

4.3.1 Identification of AP2/ERF gene families in switchgrass genome

We used representative genes from appropriate rice gene families as the basis to search for orthologues in switchgrass. The amino acid sequences of AP2 domain-containing rice genes

representing three families: AP2 (Os02g40070), ERF (Os06g40150) and RAV (Os01g04800) were used to query the amino acid sequences of all switchgrass AP2/ERF proteins using TBLASTN against the switchgrass EST database (Zhang et al., 2013) or BLASTP against the *Panicum virgatum* draft genome (Phytozome v1.1 DOE-JGI) (<http://www.phytozome.net/>). The sequences were retrieved and evaluated for the presence of AP2 domains by searching against the conserved domain database (CDD) at NCBI. The AP2-containing switchgrass sequences were further evaluated for any redundant and missing sequences by BLASTP searches using the previously identified homologous counterparts of the foxtail millet (Lata et al., 2014) and rice (Nakano et al., 2006; Rashid et al., 2012). The presence of multiple gene copies from the tetraploid switchgrass genome was addressed by the identification of only a single gene copy with the highest similarity to the corresponding homologs in foxtail millet or rice. Genes with additional domains besides the AP2 domain with no corresponding homologs in foxtail millet, rice and *Arabidopsis* AP2/ERF TFs were excluded from our subsequent analysis.

4.3.2 Cluster and protein sequence analysis of AP2/ERF TFs

The amino acid sequences of the AP2/ERF TFs were imported into the MEGA6 program and multiple sequence alignment analysis was conducted using MUSCLE with default parameters (Edgar, 2004). Construction of cluster trees was performed using the neighbor-joining (NJ) method by the MEGA6 program using a bootstrap value of 1000, Poisson correction and pairwise deletion (Tamura et al., 2013). Conserved motifs in switchgrass AP2/ERF TFs were identified with the online tool, MEME version 4.10.0 (<http://meme.sdsc.edu/meme/meme.html>) using the following parameters: optimum width, 6-200 amino acids; with any number of repetitions and maximum number of motifs set at 25 (Bailey and Elkan, 1994).

4.3.3 Analysis of gene structure and gene ontology (GO) annotation

The genomic and coding DNA sequences of the identified AP2/ERF TFs were retrieved from the Phytozome (*Panicum virgatum* v1.1 DOE-JGI). The exon-intron organizations in these genes were visualized by the gene structure display server (<http://gsds.cbi.pku.edu.cn/>) (Guo et al., 2007). To evaluate the gene ontology (GO) annotation of the identified AP2/ERF TFs, their amino acid sequences were imported into the Blast2GO suite (Conesa and Gotz, 2008). BLASTP search was performed against rice protein sequences at NCBI. The resulting hits were mapped to

obtain the GO terms, which were annotated to assign functional terms to the query sequences. Plant GOslim was used to filter the annotation to plant-related terms. The protein subcellular localization prediction tool WOLF PSORT (<http://www.genscript.com>) was used to complement the results of the cellular localization predicted by blast2GO.

4.3.4 Analysis of transcript data from the switchgrass gene expression atlas

The transcript data for the AP2/ERF superfamily TFs was extracted from the publicly available switchgrass gene expression atlas (PviGEA) (<http://switchgrassgenomics.noble.org/>) (Zhang et al., 2013), which was obtained by Affymetrix microarray analysis. The probe set IDs of 108 matching genes representing the switchgrass unitranscripts (PviUT) were identified by TBLASTN query search using the amino acid sequences of the AP2/ERF TFs. The transcript data for each tissues and stage of development were retrieved using the probe set IDs. The expression values of the genes were log2 transformed and a heatmap was created using an online graphing tool, Plotly (<https://plot.ly/plot>). Tissues used for the extraction of RNA to determine the level of expression included the following: whole seeds for seed germination at 24 h, 48 h, 72 h and 96 h intervals post-imbibition, whole shoots and roots at vegetative stages, V1 to V5, pooled leaf sheath, leaf blade and nodes, whole crown, the bottom, middle and top portions of the 4th internode, vascular bundle tissues and middle portion of the 3rd internode all at E4 (stem elongation stage 4) developmental stage. For analysis of the expression level during reproductive developmental stages, inflorescence tissues and whole seeds along with floral tissues such as lemma and palea was used.

4.3.5 Vector construction and plant transformation

Cloning and tissue culture was performed as previously described (Wuddineh et al., 2015). Briefly, the putative homolog of *Arabidopsis* AtSHN2 (At5g11190) and rice OsSHN (Os06g40150) was identified by TBLASTN or BLASTP against the switchgrass EST database or draft genome (Phytozome v1.1 DOE-JGI) followed by cluster and multiple sequence alignment analysis to discriminate the most closely related gene for cloning. For construction of overexpression cassette, the open reading frame (ORF) of *PvERF001* was isolated from cDNA obtained from ST1 clonal genotype of ‘Alamo’ switchgrass using gene specific primers flanking the ORF of the gene excluding the stop codon and cloned into pANIC-10A expression vector by

GATEWAY recombination (Mann et al., 2012). The primer pairs used for cloning are shown in Table 4-1. Embryogenic callus derived from SA1 clonal genotype of ‘Alamo’ switchgrass (King et al., 2014) was transformed with the expression vector construct through *Agrobacterium*-mediated transformation (Burris et al., 2009). Antibiotic selection was carried out for about two months on 30- 50 mg/L hygromycin followed by regeneration of orange fluorescent protein reporter-positive callus sections on regeneration medium (Li and Qu, 2011) containing 400 mg/L timentin. Regenerated plants were rooted on MSO medium (Murashige and Skoog, 1962) with 250 mg/L cefotaxime to assure elimination of *Agrobacterium* from the tissues as well as promote shoot regeneration from transgenic callus (Grewal et al., 2006), and the transgenic lines were screened based on the presence of the insert and expression of the transgene. Simultaneously a non-transgenic control line was also generated from callus.

4.3.6 Plants and growth conditions

T0 transgenic and non-transgenic control plants were grown in growth chambers under standard conditions (16 h day/8 h night light at 24°C, 390 $\mu\text{E m}^{-2} \text{s}^{-1}$) and watered three times per week, including weekly nutrient supplements with 100 mg/L Peter’s 20-20-20 fertilizer. Transgenic and non-transgenic control lines were propagated from a single tiller to produce three clonal replicates for measuring growth parameters (Hardin et al., 2013). The plants were grown in 12-L pots in Fafard 3B soil mix (Conrad Fafard, Inc., Agawam, MA) and grown for 4 months to the R1 stage, in which shoot samples were collected to assay the transgene transcript abundance (Moore et al., 1991; Shen et al., 2009). Each sample was snap frozen in liquid nitrogen and macerated with mortar and pestle. The macerated samples were used for RNA extraction as described below.

4.3.7 RNA extraction and quantitative reverse transcription polymerase chain reaction (qRT-PCR)

RNA extraction and analysis of transgene transcripts were performed as previously described (Wuddineh et al., 2015). Briefly, total RNA was extracted from shoot tip samples of transgenic and non-transgenic control lines using Tri-Reagent (Molecular Research Center, Cincinnati, OH), and 3 μg of the RNA was treated with DNase-I (Promega, Madison, WI). High-Capacity cDNA Reverse Transcription kit (Applied Biosystems, Foster City, CA) was used for the

synthesis of first-strand cDNA. Power SYBR Green PCR master mix (Applied Biosystems) was utilized to conduct qRT-PCR analysis according to the manufacturer's protocol. All the experiments were conducted in triplicates. The list of all primer pairs used for qRT-PCR is shown in Table 4-1. Analysis of the relative expression was done as previously described (Wuddineh et al., 2015). There was no amplification products observed with all the primer pairs when using only the RNA samples or the water instead of cDNA.

4.3.8 Determination of leaf water loss

The rate of water loss via leaf epidermal layer was determined as previously described (Zhou et al., 2014). The second fully expanded leaves of both transgenic and non-transgenic plants were excised and soaked in 50 ml distilled water for 2 h in the dark to saturate the leaves. Subsequently, the excess water was removed and initial leaf weight was measured and water loss determined by weighing the leaves every 30 min for at least 3 h. Subsequently, the detached leaves were dried for 24 h at 80°C to determine the final dry weight. The rate of water loss was calculated as the weight of water lost divided by the initial leaf weight.

4.3.9 Analysis of lignin content and composition

Both qualitative (phloroglucinol-HCl staining) and quantitative (pyrolysis molecular beam mass spectrometry [py-MBMS]) analysis of lignin content was performed as previously described (Wuddineh et al., 2015). Briefly, leaf samples collected at the R1 developmental stage and cleared in a 2:1 solution of ethanol and glacial acetic acid for 5 days were used for staining analysis. The cleared leaf samples were immersed in 1% phloroglucinol (in 2:1 ethanol/HCl) overnight for staining and the pictures were taken at 2x magnification. For the quantification of lignin content and S:G lignin monomer ratio by NREL high-throughput py-MBMS method, tillers were collected at R1 developmental stage, air dried for 3 weeks at room temperature and milled to 1 mm (20 mesh) particle size. Lignin content and composition were determined on extractives- and starch-free samples (Sykes et al., 2009).

4.3.10 Determination of sugar release

For analysis of sugar release efficiency, tiller samples at R1 developmental stage were collected and air-dried for 3 weeks at room temperature. The dry samples were pulverized to 1 mm (20

mesh) particle size and sugar release efficiency was determined via NREL high-throughput sugar release assays on extractives- and starch-free samples (Decker et al. 2012). Glucose and xylose release were measured by colorimetric assays, and total sugar release is the sum of glucose and xylose released.

4.3.11 Statistical analysis

To analyze the differences between treatment means, ANOVA with least significant difference (LSD) procedure was used while PROC TTEST procedure was used to examine the statistical difference between the expression of target genes in transgenic vs non-transgenic lines using SAS version 9.3 (SAS Institute Inc., Cary, NC). Pearson correlation coefficient to determine the relationship between relative transcript levels and growth parameters was calculated by SAS.

4.4 Results

4.4.1 Identification of AP2/ERF TFs in switchgrass genome

A total of 207 unique switchgrass genes containing one or two AP2 DNA binding domain were identified from the currently available switchgrass EST and genome databases. Amino acid sequence similarities within the conserved AP2 domain between these proteins and previously characterized AP2/ERF TFs from rice and *Arabidopsis* along with the presence of the conserved B3 domain suggest that these proteins might be categorized as putative AP2/ERF TFs. The characteristic features of these genes are summarized in Table 4-2. The amino acid sequences of AP2/ERF TFs showed wide variation in size (ranging from 119 to 666 amino acids) and sequence composition. Twenty-two of these TFs contained two AP2 DNA-binding domains and hence were classified under AP2 family. Five of the AP2/ERF proteins had a B3 conserved domain at the C-terminus in addition to the common AP2 domain, and these genes were grouped into the RAV family. Three of the remaining 180 proteins, namely PvERF049, PvERF160 and PvERF177 with a single AP2 domain, which is more similar to the amino acid sequences of AP2 domains in the AP2 family TFs, were also grouped under the AP2 family. Moreover, one AP2/ERF protein showed a distinct AP2 domain different from all other switchgrass AP2/ERF proteins but with higher shared sequence similarity with the previously identified genes in rice and *Arabidopsis*. The remaining 176 proteins were grouped into ERF family, which was further

subdivided into either one of two subfamilies (ERF and DREB) based on sequence similarity in the AP2 domain. The ERF subfamily members included 121 proteins while DREB had only 55 proteins (Table 4-3).

The distribution of the identified switchgrass AP2/ERF genes across the 9 chromosomes was also evaluated. Thus far, only about half of the switchgrass genomic sequences are mapped into their chromosomal locations based on the draft genome assembly by JGI-DOE available at Phytozome. Accordingly, 166 of the 207 genes could be assigned a chromosomal location. The genes were non-evenly distributed across the nine switchgrass chromosomes wherein the highest number of genes was localized on chromosomes 9, 2 and 1, with the fewest number of genes being assigned to chromosome 8 (Table 4-4).

4.4.2 Cluster analysis of switchgrass AP2/ERF TFs

To confirm the classification and evaluate the sequence similarities between the switchgrass AP2/ERF TFs, a dendrogram was constructed by NJ method using the whole amino acid sequences of the proteins based on MUSCLE alignment. The analysis showed distinct clustering of the proteins into specific groups and families as previously described in other species (Figure 4-1). Specifically, these clusters highlighted the distinction between the switchgrass AP2, ERF and RAV families as well as between the ERF and DREB subfamilies. The ERF and DREB subfamilies were further subdivided into 7 (group V to XI) and 4 (I to IV) distinct groups, respectively. The cluster analysis also resolved the RAV protein family and the singleton into separate clusters, which was in accordance with the sequence similarities in the conserved domains as well as the presence of additional domains in the families/clusters.

4.4.3 Characterization of AP2/ERF gene structures and conserved motifs

To complement the cluster analysis-based classification, the exon-intron structures of AP2/ERF genes were evaluated. The schematic representations of protein and gene structures of switchgrass AP2/ERF superfamily are presented in Figure 4-2 (ERF), Figure 4-3 (DREB) and Figure 4-4 (AP2, RAV and Singleton). The ORF lengths of these genes vary from 394 bp for the shortest gene to 5409 bp for the longest gene. Analysis of their gene structure showed highly diverse distribution of intron regions within the ORF of the different gene groups or families.

The majority of genes belonging to ERF and DREB subfamilies and all but one of the RAV genes appeared to be intronless. Only 9 DREB genes (16%) belonging to group I, III and VI had a single intron in their gene structures. Among ERF genes, 45 (37%) had a single intron in their ORF while 8 genes had 2 and 3 of them with 3 introns in its ORF. On the other hand, genes in the AP2 family contained a higher number of introns; ranging from 1 to 10. Only one gene in the AP2 family had a single intron while majority of the genes had more than 5 introns. The position and state of the introns in the ORF of ERF family genes belonging to groups V, VII and X shows relatively high conservation as compared to other groups. For instance, about half of the genes belonging to phylogenetic group V in the ERF family, showed highly conserved intron positions with an intron phase of 2, meaning the location of the intron is found between the 2nd and 3rd nucleotides in the codon. Similarly, the intron positions and splicing phases seems conserved in group VII of the ERF subfamily (Figures 4-2, 4-3 & 4-4).

Analysis of amino acid sequence conservation in the whole proteins of AP2/ERF superfamily showed the presence of unique conserved motifs shared between proteins within families, subfamilies or groups (Figures 4-2, 4-3 & 4-4). Moreover, shared conserved motifs across families, subfamilies or between groups within subfamilies were also detected, signifying the conservation of the proteins in the AP2/ERF superfamily. In general, a total of 25 conserved motifs (M1-M25) were identified in the superfamily of which 14 motifs, M1-M7, M9, M11, M12, M16, M20, M22 and M23 were related to the AP2 domain (Table 4-5). The conserved motifs from the non-AP2 domain region appear to specify individual groups within the subfamilies. Among the ERF subfamily, proteins in groups VII and IX has the most diverse set of motifs compared to others while proteins in group XI harbors merely two motifs, M1 and M23 with the last motif being unique to the group (Figure 4-2). Moreover, Group VII proteins in the ERF subfamily shared a unique motif, M25. Similarly, ERF proteins in the IX group were distinguished by the presence of specific motifs namely M10 and M15 while those in groups VI and VI-L shared a specific motif M18. Most of the DREB genes belonging to group II have only one specific motif (M12) while a few others have additional motifs such as M5 (Figure 4-3). The pattern of conserved motif distribution within the largest group in the DREB subfamily (Group III) showed the presence of two unique subgroups sharing conserved motifs, (M2, M9 and M16) and (M4, M11 and M21), respectively. Three of these motifs (11, 16 and 21) were specific to

proteins in group III DREB subfamily. Other proteins in DREB subfamily belonging to group I were also distinguished by conserved motifs-M13 and M24. The most common motif in proteins of Group IV DREB genes was motif M2 (Figure 4-3). Proteins of AP2 family genes harbor 4 family-specific motifs namely M7, M8, M20 and M22 (Figure 4-4). In addition, the majority of AP2 family proteins share M3 with ERF proteins in group IX. Similarly, RAV proteins also possess two unique motifs, M14 and M17 spanning the B3 DNA binding domain, in addition to M6 and M12 spanning the AP2 domain (Figure 4-4). M6 and M12 are also present in most proteins in the ERF and DREB (group II) subfamilies (Figures 4-2 & 4-3, Table 4-5).

4.4.4 Gene ontology annotation

GO analysis of switchgrass AP2/ERF TFs, based on rice reference sequences, predicted candidate genes' molecular functions, putative roles in the regulation of diverse biological processes, and their cellular localization (Figure 4-5). According to blast2GO outputs, over 95% of the switchgrass genes in the AP2/ERF superfamily were predicted to have sequence-specific DNA binding activities (Figure 4-5A). Furthermore, it also predicted that these genes might be involved in the regulation of various biosynthetic processes, which could include the biosynthesis of cuticle, waxes, hormones and other organic compounds. Importantly, many of these genes were also predicted to participate in the regulation of responses to various environmental stresses caused either by biotic factors such as pathogens and insect pests or abiotic factors such as flooding, water deprivation, wounding and osmotic stress (Figure 4-5B). Cellular localization of the AP2/ERF TFs was predicted by Blast2GO analysis complemented with subcellular localization prediction tool, WoLF PSORT for proteins with heretofore ambiguous results. The results showed that majority of switchgrass AP2/ERF proteins (>80%) were predicted to be at least dual targeted i.e. localized to nucleus, plastid and/or mitochondrion (Figure 4-5C). Only 39 gene products (20%) were predicted to be localized solely to the nucleus (Figure 4-5C).

4.4.5 Expression pattern of switchgrass *AP2/ERF* genes

A switchgrass gene expression atlas (PviGEA) containing expression data for about 78,000 unique transcripts in various tissues was recently developed (Zhang et al. 2013) and is publicly available at web server (<http://switchgrassgenomics.noble.org/>). To investigate whether the

identified switchgrass AP2/ERF genes may have any association with various biological processes that occur during seed germination, vegetative and reproductive development as well as lignification or cell wall development, transcript data was pooled from the PviGEA web server to assess their expression profile.

During seed germination (Figure 4-6), genes in the DREB subfamily (*PvERF101*, *PvERF102*, *PvERF148* and *PvERF158*) showed high expression at early stages of germination (radicle emergence) (48 h after imbibition) while other DREB genes (*PvERF106*, *PvERF136*, *PvERF137*, *PvERF139*, *PvERF142*, and *PvERF143*) showed increased expression at later stages of germination (mainly coleoptile emergence) (Figure 4-6). Similarly, the expression of many ERF genes (*PvERF003*, *PvERF047*, *PvERF051*, *PvERF054*, *PvERF057*, *PvERF068*, *PvERF074*, *PvERF075*, *PvERF087*, *PvERF088*, *PvERF119* and *PvERF178*) showed dramatic increase during early germination stage while others (*PvERF001*, *PvERF002*, *PvERF013*, *PvERF016*, *PvERF018*, *PvERF019*, *PvERF020*, *PvERF037*, *PvERF038*, *PvERF103*, *PvERF109*, *PvERF111*, *PvERF115* and *PvERF164*) had peak expression at later stages (coleoptile emergence (72 h) and mesocotyl elongation (96 h) stages. Genes in the AP2 family (*PvERF193*, *PvERF194*, *PvERF195* and *PvERF201*) displayed increased expression level at radicle emergence whereas few AP2 genes (*PvERF049* and *PvERF203*) showed increased expression at coleoptile emergence. The expression of the RAV genes and the singleton gene were apparently relatively less variable throughout the seed germination process (Figure 4-6).

Comparison of the expression pattern of AP2/ERF genes in roots and shoots at three vegetative phases of development (1st, 3rd or 5th fully-collared leaf stages) revealed apparent differential expression pattern between the organs and different stages of vegetative development (Figure 4-7). Moreover, the expression pattern of AP2/ERF genes during reproductive development also showed differential expression between the reproductive tissues from the initiation of inflorescence meristem to the maturation of the seeds (Figure 4-8).

4.4.6 Expression profiles of switchgrass AP2/ERF genes in lignified tissues

To evaluate whether the identified switchgrass genes coding for AP2/ERF TFs are associated with the regulation of the cell wall biosynthetic genes during cell wall formation or lignification, the transcripts of the genes extracted from the PviGEA web server were used to compare the

level of expression in the lignified tissues of vascular bundles and internode fragments against the expression level in less lignified plant tissues such as leaf blades and sheath. DREB subfamily genes in group I (*PvERF95*, *PvERF98*, *PvERF101* and *PvERF102*) and II (*PvERF148*) had their highest expression in vascular bundles and internode tissues followed by internode portions where active lignification is expected (Figure 4-9). Similarly, the majority of DREB genes belonging to group III including *PvERF133*, *PvERF135*, *PvERF136*, *PvERF137*, *PvERF139*, *PvERF140*, *PvERF142*, *PvERF143*, *PvERF145* and *PvERF146* were highly expressed mainly in the vascular bundles. Many genes in the ERF subfamily belonging to group VIII (*PvERF013*, *PvERF015*, *PvERF016*, *PvERF018*, *PvERF019* and *PvERF020*) and X (*PvERF047*, *PvERF065* and *PvERF103*) also showed the highest expression in the vascular bundles followed by youngest internode sections (Figure 4-9). In comparison, only two genes in group IX (*PvERF037*, *PvERF038*), one gene in group VI-L (*PvERF088*) and 3 genes in group VII (*PvERF111*, *PvERF112* and *PvERF116*) had high expression in vascular bundles. Contrastingly, genes in the ERF subfamily, *PvERF001* and *PvERF002* belonging to group V and *PvERF068* from group VI showed the highest expression in the basal fragments of the 4th internodes (E4) that is under less active lignification. Some genes including *PvERF178* (VI), *PvERF110* (VII), *PvERF115* (VII) and *PvERF164* (VII) and *PvERF038* (IX) had notably high relative expression in roots than in other tissues. Compared to the ERF family genes, the expression of AP2 genes was highly diverse with *PvERF189*, *PvERF194*, *PvERF204* and *PvERF205* having high specificity to roots followed by vascular bundles. The AP2 family gene, *PvERF190*, was expressed highly in the vascular bundles while *PvERF049*, *PvERF205* and *PvERF207* expression was relatively higher in basal and middle sections of the 4th internode. The expression of the two RAV genes analyzed was uniformly low throughout whereas the singleton gene was highly expressed in the leaf blades, leaf sheath as well as the vascular bundles and young internode sections (Figure 4-9).

4.4.7 Overexpression of *PvERF001* in switchgrass have enhanced plant growth and sugar release efficiency

Transgenic switchgrass is desired for less recalcitrance biomass for biofuels. To that end, we selected *PvERF001*, a putative switchgrass homolog of *Arabidopsis* *AtERF004* (*AtSHN2*) and rice *OsERF057* (*OsSHN*) in ERF subfamily group V, for overexpression analysis in switchgrass.

This gene was selected since the expression of its *Arabidopsis* homolog in transgenic rice resulted in modified cell wall composition (Ambavaram et al., 2011). Sequence grouping/cluster and sequence alignment analysis suggested that PvERF001 is closely related with its rice and *Arabidopsis* homologs, sharing two highly-conserved motifs: the middle motif (mm) and the C-terminal motif (cm) specific to the *Arabidopsis* SHINE clade of TFs (AtERF001, AtERF004 and AtERF005) and OsERF012 and OsERF057 (Figure 4-10A, B). Thus, the ORF of *PvERF001* was cloned and overexpressed in switchgrass producing multiple independent transgenic lines, which were confirmed based on genomic PCR for the insertion of the transgene and the hygromycin-resistance gene, as well as visualization of OFP in transgenic plants compared to the non-transgenic control lines (Figures 4-11A, 4-12A-C). Analysis of the transgene expression level by qRT-PCR showed up to 12-fold overexpression in transgenic lines (Figure 4-11B). The expression of the endogenous gene in transgenic lines was not affected compared to the non-transgenic control line (Figure 4-11C). All transgenic lines had equivalent or improved vegetative growth metrics relative to the non-transgenic control lines under greenhouse conditions, which was congruent with the relative transcript levels of the transgene (Pearson correlation for biomass weights [$R = 0.77$ at $P < 0.05$] and tiller height [$R = 0.73$ at $P = 0.06$]) (Figure 4-11B, Table 4-6, Figure 4-13). Three transgenic lines (3, 7, and 9) had increased biomass. Line 3 had statistically significant increases in four of the six growth traits and approximately twice the dry biomass of the control line (Table 4-6).

To investigate whether *PvERF001* overexpression could affect the leaf cuticular permeability, the water retention capacity in transgenic and non-transgenic control lines was analyzed in detached leaves measured in the dark to minimize transpirational water loss through stomata. Transgenic lines showed relative reduction in rate of water loss compared with the control lines (Figure 4-14). However, no tangible difference was observed in the rate of leaf chlorophyll leaching between transgenic and the control lines (data not shown). Subsequently, we analyzed whether the changes in cuticular permeability might be accompanied by changes in the expression level of genes in the cutin and wax biosynthesis pathway, in which none were observed (Figure 4-15). Moreover, overexpression of *PvERF001* in transgenic switchgrass showed relatively reduced expression of some lignin (*PvC4H* and *PvPAL*), hemicellulose (*PvCSLS2*) and cellulose (*PvCESA4*) biosynthetic genes, as well as some of the transcriptional

regulators (*PvMYB48/59* and *PvNST1*) of cell wall biosynthesis (Figure 4-16A-C). The total lignin content in R1 tillers determined by Py-MBMS of cell wall residues and in leaves determined by phloroglucinol-HCl staining did not show sizeable difference between the transgenic and non-transgenic control lines (Figures 4-17A and 4-18). Similarly, analysis of the S/G lignin monomer ratio in transgenic lines did not significantly change as compared to the non-transgenic control line (Figure 4-17B). However, significant improvement in glucose release efficiency was observed in lines 7 (10%) and 8 (16%) relative to the non-transgenic control line (Table 4-7). In contrast, none of the transgenic lines released significantly more xylose than the control. The total sugar release, however, was significantly increased in transgenic line 8 by 11% relative to the non-transgenic control (Table 4-7).

4.5 Discussion

4.5.1 Significance of AP2/ERF TFs for improvement of bioenergy crops

AP2/ERF TFs constitute one of the largest protein superfamilies in plants. These TFs play a role in regulating a wide array of developmental and growth processes. Thus, they are interesting targets for crop genetic engineering and breeding (Licausi et al., 2013; Bhatia and Bosch, 2014). Numerous TFs belonging to this superfamily have been characterized in various plant species and their potential biotechnological applications in crop improvement has focused primarily on biotic and abiotic stress tolerance (Xu et al., 2011; Licausi et al., 2013; Hoang et al., 2014). However, less effort has been made to utilize this potential for genetic improvement of bioenergy feedstocks such as switchgrass (Bhatia and Bosch, 2014). We found this lack of development to be somewhat anachronistic since these TFs are variably associated with plant growth and cell wall biosynthesis, which are directly related to two most important traits to a bioenergy crops such as switchgrass: biomass and cell wall recalcitrance.

4.5.2 Sequence-based classification of putative AP2/ERF TFs in switchgrass

With this in mind, we conducted a whole genome search for putative switchgrass AP2/ERF superfamily of TFs and found 207 members (Figure 4-1, Tables 4-2 & 4-3). Based on comparative genome analysis with the published results in rice, foxtail millet and *Arabidopsis*, the identified proteins were classified into 3 families, namely AP2, RAV and ERF with the later further divided into two subfamilies (ERF and DREB) (Nakano et al., 2006; Lata et al., 2014).

The number of genes in the DREB subfamily found in switchgrass (55) was comparable with that of rice (56), *Arabidopsis* (57) and *Populus* (66). All 3 species along with switchgrass have a singleton in their genome. Consistent with the previous report in rice (Nakano et al., 2006), the switchgrass DREB and ERF subfamilies comprise 4 and 7 groups, respectively. Moreover, based on comparative analysis of the AP2/ERF TFs between different plant species, it seems that group XI of ERF subfamily is specific to monocots as the Xb-L was reported only in dicots (Nakano et al., 2006; Liu et al., 2013). In general, the relative distribution of genes within the different groups in each subfamily appears to be conserved between the three plant species (Table 4-3). Classification of the switchgrass AP2/ERF TFs into distinct groups was clearly supported by the amino acid sequence-based dendrogram of the identified proteins suggesting robust evolutionary conservation between the superfamily among plant species.

4.5.3 *In silico* predicted gene functions and subcellular localization of AP2/ERF TFs in switchgrass

Consistent with the purported role of AP2/ERF proteins as transcriptional regulators of target genes (Magnani et al., 2004), GO analysis predicted that the majority of the switchgrass AP2/ERF genes appear to have DNA-binding activity consistent with the previous observation in foxtail millet (Lata et al., 2014). Therefore, these genes might be associated with the regulation of various biosynthetic processes as well as responses to environmental stimuli as previously demonstrated for numerous genes in other plant species (Xu et al., 2011; Mizoi et al., 2012; Licausi et al., 2013) (Figure 4-5A, B). The predicted subcellular localization pattern of AP2/ERF superfamily genes in switchgrass, which was mainly to the nucleus as would be expected for transcriptional regulators but also to the plastids and/or mitochondria in addition to the nucleus, was comparable to that reported in foxtail millet (Figure 4-5C) (Lata et al., 2014). Such multi-localization of the proteins could be attributed to post-translational modifications, protein folding or interactions with other proteins (Karniely and Pines, 2005), and might serve to facilitate the coordinated regulation of the expression of nuclear and organellar genomes (Duchene and Giege, 2012).

4.5.4 Gene and protein sequence diversity of switchgrass AP2/ERF TFs

The exon/intron structures of switchgrass *AP2/ERF* genes were analogous with that of foxtail millet (Lata et al., 2014), castor bean (Xu et al., 2013), rice and *Arabidopsis* (Nakano et al.,

2006). Consistent with these species we observed a high diversity in the distribution of the intron regions of AP2 genes versus a single or no intron in most genes in the ERF and RAV families (Figures 4-2 & 4-4). The pattern of intron distribution within the ORF and their splicing phases was highly conserved in genes within specific groups as reported in castor bean (Xu et al., 2013). Based on the analysis of amino acid sequence alignment of switchgrass AP2/ERF proteins, 14 conserved protein motifs located within the AP2 domain region and 11 motifs located outside the domain region were identified (Table 4-5). Consistent with the observation in rice, the majority of the groups or subfamilies in the AP2/ERF superfamily could be distinguished by the presence of one or more diagnostic motifs located outside the AP2 domain region (Rashid et al., 2012). These groups or subfamily specific conservation in gene structures and protein motifs supported the accuracy of the predicted cluster relationships between the switchgrass AP2/ERF TFs.

AP2/ERF TFs that function as repressors or activators of specific target genes are distinguished by the presence of conserved motifs called repression domains (RD) that are highly conserved, or by the presence of activation domains which are generally less conserved (Licausi et al., 2013). One of the characteristic motif in AP2/ERF transcriptional activators is the activation domain, EDLL motif (Tiwari et al., 2012) while repressors have unique RD namely the ERF-associated amphiphilic repression (EAR) motif (LxLxL or DLNxxP) (Kagale and Rozwadowski, 2011) and B3 repression domain (BRD: R/KLFGV) motif (Ikeda and Ohme-Takagi, 2009). Analysis of the switchgrass AP2/ERF TF sequences also indicated the presence of these motifs in many proteins (Table 4-5). For instance, many genes in group IX of ERF subfamily appear to be transcriptional activators due to the presence of motif M10, which is an EDLL-like motif. Moreover, this motif is rich in acidic amino residues which has been suggested as the characteristics of transcriptional activators (Licausi et al., 2013). Majority of the ERF subfamily TFs in group VIII and DREB subfamily TFs in group I displayed a DLNxxP-like motifs. Four TFs belonging to the AP2 family (PvERF204, PvERF205, PvERF206 and PvERF207) also displayed similar EAR motif while PvERF203 and PvERF207 harbors DLELSL and NLDLS-like repression domains, respectively. Similarly, switchgrass TFs in RAV family also displayed unique repression domain, RLFGV (Ikeda and Ohme-Takagi, 2009). ERF subfamily TFs in groups VI and VI-L share a characteristic motif at the N-terminus (M18), also known as the cytokinin responsive factor (CRF) domain in *Arabidopsis* that is also shared by rice ERF genes

belonging to same group in rice ERF subfamily (Nakano et al., 2006). Genes containing the CRF domain (VI and VI-L) were shown to be responsive to cytokinin (Rashotte et al., 2006). The distinguishing N-terminal motif in group VII ERF subfamily proteins, M25 was conserved in both *Arabidopsis* and rice as described previously (Nakano et al., 2006). This motif was shown to dictate the stability of proteins based on the level of oxygen via N-end rule pathway (Dubouzet et al., 2003; Licausi et al., 2011). DREB genes in rice with characteristic LWSY motif has been shown to function in regulation of drought, cold and salinity responsive gene expression (Dubouzet et al., 2003). Switchgrass genes belonging to group III in DREB subfamily (*PvERF133*, *PvERF134*, *PvERF135*, *PvERF136*, *PvERF137*, *PvERF139*, *PvERF140*, *PvERF141*, *PvERF142*, *PvERF143*, *PvERF145* and *PvERF146*) displayed LWSY conserved motif (M21) at the C-terminal and thus may play similar roles. No information is available in the literature on some of the conserved motifs identified here including M8, M13, M14, M15, M17 and M24 (Table 4-5), which might potentially be specific to switchgrass.

4.5.5 Diverse expression profiles of switchgrass AP2/ERF TFs and functional implications

Differential expression of genes according to developmental stages and tissue or organ types may provide an insight into the specialized biological processes that are taking place in the specific plant parts (Zhang et al., 2013; Cassan-Wang et al., 2013). Therefore, exploring patterns of gene expression and plant development may be useful to determine which AP2/ERF TF genes should be most useful to alter for bioenergy applications.

Based on the observed pattern of gene expression during seed germination, it seems that several genes belonging to the DREB and ERF subfamilies as well as AP2 family might be associated with the regulation of seed germination processes in switchgrass (Figure 4-6). The dramatic increase in the expression level of some genes in the DREB and ERF subfamilies and AP2 family during the first 48 h after imbibition implies potentially important roles in the regulation of water uptake into the quiescent seeds and ABA signaling during germination (Sreenivasulu et al., 2008; Howell et al., 2009). Similarly, an ERF gene from *Fagus sylvatica* has been suggested to be associated more with the transitioning from dormancy to the initiation of germination than regulation of stress response (Jiménez et al., 2005). The fact that some of the DREB genes are highly expressed after imbibition while others are expressed at later stages might have functional

significance, such as transcriptional regulation of activation of storage reserve mobilization during seed germination (Sreenivasulu et al., 2008) (Figure 4-6). These responses may also serve as part of defending seedlings against attack and stress. DREB TFs have been reported to regulate the expression of several stress-inducible genes in early seed germination (Krishnaswamy et al., 2011). Activation of genes responsible for cell wall modification have been reported to be key during the initiation of seed germination (Sreenivasulu et al., 2008). Transcriptional upregulation of ERF (*PvERF057*, *PvERF068*, *PvERF088* and *PvERF119*) and DREB (*PvERF101*, *PvERF102* and *PvERF148*) subfamily genes as well as AP2 family genes (*PvERF193*, *PvERF201* and *PvERF204*) during early phase of seed germination (radicle emergence) as well as in vascular bundle tissues and internode sections may also hint that these genes may have some intrinsic association with the regulatory machinery of cell wall formation/lignification. Consistently, the expression of many genes in cell wall biosynthesis and remodeling pathways along with related TFs were upregulated during early germination phase in barley (An and Lin, 2011).

Intriguingly, the role of AP2/ERF TFs in the regulation of cell wall biosynthesis is rather less known compared to their role in stress adaptation. The expression profiles of the switchgrass AP2/ERF TFs suggested that several putative TFs may be associated with the transcriptional regulation of cell wall biosynthetic machinery. The observation that about 14 genes in the DREB and 17 in ERF subfamilies as well as 3 genes belonging to the AP2 family showing robust expression in tissues or organs undergoing active lignification (vascular bundles, top or middle internode sections as well as roots) but less robust expression in less lignified tissues (leaves) support this assertion (Figure 4-9). It should be noted that the transcript levels of several of these genes showed a relative increase with the developmental stage of the plants (Figures 4-7 & 4-9) while exhibiting only marginal expression in less lignified tissues such as inflorescence meristem and germinating seedlings (Figures 4-6 & 4-8). Differential gene expression profiling between elongating and non-elongating internodes in maize was used to identify a total of 7 Ap2/ERF TFs that are highly expressed in non-elongating internodes undergoing secondary wall development suggesting that these genes may involve in the regulation of secondary cell wall formation (Bosch et al., 2011). Moreover, recent study in *Arabidopsis* and rice identified several putative secondary cell wall-related AP2/ERF TFs based on preferential expression in secondary

cell wall-related tissues and coexpression analysis (Cassan-Wang et al., 2013; Hirano et al., 2013a; Bhatia and Bosch, 2014). Some of the switchgrass genes identified in this study (PvERF037, PvERF115, PvERF116, PvERF143, PvERF148 and PvERF164) appear to be putative homologs of maize, rice and *Arabidopsis* genes identified in the aforementioned studies. Overexpression of *Populus* ERF genes in wood-forming tissues of hybrid aspen was recently shown to result in modified stem growth (including increased stem diameter following the overexpression of 5 different ERF genes), reduced lignification and enhanced carbohydrate content (cellulose) in the wood of transgenic lines hinting that these TFs may indeed interact with the transcriptional machinery regulating cell wall biosynthesis (Vahala et al., 2013). Another evidence supporting this is a recent study suggesting that an ERF TF from loquat fruit (*Eriobotrya japonica*) (EjAP2-1) is an indirect transcriptional repressor of lignin biosynthesis via interaction with EjMYB1 TFs (Zeng et al., 2015).

4.5.6 Overexpression of *PvERF001* improved biomass productivity and sugar release efficiency in switchgrass

Based on global gene coexpression analysis, the rice homolog of AtSHN2, OsSHN (OsERF057) was proposed to have a native association with cell wall regulatory and biosynthetic pathways, yet this was not experimentally verified (Ambavaram et al., 2011). In this study, we investigated whether PvERF001, the closest putative switchgrass homolog of these genes based on clustering, sequence alignment analysis and the sharing of conserved motifs (mm and cm) specific to *Arabidopsis* SHN clade of TFs and the rice SHN, may participate in the regulation of cell wall biosynthesis (Figure 4-10). Our results suggest that PvERF001 may not be directly involved in the regulation of cell wall biosynthesis though its transgenic overexpression resulted in increased sugar release efficiency (Figure 4-16, Table 4-7). Despite the observed reduction in relative expression of some lignin biosynthetic genes and their transcriptional regulators in switchgrass that seem to relate with the results in rice overexpressing *AtSHN2*, no significant changes in the lignin content and composition was detected in transgenic switchgrass in contrast to the reduced lignin content observed in rice overexpressing *AtSHN2* (Ambavaram et al., 2011) (Figures 4-16, 4-17A & 4-18). The increased sugar release might be attributed to altered storage carbohydrates such as starches as recently reported in *Arabidopsis* where ectopic expression of rice ERF TF (SUB1A-1) gene resulted in improved enzymatic saccharification efficiency via increased level

of starch (Nunez-Lopez et al., 2015). Similar results were obtained from overexpression of maize *corngrass1* microRNA in switchgrass (Chuck et al., 2011). However, whether PvERF001 is associated with starch biosynthesis remains to be determined. Moreover, in contrast to the previous reports where heterologous expression of *AtSHN2* in rice did not significantly affect the growth characteristics of transgenic lines (Ambavaram et al., 2011), overexpression of *PvERF001* resulted in increased plant growth including plant height, stem diameter, aboveground biomass weight in transgenic lines (Table 4-6). The discrepancy in lignin content and biomass productivity traits between the *AtSHN2* and *PvERF001* may indicate the differences in functional specialization between the two genes in monocots and dicots even though sequence analysis seem to suggest that they might be orthologs. The fact that overexpression of *AtSHN* genes in *Arabidopsis* rather showed association with the regulation of wax, cutin and pectin biosynthesis supports this assertion (Aharoni et al., 2004; Shi et al., 2011). Moreover, recent study showed that the homolog of *Arabidopsis SHN* genes in tomato (*SlERF52*) was expressed mainly in the abscission zone and functionally associated with the regulation of the pedicel abscission zone-specific transcription of genes including cell wall-hydrolytic enzymes (polygalacturonase and Cellulase) required for abscission (Nakano et al., 2014). These differences in the expression pattern and function may suggest functional divergence between *SlERF52* and its *Arabidopsis* homologs. Functional divergence between homologous TFs in monocots and dicots has also been reported in previous studies involving the homologs of *AtMYB58/63*, which is a known activator of lignin biosynthesis that did not appear to play similar roles in rice (Hirano et al., 2013b).

A recent study involving overexpression of rice homolog of *AtSHN2*, *OsSHN*, in rice showed enhanced tolerance of transgenic plants to water deprivation and association of the gene with the regulation of wax and cutin biosynthesis and hence named rice wax synthesis regulatory gene (*OsWR2*) (Zhou et al., 2014). The closest homolog of this gene, *OsERF012* (*OsWR1*), was also shown to be induced by drought stress and involved in the regulation of wax synthesis (Wang et al., 2012). Therefore, we examined whether *PvERF001* might be involved in the regulation of wax and cutin biosynthesis. Consistent with previous studies in rice, relative increase in leaf water retention capacity was detected in transgenic plants though the effect on the expression of wax and cutin biosynthetic genes was minimal (Figures 4-14 & 4-15). Possible explanation for

the observed differences between overexpression of rice and switchgrass homologs might be an indication of the functional divergence in the switchgrass genes due to gene duplication. This may explain the discrepancy between transgenic rice overexpressing rice *SHN* (*OsWR2*) exhibiting reduction in plant height but increase in the number of tillers (Zhou et al., 2014) and transgenic switchgrass overexpressing *PvERF001* showing increased plant height but no difference in number of tillers. This suggests that ERF genes might functionally be highly diversified and *PvERF001* may be part of a different pathway than we anticipated such as regulation of responses to biotic stress or other abiotic stress or regulation of cell elongation or division in coordination with the cytokinin pathway, with the latter perhaps explaining the observed increase in biomass and vegetative growth in transgenic lines.

In summary, the expression profiling of the switchgrass AP2/ERF genes provides baseline information as to the putative roles of these genes and thus a useful resource for future reverse genetic studies to characterize genes for economically important bioenergy crops. With the current advancements in switchgrass research and establishment of efficient transformation system, this inventory of genes along with the information provided here could facilitate our understanding regarding the functional roles of AP2/ERF TFs in plant growth and development. Furthermore, it would aid in the identification of potential target genes that may be used to improve stress adaptation, plant productivity and sugar release efficiency in bioenergy feedstocks such as switchgrass. The increased biomass yield and sugar release efficiency from overexpressing *PvERF001* highlight the potential of these TFs for improvement of bioenergy feedstocks.

4.6 Acknowledgements

We thank Crissa Doeppke, Geoff Turner, Steve Decker, Melvin Tucker, and Erica Gjersing for their assistance with lignin and sugar release assays. We also thank Susan Holladay for her assistance with data entry into LIMS. This work was supported by funding from the BioEnergy Science Center. The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. We also thank Tennessee Agricultural Experiment Station for providing partial financial support for WAW.

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.
- Ambavaram MM, Krishnan A, Trijatmiko KR, Pereira A. 2011. Coordinated activation of cellulose and repression of lignin biosynthesis pathways in rice. *Plant Physiology* 155:916-931.
- Ambavaram MMR, Basu S, Krishnan A, Ramegowda V, Batlang U, Rahman L, Baisakh N, Pereira A. 2014. Coordinated regulation of photosynthesis in rice increases yield and tolerance to environmental stress. *Nature Communication* 5:1-14.
- An Y-Q, and Lin L. 2011. Transcriptional regulatory programs underlying barley germination and regulatory functions of gibberellin and abscisic acid. *BMC Plant Biology* 11:105-105.
- Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings /International Conference on Intelligent Systems for Molecular Biology; ISMB International Conference on Intelligent Systems for Molecular Biology* 2:28-36
- Baxter H, Poovaiah C, Yee K, Mazarei M, Rodriguez M, Jr., Thompson O, Shen H, Turner G, Decker S, Sykes RW, Chen F, Davis M, Mielenz J, Davison B, Dixon R, Stewart CN Jr. 2015. Field evaluation of transgenic switchgrass plants overexpressing *PvMYB4* for reduced biomass recalcitrance. *BioEnergy Research*:1-12.
- Baxter HL, Mazarei M, Labbe N, Kline LM, Cheng Q, Windham MT, Mann DG, Fu C, Ziebell A, Sykes RW, Rodriguez M, Jr., Davis MF, Mielenz JR, Dixon RA, Wang ZY, Stewart CN Jr. 2014. Two-year field analysis of reduced recalcitrance transgenic switchgrass. *Plant Biotechnology Journal* 12:914-924.
- Bhatia R, and Bosch M. 2014. Transcriptional regulators of *Arabidopsis* secondary cell wall formation: tools to re-program and improve cell wall traits. *Frontiers in Plant Science* 5:192.

- Bosch M, Mayer C-D, Cookson A, Donnison IS. 2011. Identification of genes involved in cell wall biogenesis in grasses by differential gene expression profiling of elongating and non-elongating maize internodes. *Journal of Experimental Botany* 62:3545–3561.
- Bouaziz D, Pirrello J, Charfeddine M, Hammami A, Jbir R, Dhieb A, Bouzayen M, and Gargouri-Bouazid R. 2013. Overexpression of *StDREB1* transcription factor increases tolerance to salt in transgenic potato plants. *Molecular Biotechnology* 54:803-817.
- Burris J, Mann DJ, Joyce B, Stewart CN Jr. 2009. An improved tissue culture system for embryogenic callus production and plant regeneration in switchgrass (*Panicum virgatum* L.). *BioEnergy Research* 2:267-274.
- Cassan-Wang H, Goué N, Saidi MN, Legay S, Sivadon P, Goffner D, Grima-Pettenati J. 2013. Identification of novel transcription factors regulating secondary cell wall formation in *Arabidopsis*. *Frontiers in Plant Science* 4:189.
- Chuck GS, Tobias C, Sun L, Kraemer F, Li C, Dibble D, Arora R, Bragg JN, Vogel JP, Singh S, Simmons BA, Pauly M, and Hake S. 2011. Overexpression of the maize *Corngrass1* microRNA prevents flowering, improves digestibility, and increases starch content of switchgrass. *Proceedings of the National Academy of Sciences USA* 108:17550-17555.
- Conesa A, Gotz S. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics* 2008:619832.
- Decker SR, Carlile M, Selig MJ, Doeppke C, Davis M, Sykes R, Turner G, Ziebell A. 2012. Reducing the effect of variable starch levels in biomass recalcitrance screening. *Methods in Molecular Biology* (Clifton, NJ) 908:181-195.
- Dong L, Cheng Y, Wu J, Cheng Q, Li W, Fan S, Jiang L, Xu Z, Kong F, Zhang D, Xu P, and Zhang S. 2015. Overexpression of *GmERF5*, a new member of the soybean EAR motif-containing ERF transcription factor, enhances resistance to *Phytophthora sojae* in soybean. *Journal of Experimental Botany* 66:2635-2647.

Dubouzet JG, Sakuma Y, Ito Y, Kasuga M, Dubouzet EG, Miura S, Seki M, Shinozaki K, and Yamaguchi-Shinozaki K. 2003. *OsDREB* genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *Plant Journal* 33:751-763.

Duchene AM, Giege P. 2012. Dual localized mitochondrial and nuclear proteins as gene expression regulators in plants? *Frontiers in Plant Science* 3:221.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792-1797.

Elliott RC, Betzner AS, Huttner E, Oakes MP, Tucker WQ, Gerentes D, Perez P, Smyth DR. 1996. *AINTEGUMENTA*, an *APETALA2*-like gene of *Arabidopsis* with pleiotropic roles in ovule development and floral organ growth. *Plant Cell* 8:155-168

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

Fu C, Xiao X, Xi Y, Ge Y, Chen F, Bouton J, Dixon R, Wang ZY. 2011b. Downregulation of *Cinnamyl Alcohol Dehydrogenase* (*CAD*) leads to improved saccharification efficiency in switchgrass. *BioEnergy Research* 4:153-164.

Fujita Y, Fujita M, Shinozaki K, Yamaguchi-Shinozaki K. 2011. ABA-mediated transcriptional regulation in response to osmotic stress in plants. *Journal of Plant Research* 124:509-525.

Grewal D, Gill R, Gosal SS. 2006. Influence of antibiotic cefotaxime on somatic embryogenesis and plant regeneration in indica rice. *Biotechnology Journal* 1:1158-1162.

Guo AY, Zhu QH, Chen X, Luo JC. 2007. GSDS: a gene structure display server. *Yi chuan = Hereditas / Zhongguo yi chuan xue hui bian ji* 29:1023-1026.

- Guo ZJ, Chen XJ, Wu XL, Ling JQ, Xu P. 2004. Overexpression of the AP2/EREBP transcription factor *OPBP1* enhances disease resistance and salt tolerance in tobacco. *Plant Molecular Biology* 55:607-618.
- Hardin CF, Fu C, Hisano H, Xiao X, Shen H, Stewart CN Jr., Parrott W, Dixon R, Wang ZY. 2013. Standardization of switchgrass sample collection for cell wall and biomass trait analysis. *BioEnergy Research* 6:755-762.
- Hattori Y, Nagai K, Furukawa S, Song XJ, Kawano R, Sakakibara H, Wu J, Matsumoto T, Yoshimura A, Kitano H, Matsuoka M, Mori H, Ashikari M. 2009. The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water. *Nature* 460:1026-1030.
- Hirano K, Aya K, Morinaka Y, Nagamatsu S, Sato Y, Antonio BA, Namiki N, Nagamura Y, Matsuoka M. 2013a. Survey of genes involved in rice secondary cell wall formation through a co-expression network. *Plant & Cell Physiology* 54:1803-1821.
- Hirano K, Kondo M, Aya K, Miyao A, Sato Y, Antonio BA, Namiki N, Nagamura Y, Matsuoka M. 2013b. Identification of transcription factors involved in rice secondary cell wall formation. *Plant & Cell Physiology* 54:1791-1802.
- Hoang XLT, Thu NBA, Thao NP, and Tran L-SP. 2014. Transcription factors in abiotic stress responses: their potentials in crop improvement. In: Ahmad P, Wani MR, Azooz MM & Tran L-SP. (eds.), *Improvement of crops in the era of climatic changes*. Springer, New York, 337-366.
- Hong JP, Kim WT. 2005. Isolation and functional characterization of the *Ca-DREBLP1* gene encoding a dehydration-responsive element binding-factor-like protein 1 in hot pepper (*Capsicum annuum* L. cv. *Pukang*). *Planta* 220 (6):875-888.
- Horstman A, Willemsen V, Boutilier K, and Heidstra R. 2014. AINTEGUMENTA-LIKE proteins: hubs in a plethora of networks. *Trends in Plant Science* 19:146-157.
- Howell KA, Narsai R, Carroll A, Ivanova A, Lohse M, Usadel B, Millar AH, Whelan J. 2009. Mapping metabolic and transcript temporal switches during germination in rice highlights

- specific transcription factors and the role of RNA instability in the germination process. *Plant Physiology* 149:961-980.
- Ikeda M, Ohme-Takagi M. 2009. A novel group of transcriptional repressors in *Arabidopsis*. *Plant & Cell Physiology* 50:970-975.
- Ito Y, Katsura K, Maruyama K, Taji T, Kobayashi M, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. 2006. Functional analysis of rice DREB1/CBF-type transcription factors involved in cold-responsive gene expression in transgenic rice. *Plant & Cell Physiology* 47:141-153.
- Jaglo-Ottosen KR, Gilmour SJ, Zarka DG, Schabenberger O, Thomashow MF. 1998. *Arabidopsis* CBF1 overexpression induces COR genes and enhances freezing tolerance. *Science* 280:104-106.
- Jiménez JA, Rodríguez D, Calvo AP, Mortensen LC, Nicolás G, Nicolás C. 2005. Expression of a transcription factor (FsERF1) involved in ethylene signalling during the breaking of dormancy in *Fagus sylvatica* seeds. *Physiologia Plantarum* 125:373-380.
- Jofuku KD, Omidyar PK, Gee Z, Okamuro JK. 2005. Control of seed mass and seed yield by the floral homeotic gene *APETALA2*. *Proceedings of the National Academy of Sciences USA* 102:3117-3122.
- Kagale S, Rozwadowski K. 2011. EAR motif-mediated transcriptional repression in plants: an underlying mechanism for epigenetic regulation of gene expression. *Epigenetics : Official Journal of the DNA Methylation Society* 6:141-146.
- Karniely S, Pines O. 2005. Single translation--dual destination: mechanisms of dual protein targeting in eukaryotes. *EMBO Reports* 6:420-425.
- King ZR, Bray AL, Lafayette PR, Parrott WA. 2014. Biolistic transformation of elite genotypes of switchgrass (*Panicum virgatum* L.). *Plant Cell Reports* 33:313-322.

Krishnaswamy S, Verma S, Rahman MH, Kav NN. 2011. Functional characterization of four APETALA2-family genes (*RAP2.6*, *RAP2.6L*, *DREB19* and *DREB26*) in *Arabidopsis*. *Plant Molecular Biology* 75:107-127.

Lasserre E, Jobet E, Llauro C, Delseny M. 2008. AtERF38 (At2g35700), an AP2/ERF family transcription factor gene from *Arabidopsis thaliana*, is expressed in specific cell types of roots, stems and seeds that undergo suberization. *Plant Physiology and Biochemistry* 46:1051-1061.

Lata C, Mishra AK, Muthamilarasan M, Bonthala VS, Khan Y, Prasad M. 2014. Genome-wide investigation and expression profiling of AP2/ERF transcription factor superfamily in foxtail millet (*Setaria italica* L.). *PloS One* 9:e113092.

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

Licausi F, Giorgi FM, Zenoni S, Osti F, Pezzotti M, Perata P. 2010. Genomic and transcriptomic analysis of the AP2/ERF superfamily in *Vitis vinifera*. *BMC Genomics* 11:719.

Licausi F, Kosmacz M, Weits DA, Giuntoli B, Giorgi FM, Voesenek LACJ, Perata P, and Van Dongen JT. 2011. Oxygen sensing in plants is mediated by an N-end rule pathway for protein destabilization. *Nature* 479:419-422.

Licausi F, Ohme-Takagi M, Perata P. 2013. APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. *New Phytologist* 199:639-649.

Liu Z, Kong L, Zhang M, Lv Y, Liu Y, Zou M, Lu G, Cao J, Yu X. 2013. Genome-wide identification, phylogeny, evolution and expression patterns of AP2/ERF genes and cytokinin response factors in *Brassica rapa* ssp. *pekinensis*. *PloS One* 8:e83444.

Magnani E, Sjolander K, Hake S. 2004. From endonucleases to transcription factors: evolution of the AP2 DNA binding domain in plants. *Plant Cell* 16:2265-2277.

- Mann DG, Lafayette PR, Abercrombie LL, King ZR, Mazarei M, Halter MC, Poovaiah CR, Baxter H, Shen H, Dixon RA, Parrott WA, Stewart CN Jr. 2012. Gateway-compatible vectors for high-throughput gene functional analysis in switchgrass (*Panicum virgatum* L.) and other monocot species. *Plant Biotechnology Journal* 10:226-236.
- Mittal A, Gampala SS, Ritchie GL, Payton P, Burke JJ, Rock CD. 2014. Related to ABA-Insensitive3 (ABI3)/Viviparous1 and AtABI5 transcription factor coexpression in cotton enhances drought stress adaptation. *Plant Biotechnology Journal* 12:578-589.
- Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. 2012. AP2/ERF family transcription factors in plant abiotic stress responses. *BBA-Gene Regulatory Mechanisms* 1819:86-96.
- Moore KJ, Moser LE, Vogel KP, Waller SS, Johnson BE, Pedersen JF. 1991. Describing and quantifying growth stages of perennial forage grasses. *Agronomy Journal* 83:1073-1077.
- Murashige T, Skoog F. 1962. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia Plantarum* 15:473-497.
- Nakano T, Fujisawa M, Shima Y, Ito Y. 2014. The AP2/ERF transcription factor SlERF52 functions in flower pedicel abscission in tomato. *Journal of Experimental Botany* 65:3111-3119.
- Nakano T, Suzuki K, Fujimura T, Shinshi H. 2006. Genome-wide analysis of the ERF gene family in *Arabidopsis* and rice. *Plant Physiology* 140:411-432.
- Nunez-Lopez L, Aguirre-Cruz A, Barrera-Figueroa BE, and Pena-Castro JM. 2015. Improvement of enzymatic saccharification yield in *Arabidopsis thaliana* by ectopic expression of the rice SUB1A-1 transcription factor. *PeerJ* 3:e817.
- Oh SJ, Kim YS, Kwon CW, Park HK, Jeong JS, Kim JK. 2009. Overexpression of the transcription factor AP37 in rice improves grain yield under drought conditions. *Plant Physiology* 150:1368-1379.
- Ohme-Takagi M, Shinshi H. 1995. Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. *Plant Cell* 7:173-182.

- Qin F, Kakimoto M, Sakuma Y, Maruyama K, Osakabe Y, Tran LS, Shinozaki K, Yamaguchi-Shinozaki K. 2007. Regulation and functional analysis of ZmDREB2A in response to drought and heat stresses in *Zea mays* L. *Plant Journal: for Cell and Molecular Biology* 50:54-69.
- Rashid M, Guangyuan H, Guangxiao Y, Hussain J, Xu Y. 2012. AP2/ERF transcription factor in rice: genome-wide canvas and syntenic relationships between monocots and eudicots. *Evolutionary Bioinformatics Online* 8:321-355.
- Rashotte AM, Mason MG, Hutchison CE, Ferreira FJ, Schaller GE, Kieber JJ. 2006. A subset of *Arabidopsis* AP2 transcription factors mediates cytokinin responses in concert with a two-component pathway. *Proceedings of the National Academy of Sciences USA* 103:11081-11085.
- Shen H, Fu C, Xiao X, Ray T, Tang Y, Wang Z, Chen F. 2009. Developmental control of lignification in stems of lowland switchgrass variety Alamo and the effects on saccharification efficiency. *BioEnergy Research* 2:233-245.
- Shen H, He X, Poovaiah CR, Wuddineh WA, Ma J, Mann DG, Wang H, Jackson L, Tang Y, Stewart CN Jr., Chen F, Dixon RA. 2012. Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks. *New Phytologist* 193:121-136.
- Shen H, Mazarei M, Hisano H, Escamilla-Trevino L, Fu C, Pu Y, Rudis MR, Tang Y, Xiao X, Jackson L, Li G, Hernandez T, Chen F, Ragauskas AJ, Stewart CN Jr., Wang ZY, Dixon RA. 2013a. A genomics approach to deciphering lignin biosynthesis in switchgrass. *Plant Cell* 25:4342-4361.
- Shen H, Poovaiah CR, Ziebell A, Tschaplinski TJ, Pattathil S, Gjersing E, Engle NL, Katahira R, Pu Y, Sykes RW, Chen F, Ragauskas AJ, Mielenz JR, Hahn MG, Davis M, Stewart CN Jr., Dixon RA. 2013b. Enhanced characteristics of genetically modified switchgrass (*Panicum virgatum* L.) for high biofuel production. *Biotechnology for Biofuels* 6:71.

- Shi JX, Malitsky S, De Oliveira S, Branigan C, Franke RB, Schreiber L, and Aharoni A. 2011. SHINE transcription factors act redundantly to pattern the archetypal surface of *Arabidopsis* flower organs. *PLoS Genetics* 7:e1001388.
- Sreenivasulu N, Usadel B, Winter A, Radchuk V, Scholz U, Stein N, Weschke W, Strickert M, Close TJ, Stitt M, Graner A, Wobus U. 2008. Barley grain maturation and germination: metabolic pathway and regulatory network commonalities and differences highlighted by new MapMan/PageMan profiling tools. *Plant Physiology* 146:1738-1758.
- Sun S, Yu JP, Chen F, Zhao TJ, Fang XH, Li YQ, Sui SF. 2008. TINY, a dehydration-responsive element (DRE)-binding protein-like transcription factor connecting the DRE- and ethylene-responsive element-mediated signalling pathways in *Arabidopsis*. *The Journal of Biological Chemistry* 283:6261-6271.
- Sykes RW, Yung M, Novaes E, Kirst M, Peter G, Davis M. 2009. High-throughput screening of plant cell-wall composition using pyrolysis molecular beam mass spectroscopy. *Methods in Molecular Biology* (Clifton, NJ) 581:169-183.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30:2725-2729.
- Tiwari SB, Belachew A, Ma SF, Young M, Ade J, Shen Y, Marion CM, Holtan HE, Bailey A, Stone JK, Edwards L, Wallace AD, Canales RD, Adam L, Ratcliffe OJ, Repetti PP. 2012. The EDLL motif: a potent plant transcriptional activation domain from AP2/ERF transcription factors. *Plant Journal: for Cell and Molecular Biology* 70:855-865.
- Tschaplinski TJ, Standaert RF, Engle NL, Martin MZ, Sangha AK, Parks JM, Smith JC, Samuel R, Jiang N, Pu Y, Ragauskas AJ, Hamilton CY, Fu C, Wang ZY, Davison BH, Dixon RA, Mielenz JR. 2012. Down-regulation of the *caffeic acid O-methyltransferase* gene in switchgrass reveals a novel monolignol analog. *Biotechnology for Biofuels* 5:71.
- Vahala J, Felten J, Love J, Gorzsas A, Gerber L, Lamminmaki A, Kangasjarvi J, Sundberg B. 2013. A genome-wide screen for ethylene-induced ethylene response factors (ERFs) in hybrid

aspen stem identifies ERF genes that modify stem growth and wood properties. *New Phytologist* 200:511-522.

Van Raemdonck D, Pesquet E, Cloquet S, Beeckman H, Boerjan W, Goffner D, El Jaziri M, Baucher M. 2005. Molecular changes associated with the setting up of secondary growth in aspen. *Journal of Experimental Botany* 56:2211–2227.

Wang L, Wang C, Qin L, Liu W, and Wang Y. 2015. ThERF1 regulates its target genes via binding to a novel cis-acting element in response to salt stress. *Journal of Integrative Plant Biology*, in press. doi: 10.1111/jipb.12335.

Wang Y, Wan L, Zhang L, Zhang Z, Zhang H, Quan R, Zhou S, Huang R. 2012. An ethylene response factor OsWR1 responsive to drought stress transcriptionally activates wax synthesis related genes and increases wax production in rice. *Plant Molecular Biology* 78:275-288.

Wuddineh WA, Mazarei M, Zhang J, Poovaiah CR, Mann DG, Ziebell A, Sykes RW, Davis MF, Udvardi MK, Stewart CN Jr. 2015. Identification and overexpression of *gibberellin 2-oxidase* (GA2ox) in switchgrass (*Panicum virgatum* L.) for improved plant architecture and reduced biomass recalcitrance. *Plant Biotechnology Journal*. 13:636-647.

Xu W, Li F, Ling L, Liu A. 2013. Genome-wide survey and expression profiles of the AP2/ERF family in castor bean (*Ricinus communis* L.). *BMC Genomics* 14:785.

Xu ZS, Chen M, Li LC, Ma YZ. 2011. Functions and application of the AP2/ERF transcription factor family in crop improvement. *Journal of Integrative Plant Biology* 53:570-585.

Yuan JS, Tiller KH, Al-Ahmad H, Stewart NR, Stewart CN Jr. 2008. Plants to power: bioenergy to fuel the future. *Trends in Plant Science* 13:421-429.

Zeng JK, Li X, Xu Q, Chen JY, Yin XR, Ferguson IB, and Chen KS. 2015. *EjAP2-1*, an AP2/ERF gene, is a novel regulator of fruit lignification induced by chilling injury, via interaction with EjMYB transcription factors. *Plant Biotechnology Journal* doi: 10.1111/pbi.12351.

- Zhang CH, Shangguan LF, Ma RJ, Sun X, Tao R, Guo L, Korir NK, Yu ML. 2012. Genome-wide analysis of the AP2/ERF superfamily in peach (*Prunus persica*). *Genetics and Molecular Research: GMR* 114789-4809.
- Zhang H, Liu W, Wan L, Li F, Dai L, Li D, Zhang Z, Huang R. 2010a. Functional analyses of ethylene response factor JERF3 with the aim of improving tolerance to drought and osmotic stress in transgenic rice. *Transgenic Research* 19:809-818.
- 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 Journal: for Cell and Molecular Biology* 74:160-173.
- Zhang Z, Huang R. 2010. Enhanced tolerance to freezing in tobacco and tomato overexpressing transcription factor *TERF2/LeERF2* is modulated by ethylene biosynthesis. *Plant Molecular Biology* 73:241-249.
- Zhang Z, Li F, Li D, Zhang H, Huang R. 2010b. Expression of ethylene response factor *JERF1* in rice improves tolerance to drought. *Planta* 232:765-774.
- Zhou X, Jenks M, Liu J, Liu A, Zhang X, Xiang J, Zou J, Peng Y, Chen X. 2014. Overexpression of transcription factor *OsWR2* regulates wax and cutin biosynthesis in rice and enhances its tolerance to water deficit. *Plant Molecular Biology Reporter* 32:719-731.
- Zhuang J, Cai B, Peng RH, Zhu B, Jin XF, Xue Y, Gao F, Fu XY, Tian YS, Zhao W, Qiao YS, Zhang Z, Xiong AS, Yao QH. 2008. Genome-wide analysis of the AP2/ERF gene family in *Populus trichocarpa*. *Biochemical and Biophysical Research Communications* 371:468-474.

Appendix

Figures and tables

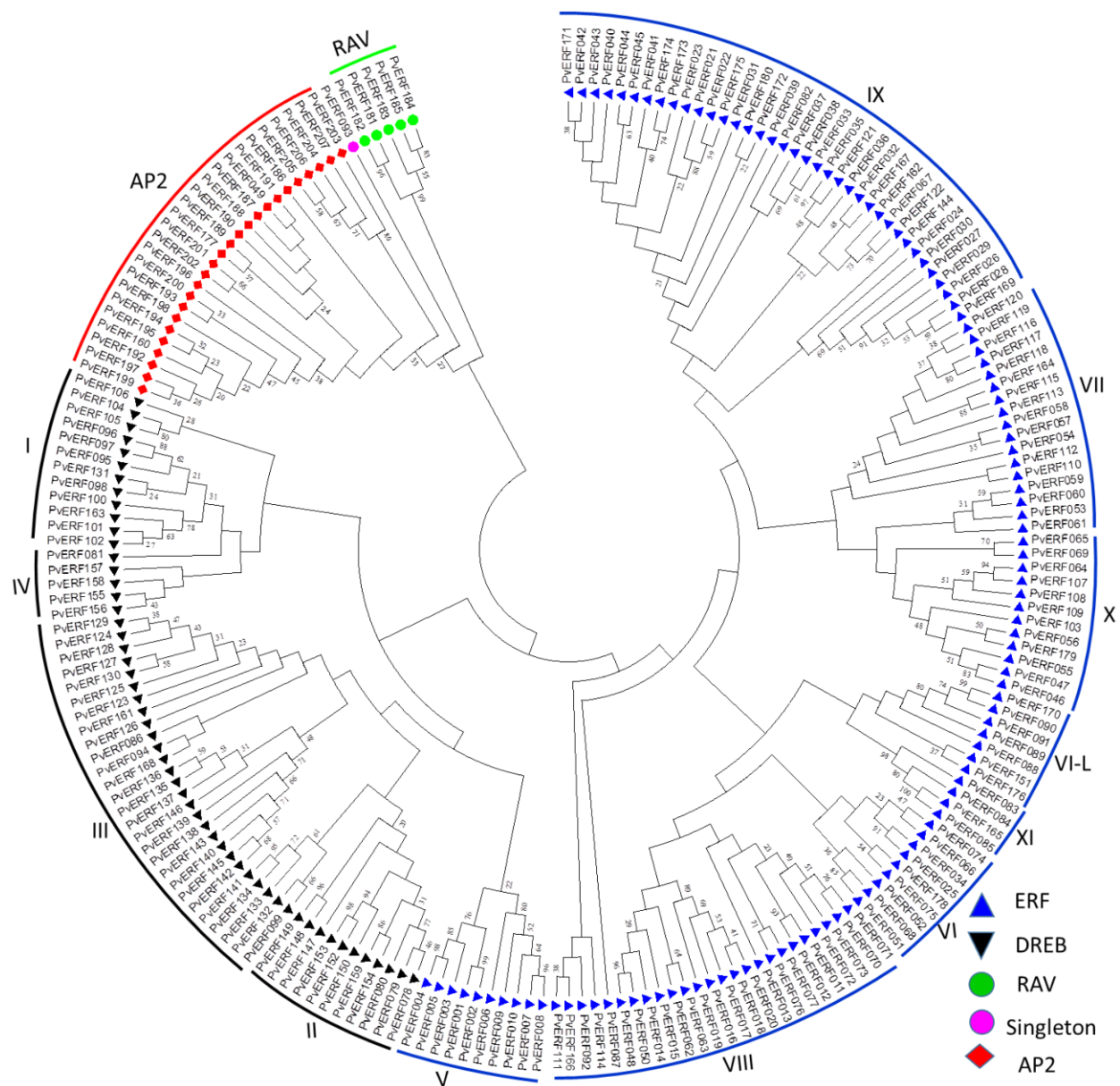


Figure 4-1 An unrooted phylogenetic tree of switchgrass AP2/ERF proteins. The deduced amino acid sequences were imported into MEGA6.0 program and aligned using MUSCLE program. The tree was constructed by a neighbor joining method with bootstrap replicates of 5000. The numbers on branches indicate bootstrap support values. The families, subfamilies and groups within each subfamilies are indicated in the tree. The list of switchgrass sequences used to construct this tree along with their gene identifier names were presented in Table 4-2.

Figure 4-2 The schematic representation of protein and gene structures of switchgrass ERF subfamily. A) Distribution of conserved motifs within the deduced amino acid sequences as determined by MEME tool (Bailey and Elkan, 1994). The colored boxes represent the conserved motifs. B) The gene features as visualized by the gene structure display server (Guo et al., 2007). The coding DNA sequence (CDS) and the untranslated regions (UTR) are shown by filled dark-blue and red boxes, respectively. The introns are shown by thick black lines. The splicing phases of the introns are indicated by numbers. The Roman numerals indicate the group of the genes within the subfamily.

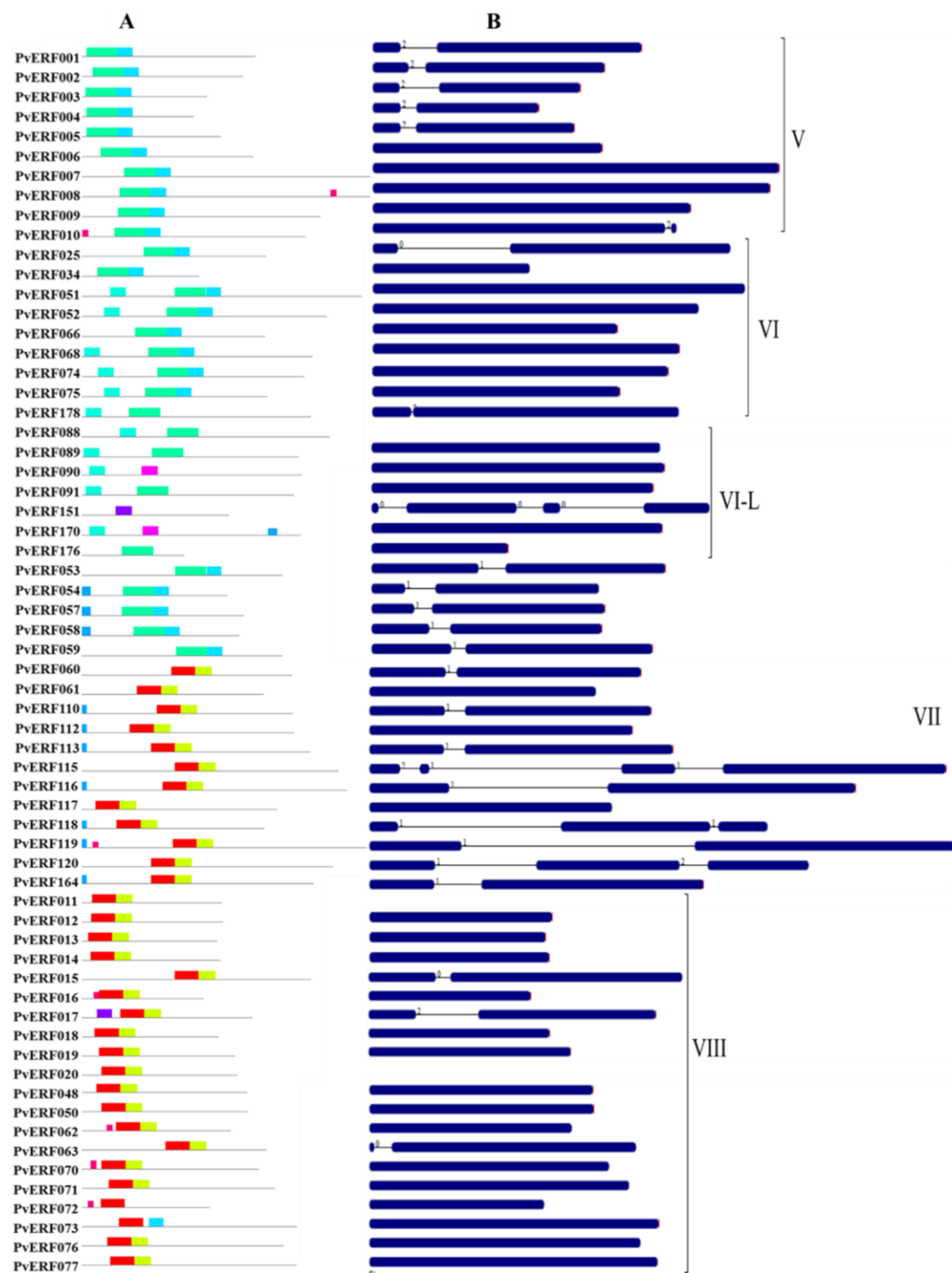


Figure 4-2 Continued.

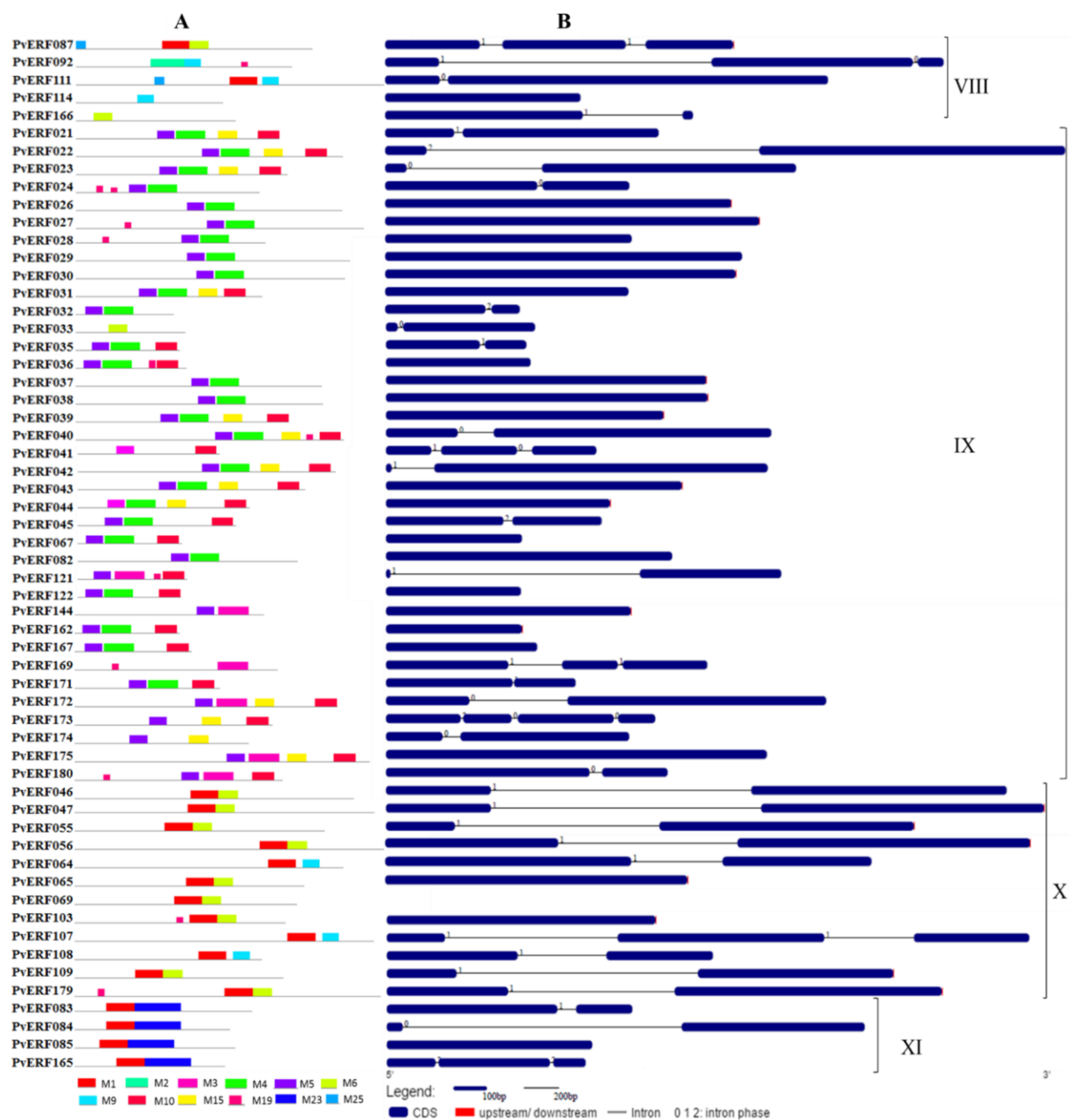


Figure 4-2 Continued.

Figure 4-3 The schematic representation of protein and gene structures of switchgrass DREB subfamily. A) Distribution of conserved motifs within the deduced amino acid sequences as determined by MEME tool (Bailey and Elkan, 1994). The colored boxes represent the conserved motifs. B) The gene features as visualized by the gene structure display server (Guo et al., 2007). The coding DNA sequence (CDS) and the untranslated regions (UTR) are shown by filled dark-blue and red boxes, respectively. The introns are shown by thick black lines. The splicing phases of the introns are indicated by numbers. The Roman numerals indicate the group of the genes within the subfamily.

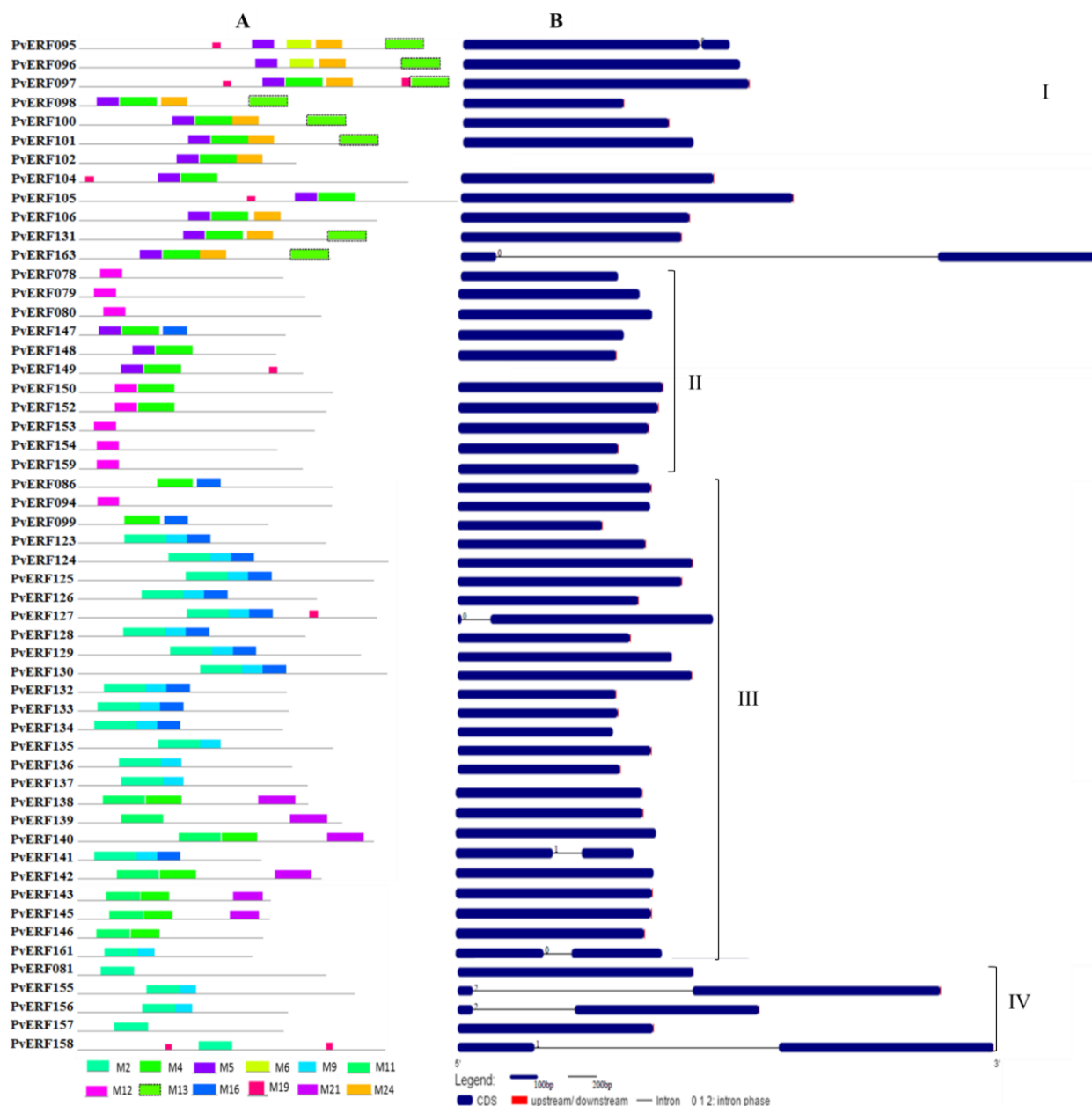


Figure 4-3 Continued.

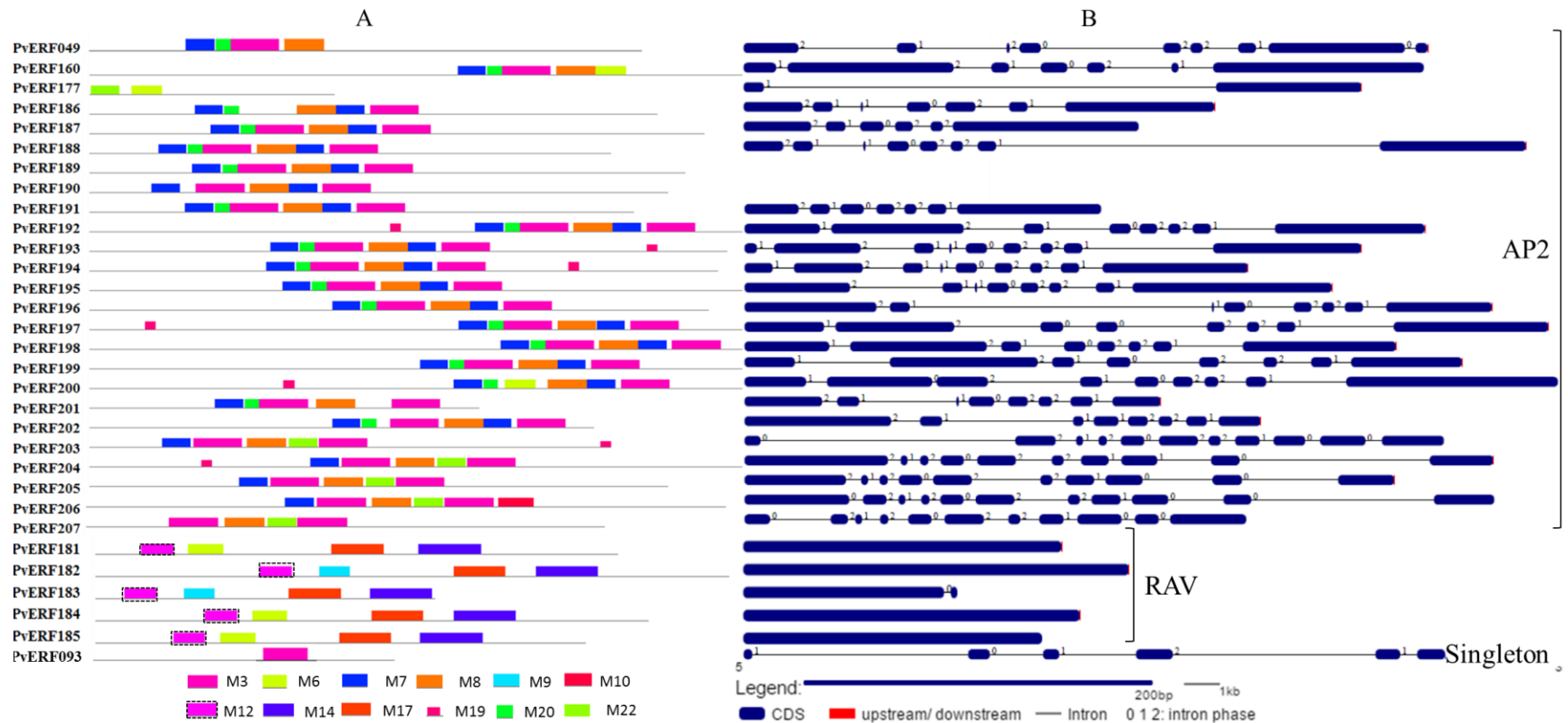


Figure 4-4 The schematic representation of protein and gene structures of switchgrass AP2 and RAV families and the singleton. A) Distribution of conserved motifs within the deduced amino acid sequences as determined by MEME tool (Bailey and Elkan, 1994). The colored boxes represent the conserved motifs. B) The gene features as visualized by the gene structure display server (Guo et al., 2007). The coding DNA sequence (CDS) and the untranslated regions (UTR) are shown by filled dark-blue and red boxes, respectively. The introns are shown by thick black lines. The splicing phases of the introns are indicated by numbers.

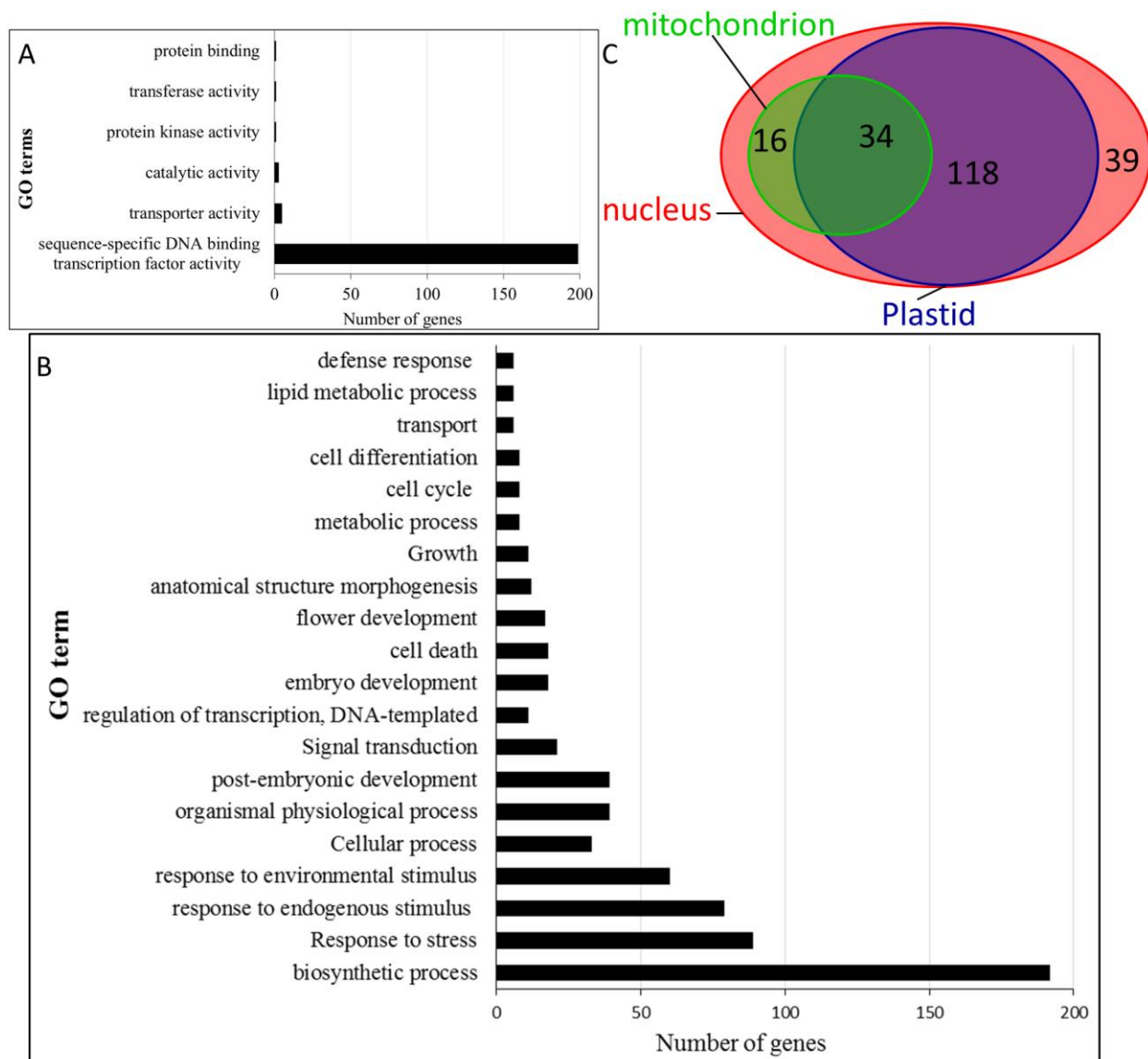


Figure 4-5 Summary of the gene ontology (GO) annotation as defined by blast2go. The switchgrass AP2/ERF genes are categorized according to biological processes (A), molecular function (B) and cellular localization (C).

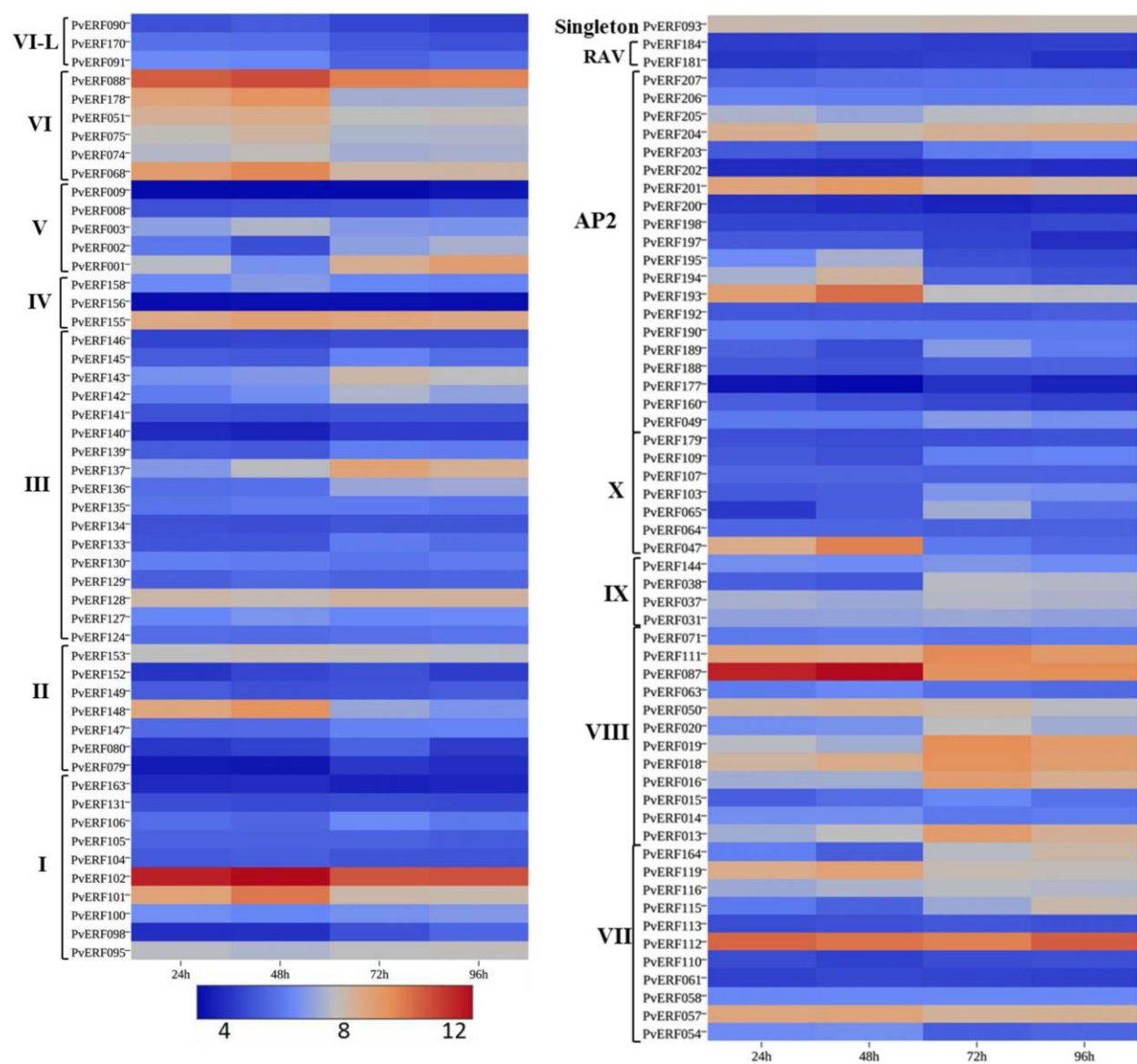


Figure 4-6 The expression pattern of putative switchgrass AP2/ERF genes at 24 h, 48 h, 72 h and 96 h after imbibition. The heat-map depicting the log2 transformed values of the expression level of each gene was obtained from the switchgrass gene expression atlas (PviGEA). The color scale represents the log2 values of gene expression with blue color denoting low expression and red for high expression. The Roman numerals I-IV represent the groups of the genes in DREB subfamily while V-X showing the groups of genes in the ERF subfamily.

Figure 4-7 The expression pattern of putative switchgrass AP2/ERF genes in roots and shoots during vegetative development. The heat-map depicting the log2 transformed values of expression level of each gene was obtained from the switchgrass gene expression atlas (PviGEA). The color scale represents the log2 values of gene expression with blue color denoting low expression and red for high expression. V1 to V5 represents the vegetative developmental stages from first fully collared leaf stage to the fifth leaf stage when the level of expression both in root and shoot samples was determined. The Roman numerals I-IV represent the groups of the genes in DREB subfamily while V-X showing the groups of genes in the ERF subfamily.

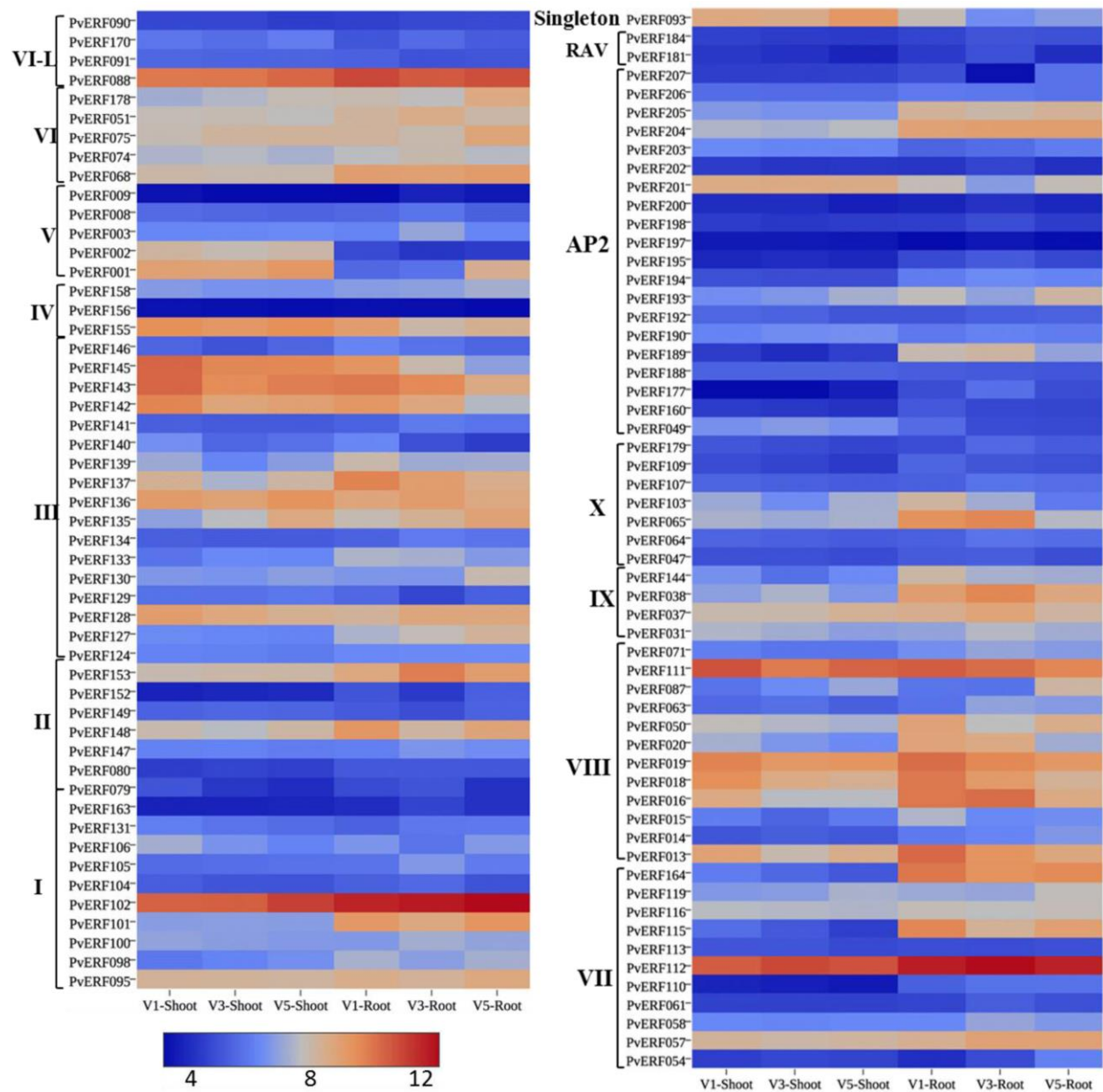


Figure 4-7 Continued.

Figure 4-8 The expression pattern of putative switchgrass AP2/ERF genes during reproductive development from floret meristem initiation to mature seed development. The heat-map depicting the log₂ transformed values of the expression level of each gene was obtained from the switchgrass gene expression atlas (PviGEA). The color scale represents the log₂ values of gene expression with blue color denoting low expression and red for high expression. The expression level was shown for developmental stages of inflorescence including meristem initiation (inflo-meristem), floret development (inflo-floret), rachis and inflorescence branch elongation (inflo-REL) and panicle emergence (inflo-PEM), and pollination and seed development from anthesis (0 days after pollination (dap)) to physiological maturity (30 dap). The Roman numerals I-IV represent the groups of the genes in DREB subfamily while V-X showing the groups of genes in the ERF subfamily.

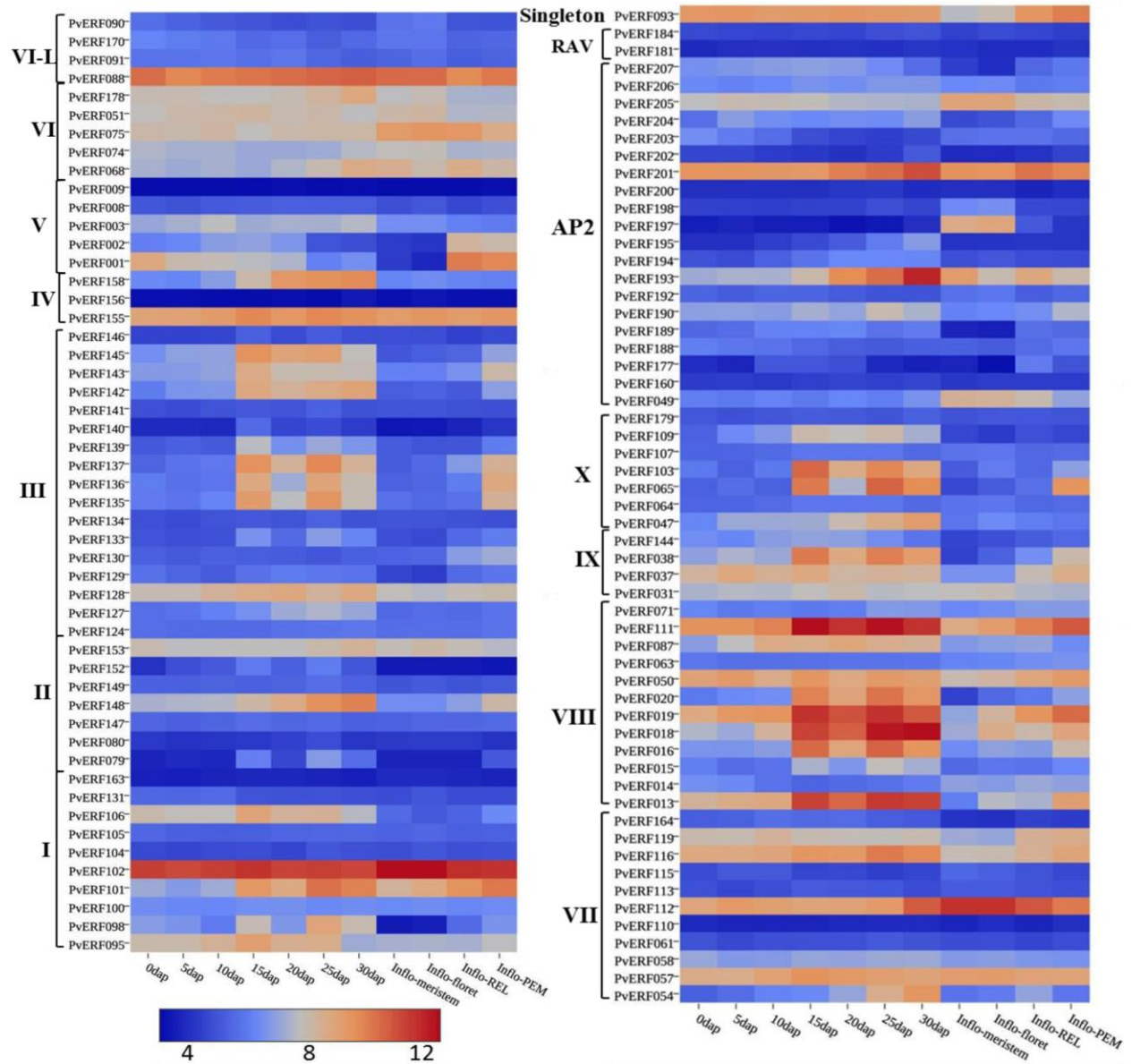


Figure 4-8 Continued.

Figure 4-9 The expression pattern of putative switchgrass AP2/ERF genes in roots and shoot parts including portions of developing internodes and vascular bundles at stem elongation stage 4 (E4). The heat-map depicting the log2 transformed values of the expression level of each gene was obtained from the switchgrass gene expression atlas (PviGEA). The color scale represents the log2 values of gene expression with blue color denoting low expression and red for high expression. The level of expression was reported for roots, nodes, leaf sheath (LSH), leaf blade (LB), whole crown (E4-crown), vascular bundle isolated from fragments of the 3rd internode (E4-I3mVB), middle fragments of the 3rd internode (E4-I3mdl) and from the bottom (E4-I4btm), middle (E4-I4mdl) and top (E4-I4top) fragments of the 4th internode. The Roman numerals I-IV represent the groups of the genes in DREB subfamily while V-X showing the groups of genes in the ERF subfamily.

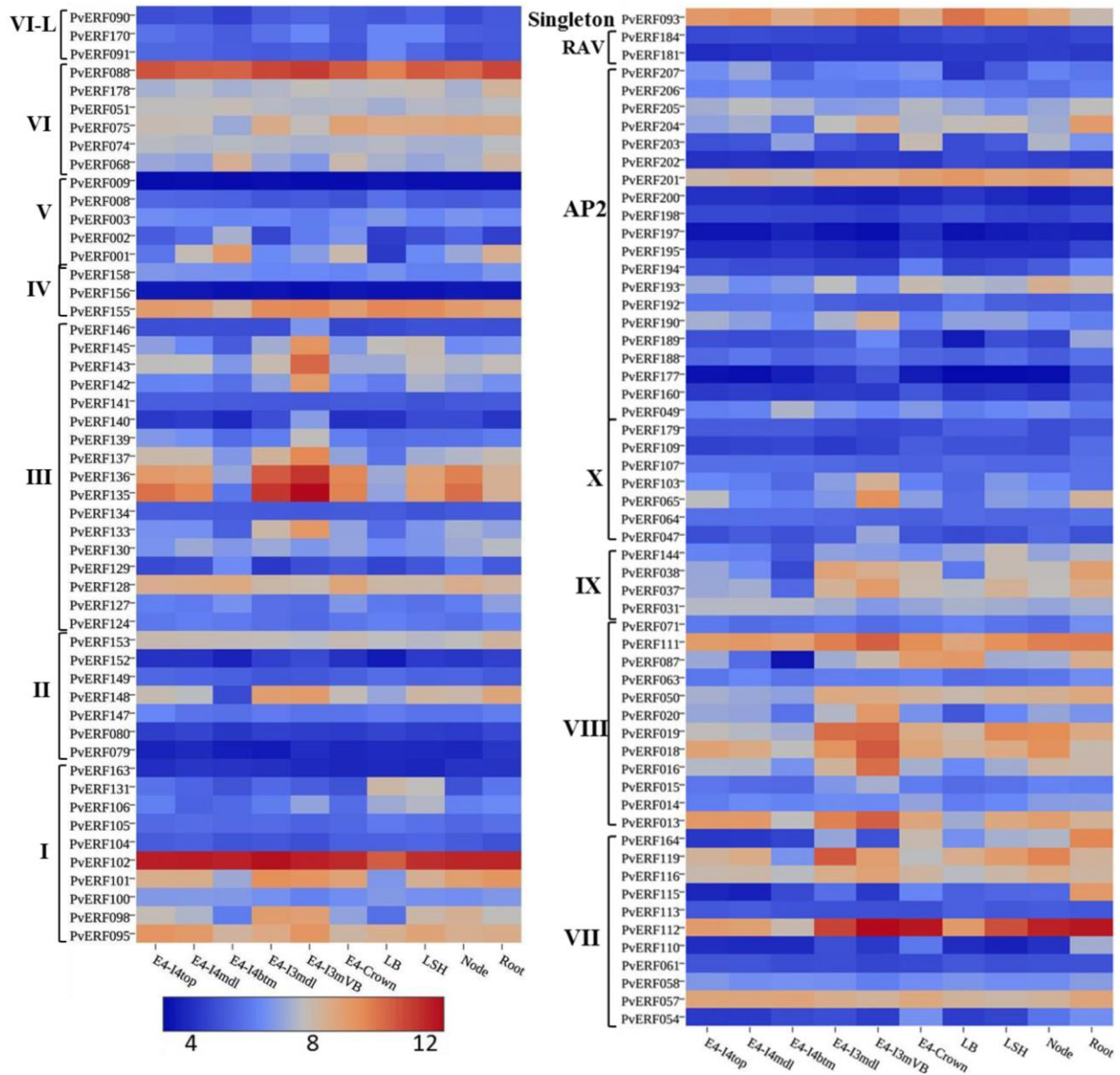


Figure 4-9 Continued.

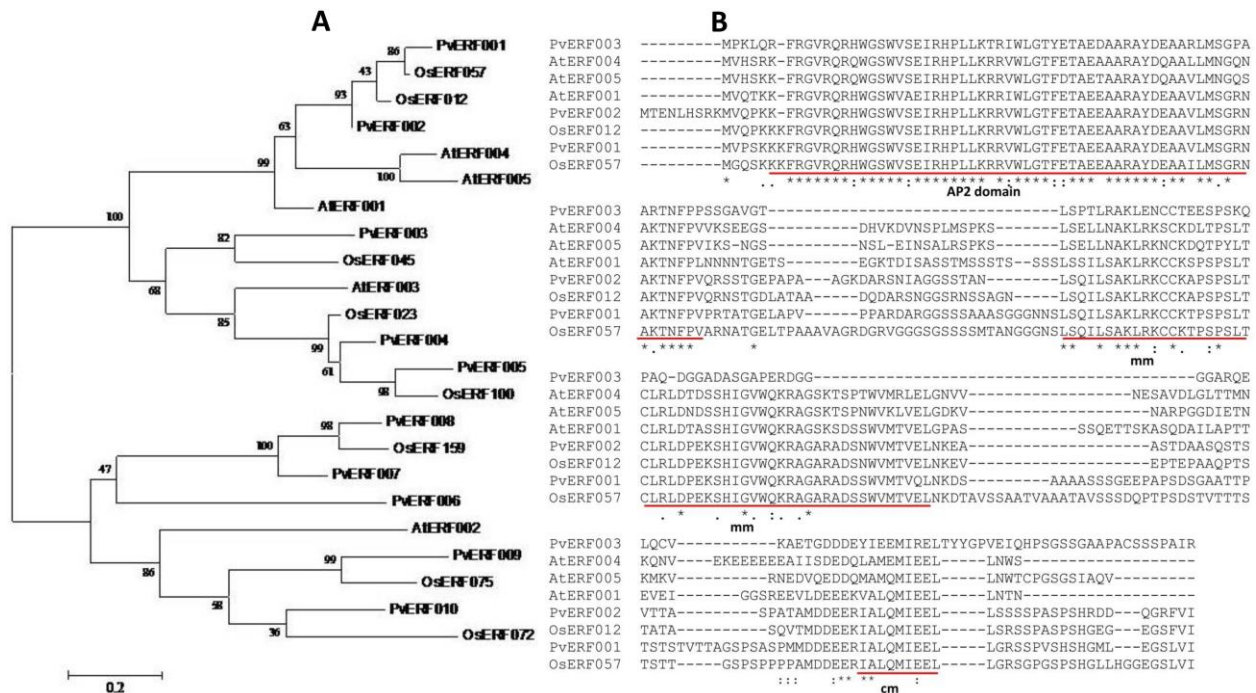


Figure 4-10 Sequence analysis of group V transcription factors in ERF subfamily. Cluster analysis of group V transcription factors in ERF subfamily using the deduced amino acid sequences of switchgrass, rice and *Arabidopsis*. The sequences were aligned using MUSCLE program and the tree was constructed by maximum likelihood method with Poisson correction and bootstrap values of 1000 in MEGA6.0 program (Tamura et al., 2013). The numbers on branches indicate bootstrap support values. The scale bar shows 0.2 amino acid substitutions per site. B) Multiple amino acid sequence alignment of switchgrass, rice and *Arabidopsis* ERF TFs. The three conserved motifs that are unique to the rice and *Arabidopsis* ERF TFs are underlined in red. The multiple sequence alignment was constructed using the amino acid sequences of respective genes by MUSCLE program (Edgar, 2004). The locus names of the switchgrass sequences and GenBank accession numbers of the sequences used in this tree are listed in Tables 4-2 & 4-8.

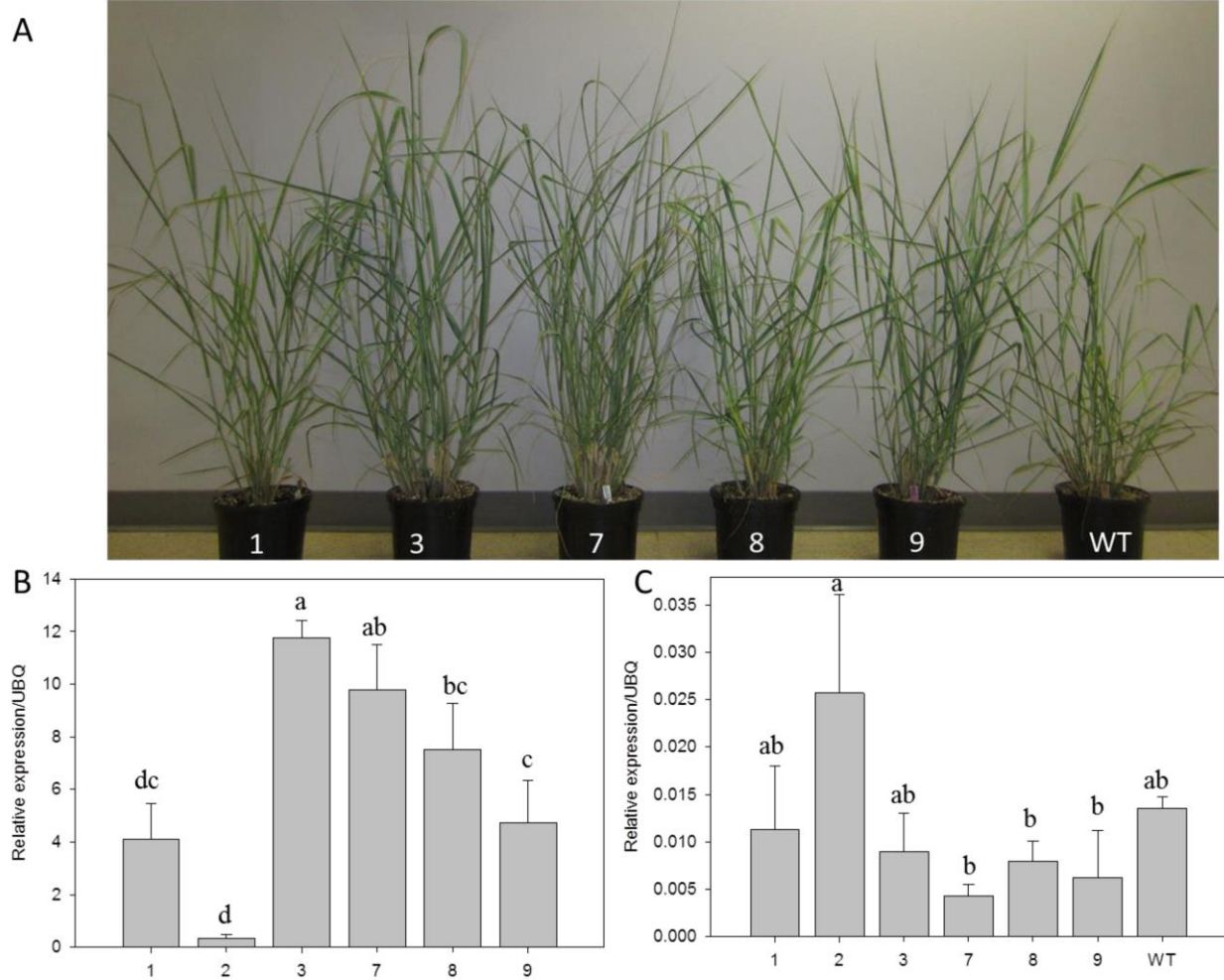


Figure 4-11 Representative *PvERF001* overexpressing and non-transgenic control (WT) switchgrass lines (A). Relative transcript levels of the transgene (B) and endogenous gene (C) in *PvERF001* overexpressing and non-transgenic (WT) plants. The expression analysis was done using RNA from the shoot tips at E4 developmental stage. The dissociation curve for the qRT-PCR products showed that the primers were gene-specific. The relative levels of transcripts were normalized to ubiquitin (UBQ). Bars represent mean values of 3 replicates $\pm$ standard error. Bars represented by different letters are significantly different at $P \leq 0.05$ as tested by LSD method with SAS software (SAS Institute Inc.).

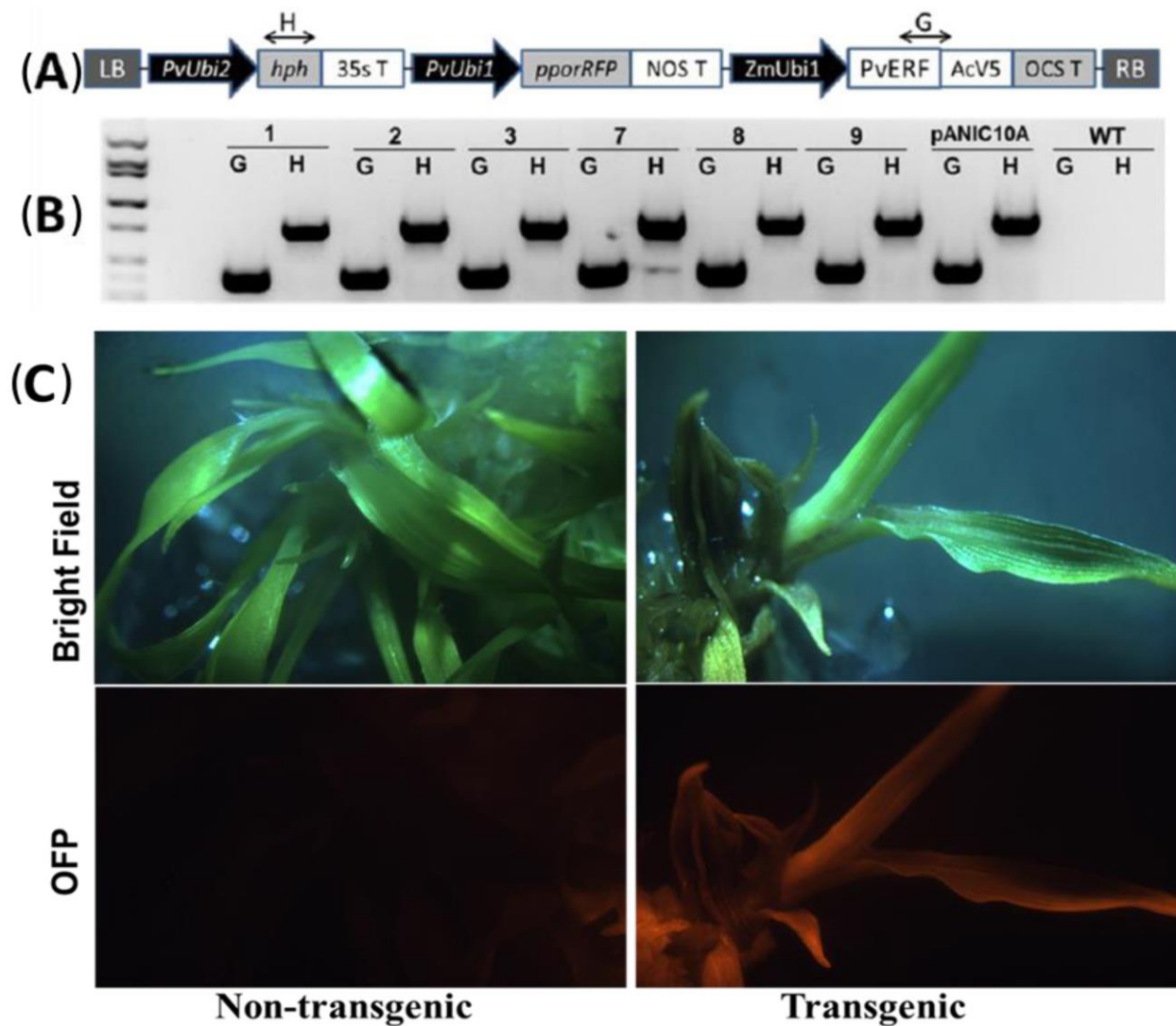


Figure 4-12 Molecular characterization of transgenic switchgrass plants overexpressing the *PvERF001* gene. A) pANIC10A vector construct used for overexpression of *PvERF001*. B) Genomic PCR confirming the insertion of the transgene (G) and the hygromycin-resistance (H) genes in transgenic lines. No amplification was observed in DNA samples from the non-transgenic (WT) plants. C) Orange fluorescence protein (*pporRFP*; OFP) visualization in transgenic plants compared to the non-transgenic control.

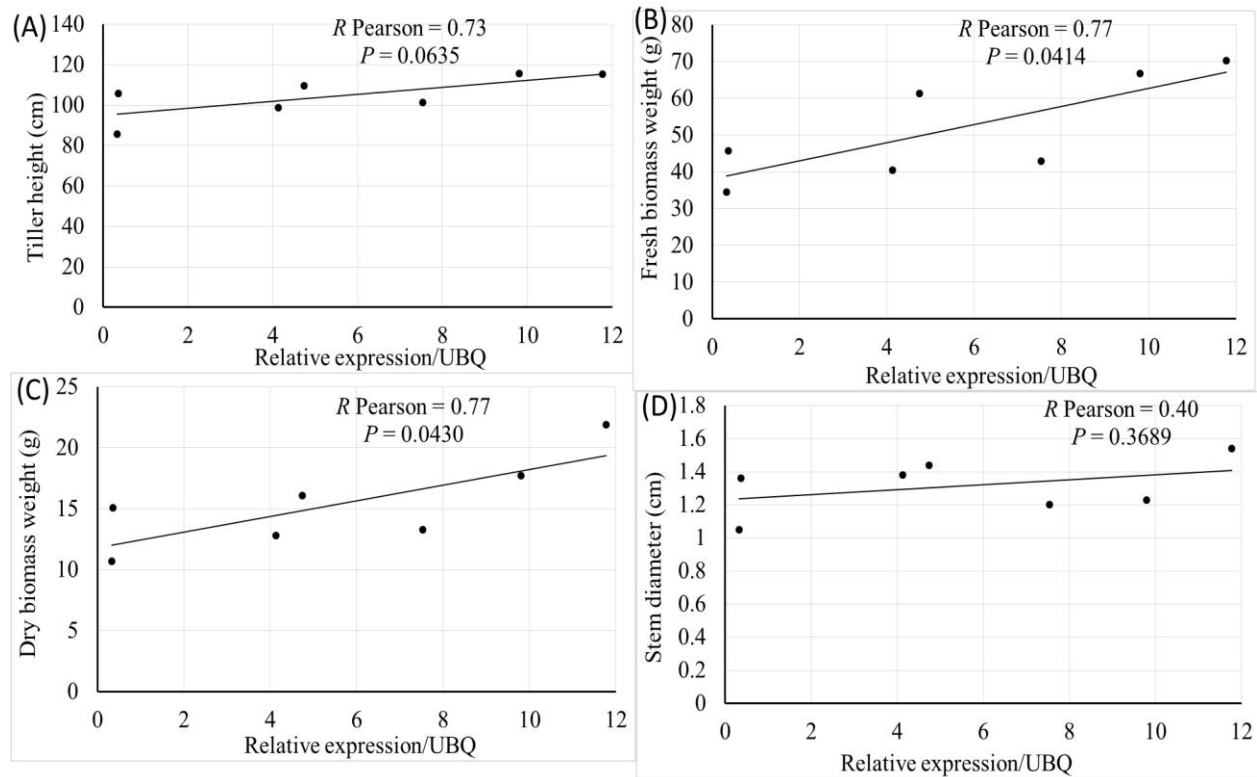


Figure 4-13 Correlation between relative transcript levels of *PvERF001* and growth metrics: tiller height (A), fresh biomass weight (B), dry biomass weight (C) and stem diameter (D) in switchgrass lines.

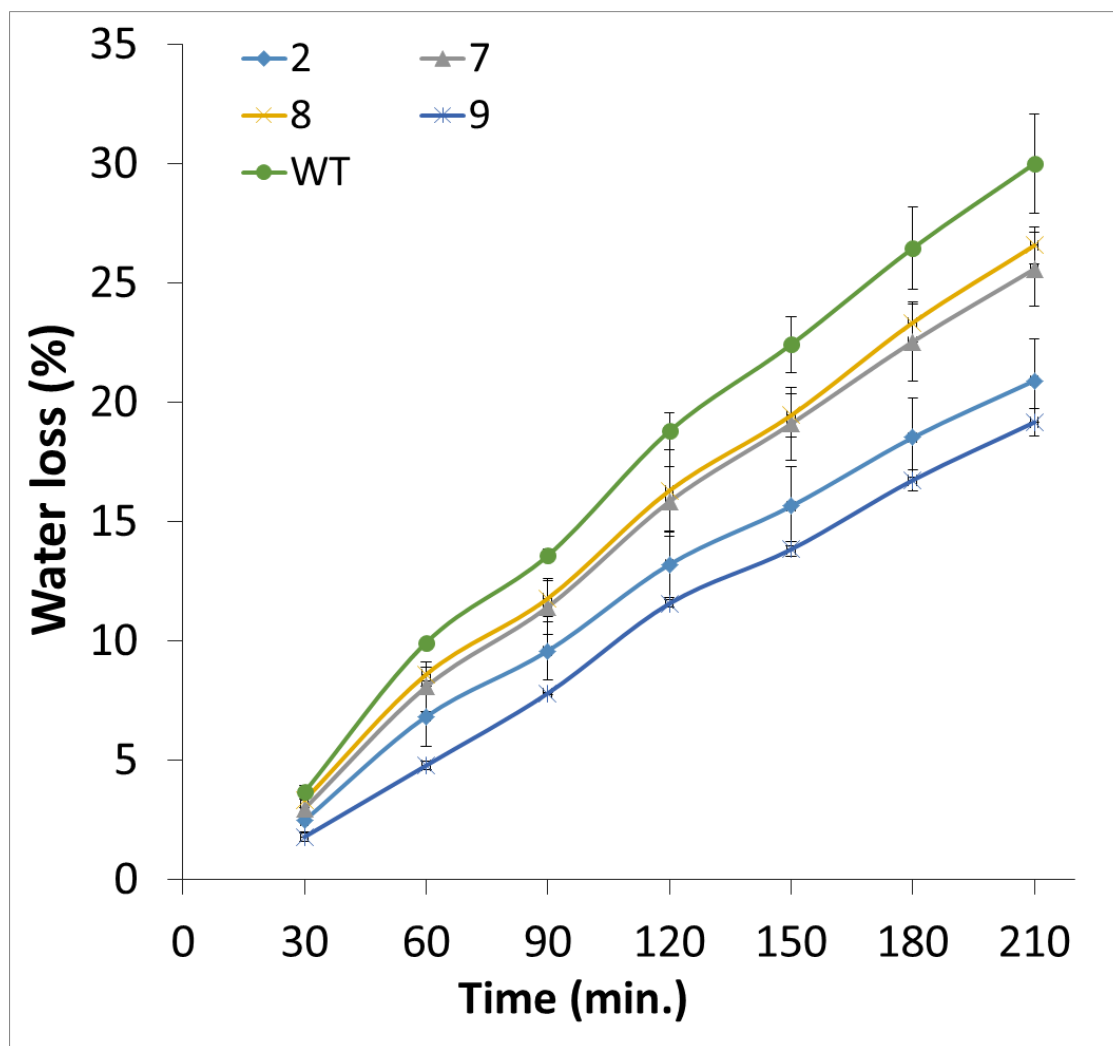


Figure 4-14 Comparison of the relative rate of water loss between the transgenic switchgrass overexpressing *PvERF001* and non-transgenic control lines. The data points indicated are means of three replicates. Error bars represent mean $\pm$ SD.

Figure 4-15 The relative expression of putative target genes of *PvERF001* in transgenic vs the non-transgenic (WT) plants as determined by qRT-PCR. A) Wax biosynthetic genes B) Cutin biosynthetic genes. The relative levels of transcripts were normalized to ubiquitin (UBQ). Bars represent mean values of 3 replicates $\pm$ standard error. Asterisks indicate significant differences from non-transgenic control plants at $P \leq 0.05$ as determined by PROC TTEST procedure using SAS software (SAS Institute Inc.). (*Pv*)*LACS1* (Long-chain acyl-CoA synthase 1); (*Pv*)*CER6* (Eceriferum 6); (*Pv*)*FAE1* (fatty acid elongase1); (*Pv*)*KCS1* (3-ketoacyl-CoA synthase 1); (*Pv*)*FDH2* (Formate dehydrogenase 2); (*Pv*)*CYP86A7-1* (Cytochrome P450); (*Pv*)*HTH* (HOTHEAD protein).

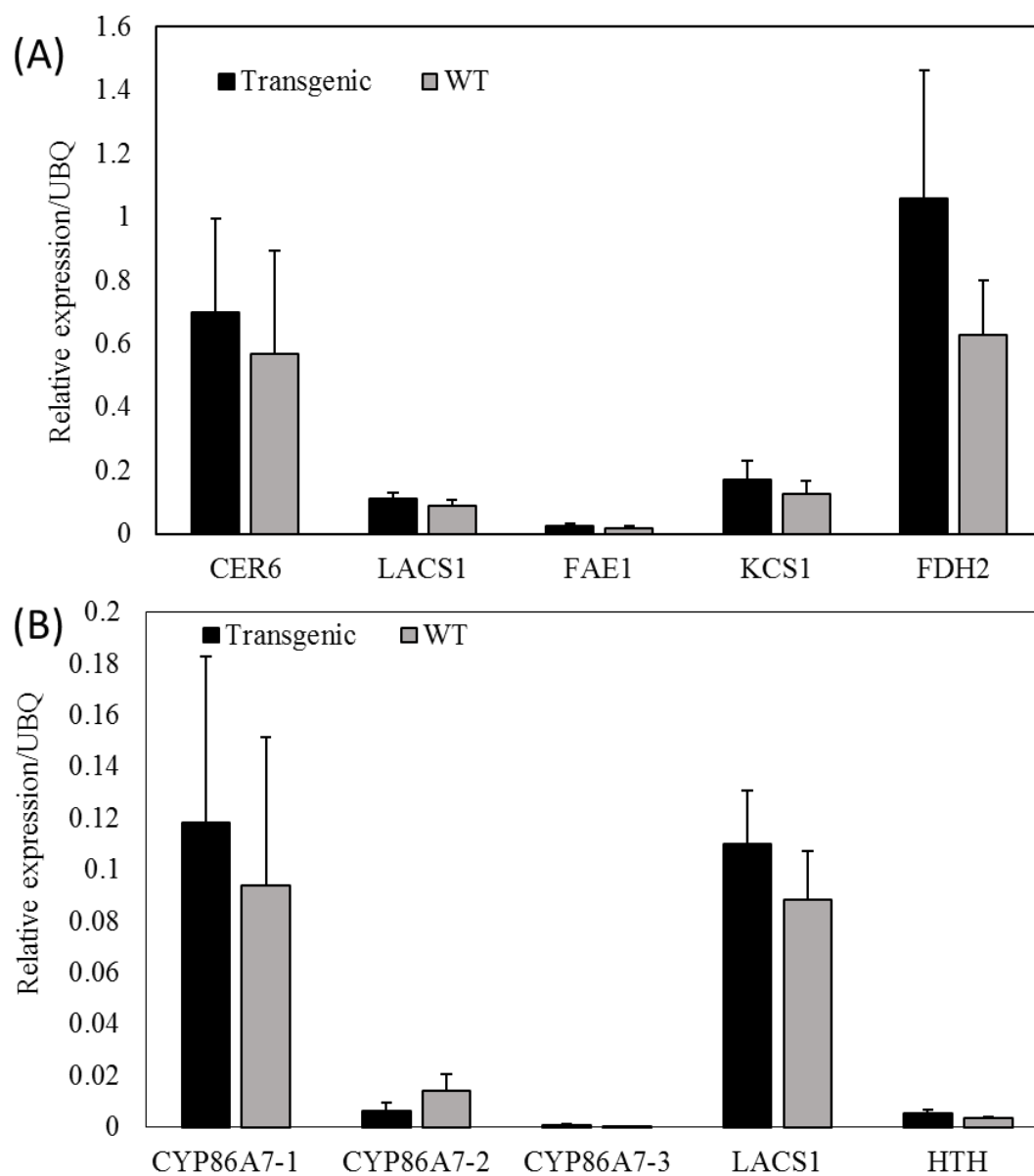


Figure 4-15 Continued.

Figure 4-16 The relative expression of putative target genes of *PvERF001* in transgenic vs the non-transgenic (WT) plants as determined by qRT-PCR. A) Lignin biosynthetic genes B) Cellulose and hemicellulose biosynthetic genes. C) Transcriptional regulators of cell wall biosynthetic genes. The relative levels of transcripts were normalized to ubiquitin (UBQ). Bars represent mean values of 3 replicates $\pm$ standard error. Asterisks indicate significant differences from non-transgenic control plants at $P \leq 0.05$ as determined by PROC TTEST procedure using SAS software (SAS Institute Inc.). (*Pv*)*4CL* (4-coumarate: CoA ligase); (*Pv*)*C3H* (coumaroyl shikimate 3-hydroxylase); (*Pv*)*C4H* (coumaroyl shikimate 4-hydroxylase); (*Pv*)*CAD* (cinnamyl alcohol dehydrogenase); (*Pv*)*CCR* (cinnamoyl CoA reductase); (*Pv*)*COMT* (caffeic acid 3-O-methyltransferase); (*Pv*)*F5H* (ferulate 5-hydroxylase); (*Pv*)*PAL* (phenylalanine ammonia-lyase); (*Pv*)*CESA* (cellulose synthase); (*Pv*)*CSL* (cellulose synthase-like). The cellulose and hemicellulose biosynthetic genes as well as the cell wall related TFs were labeled according to the naming from the closest rice or *Arabidopsis* homologs used in Ambavaram et al. (2011).

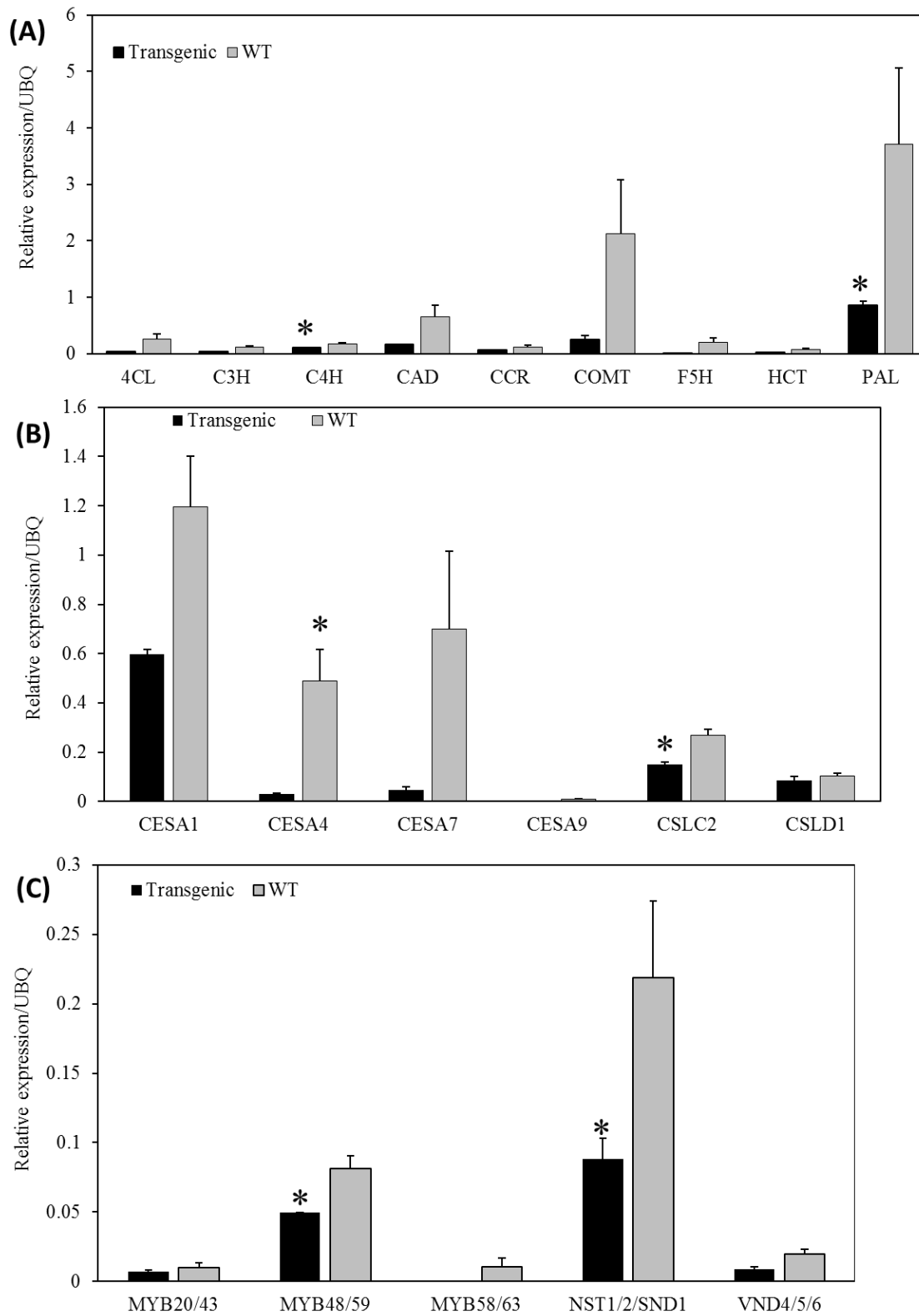


Figure 4-16 Continued.

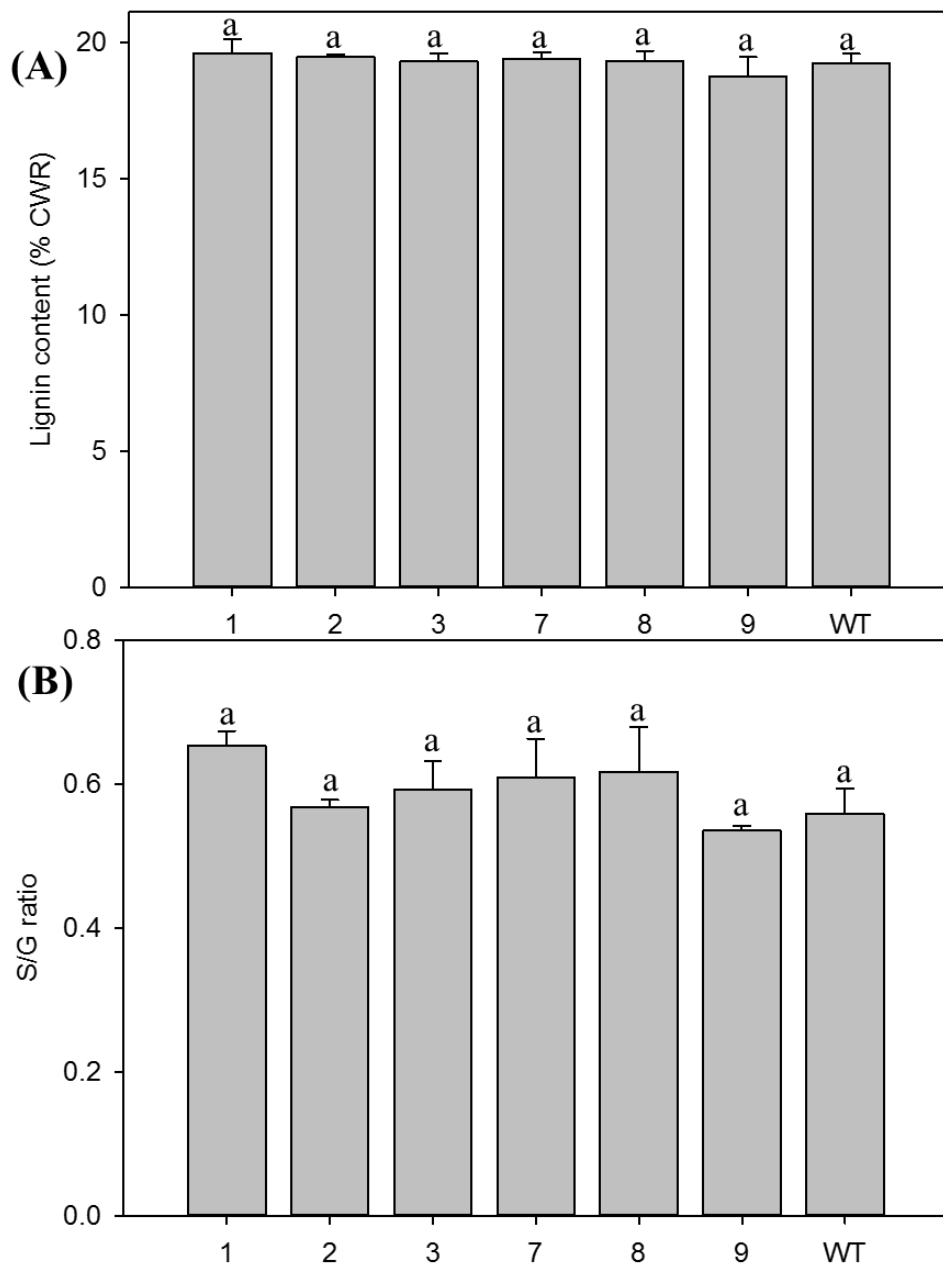


Figure 4-17 Lignin content (A) and S/G ratio (B) of transgenic plants overexpressing *PvERF001* and non-transgenic (WT) switchgrass lines as determined via py-MBMS. Bars represent the average of the replicates $\pm$ standard error. Bars represented by different letters are significantly different at $P \leq 0.05$ as tested by LSD method with SAS software (SAS Institute Inc.). CWR, cell wall residues.

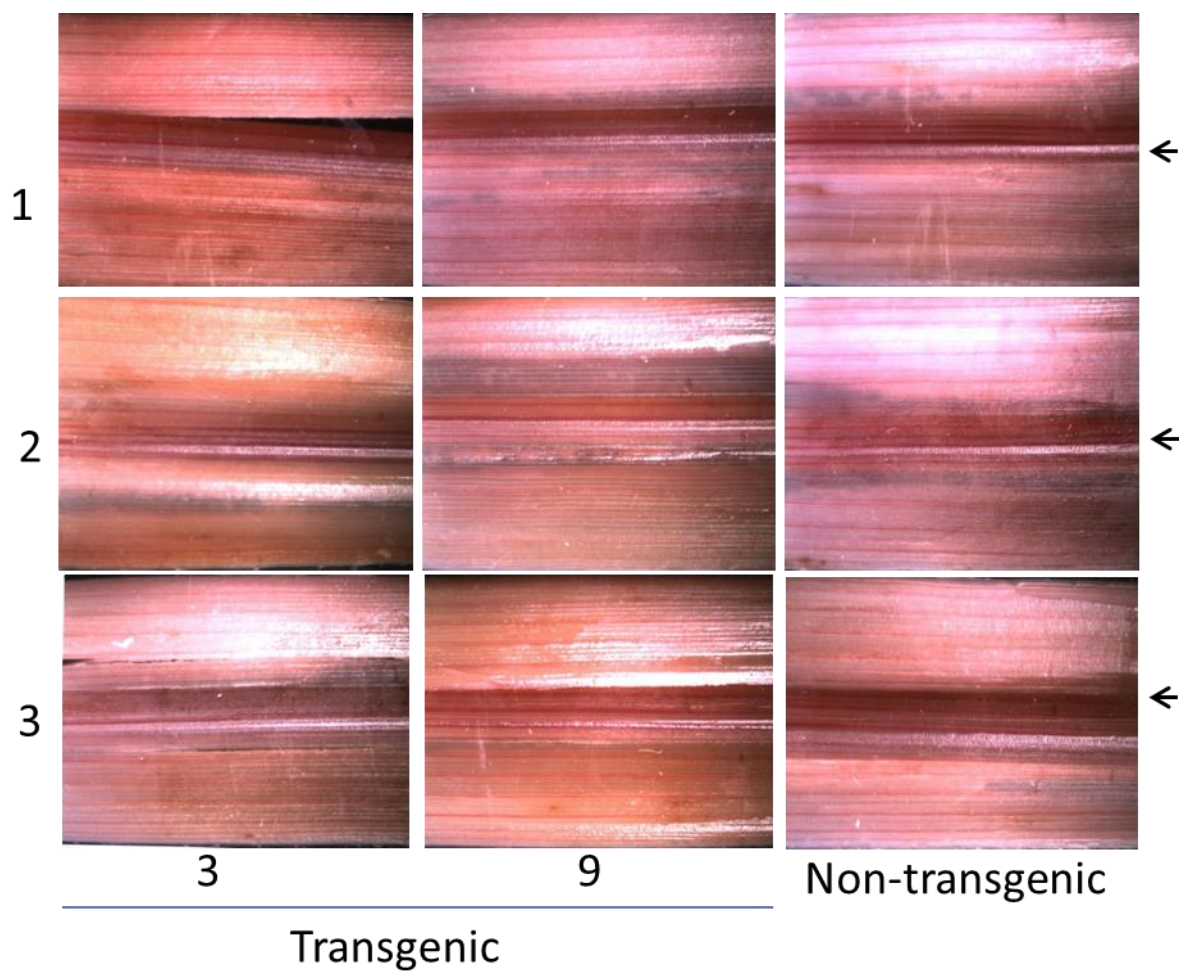


Figure 4-18 Histochemical detection of lignin in leaves of *PvERF001* overexpressing lines compared to non-transgenic control. The phloroglucinol-HCl staining was done on leaves from three independent tillers and the pictures were taken at 2x magnification.

Table 4-1 List of primers used in this study.

Primer name	Unitranscript ID	Primer sequence	Reference
Primers for Cloning <i>PvERF001</i>			
PvSHN2F	Pavir.Da00422	ATGGTGCCGTCGAAGAAGAAGTTC	
PvSHN2R		GATGACGAGGCTGCCTTCCAGC	
Primers for genomic DNA PCR			
Hygro_F		TTGCATCTCCCGCCGTTACAG	
Hygro_R		CTGGGGCGTCGGTTTCCACTAT	
837_Fnew		ATGACCGTCCAGCTCAACAAGGA	
837_R/common		ACAGCGACTTCCTGACCATCCT	
Gene-specific primers for qRT-PCR			
Primers for confirming the overexpression of <i>PvERF001</i>			
F_837		TCTCGCACTCACATGGGATGCTG	
R_AcV5		ACCAGCCGCTCGCATCTTTC	
Gene specific primers for endogenous <i>PvERF001</i> gene expression:			
PvSHN2_F1	Pavir.Da00422	ACTCACATGGGATGCTGGAAGG	
PvSHN2_R1		GTGCAACAACCACGCTCTACCA	
Primers for reference gene			
F_PvUBIQUITIN		CAGCGAGGGCTCAATAATTCCA	
R_PvUBIQUITIN	AP13CTG25905	TCTGGCGGACTACAATATCCA	Xu et al., 2011
Primers for the expression of lignin biosynthesis genes			
C4H1_1534F	AP13CTG28733	GGGCAGTTCAGCAACCAGAT	Shen et al., 2012
C4H1_1611R		CGCGTTTCCGGGACTCTAG	
PvCOMT_F461	KanlCTG02872	CAACCGCGTGTTCAACGA	Shen et al., 2012
PvCOMT_R534		CGGTGTAGAACTCGAGCAGCTT	
4CL1_1179_F	AP13CTG06049	CGAGCAGATCATGAAAGGTTACC	Shen et al., 2012
4CL1_1251_R		CAGCCAGCCGTCCTTGTC	

Table 4-1 Continued.

Primer name	Unitranscript ID	Primer sequence	Reference
PvCCR1. 112_F	AP13ISTG52570	GCGTCGTGGCTCGTCAA	Shen et al., 2012
PvCCR1. 187_R		TCGGGTCATCTGGGTTCCT	
PvCAD_F116	KanlCTG19538	TCACATCAAGCATCCACCATCT	Shen et al., 2012
PvCAD_R184		GTTCTCGTGTCCGAGGTGTGT	
PAL_F1	KanlCTG00004	CATATAGTGTGCGTGCGTGTGT	
PAL_R1		CTGGCCCGCCAATCG	
C3H_F1	AP13ISTG41630	CGTGAACAATGGGATCAGGATAG	
C3H_R1		GCGGACACAACCATCTCAAATAC	
F5H_F1	AP13ISTG56842	CCCCGTGCACTGACGATCTAT	
F5H_R1		CCAAGCCAAGGGAAAACACAGTTA	
Laccase_F2	AP13CTG11594	AGCATCCTGGGCATCGAGAG	
Laccase_R2		GTCGTAGTTGCCGTACCCTTG	
Primers for cellulose and hemicellulose synthetic genes			
CESA1_F1	AP13CTG06092	GCATCCAGGGTCCAGTTTATGTG	
CESA1_R1		CCAGATCGGCTTCGGTCAATAC	
CESA3_F1	AP13CTG00607	GCATTTCCTCCTCGTCGTC	
CESA3_R1		ACTTGCTCCTCTCGCTCTAGC	
CESA4_F1	AP13CTG01684	GAATGCTCTGGTCCGAGTGTC	
CESA4_R1		TCTGCCAACAGTAGGGTCCATC	
CESA7_F1	AP13CTG19403	CTTCTCGCTCGTCTGGGTTAG	
CESA7_R1		AGCTCGATCAATTCAGCACTCG	
CESA8_F1	AP13CTG00048_1	CGTGAGAAGCGTCCTGGATTC	
CESA8_R1		CCTGAGAGCCTTGCTGTTGTTG	
CSLA6_F1	AP13CTG14018	TGCACATTACGGAGCTTGGTG	
CSLA6_R1		TGGTCCCTCCCATACGCAAG	

Table 4-1 Continued.

Primer name	Unitranscript ID	Primer sequence	Reference
CSLC2_F1	AP13CTG06284	GTAGAAGCAGCCAAGGCACTG	
CSLC2_R1		GAGTCGCTGTACGCCTCTTG	
CSLD1_F1	AP13CTG08033	CACAGCTACCACGTCCACATC	
CSLD1_R1		CGATGACCTTGTCCATGAGGTG	
Primers for the expression of wax and cutin biosynthesis genes			
CER6_F1	AP13ISTG69567	ACCTCGTCCACATCCTCTGCTC	
CER6_R1		CTTGTAGCAGGCGTAGTCCACCA	
CYP86A7-1_F1	AP13CTG23248	ACTCGTACAGGTTCTGTCGCTTC	
CYP86A7-1_R1		GGTGAGCGACATCTTCTGCTCCA	
FAE1_F1	AP13CTG28660	CTCAACCTCGTCTCCGTGCT	
FAE1_R1		GGAGCATCGCATGAAGGTGTC	
FDH2_F1	AP13CTG03242	GTCTCCGTGGCAGAAGATGAGC	
FDH2_R1		CTGCGACGTTCCACTCTCCTTG	
HTH_F1	Pavir.Ib02493.1	GTGATCGACAGCTCCACCTTC	
HTH_R1		TCCTCCATCTCTCTGCCTGGA	
KCS1_F2	AP13ISTG42547	GACGCTCCTCCCTCCAATACAG	
KCS1_R2		AGCAGCCAGCCTCCATTGTGT	
LACS1_F1	AP13CTG04675	GGAGTGCAAGTCGAGGTTGGTG	
LACS1_R1		ACACCCCTTCCCAGTGTGTCTC	

Table 4-2 Characteristic features of switchgrass AP2/ERF transcription factor gene family members identified in *Panicum virgatum*.

Gene name	Phytozome identifier or unitranscript ID	Sub-family	Chromosome	group	Amino acid sequence length	NCBI BLASTP annotation	Minimum eValue	Mean Similarity (%)
PvERF001	Pavir.Da00422	ERF	4	V	230	ethylene response factor 1	1.76E-102	74.60
PvERF002	Pavir.Aa02977	ERF	1	V	213	ethylene response factor 1	3.28E-118	76.45
PvERF003	Pavir.Ga00062	ERF	7	V	166	ap2-domain dre binding factor dbf1	1.41E-51	71.75
PvERF004	Pavir.Aa00196	ERF	1	V	148	ethylene response factor 1	9.73E-69	73.35
PvERF005	Pavir.Da02059	ERF	4	V	184	ethylene response factor 1	2.88E-68	73.90
PvERF006	Pavir.Ba01277	ERF	2	V	227	dre binding factor 2	7.09E-50	78.35
PvERF007	Pavir.Ia01300	ERF	9	V	403	pti6 -like	1.57E-69	68.90
PvERF008	Pavir.Cb00448	ERF	3	V	394	ap2 domain containing protein	5.26E-94	74.85
PvERF009	Pavir.Ba00533	ERF	2	V	315	ap2 domain containing protein	6.99E-120	69.55
PvERF010	Pavir.Ba03636	ERF	2	V	295	ap2 domain containing protein	2.24E-80	74.40
PvERF011	AP13CTG22194	ERF		VIII	198	ethylene responsive element binding factor	1.25E-52	71.20
PvERF012	Pavir.Db00303	ERF	4	VIII	199	ethylene responsive element binding factor	1.72E-78	74.25
PvERF013	Pavir.Ga00187	ERF	7	VIII	192	ethylene responsive element binding factor	2.86E-52	75.85
PvERF014	Pavir.Aa03102	ERF	1	VIII	196	ethylene responsive element binding factor	2.13E-44	64.60
PvERF015	Pavir.Ga00485	ERF	7	VIII	320	pti6 -like	3.12E-65	73.55
PvERF016	Pavir.Fa01728	ERF	6	VIII	173	erebp-like protein	1.78E-25	72.45
PvERF017	Pavir.Fa01794	ERF	6	VIII	240	pti6 -like	1.06E-27	75.15
PvERF018	Pavir.Fa01730	ERF	6	VIII	193	erebp-like protein	5.83E-27	65.85
PvERF019	Pavir.Fb00951	ERF	6	VIII	216	pti6 -like	4.11E-36	70.90
PvERF020	AP13CTG21592	ERF		VIII	219	pti6 -like	7.45E-31	74.40
PvERF021	Pavir.Bb00653	ERF	2	IX	246	ethylene response factor erf1	6.32E-51	68.45
PvERF022	Pavir.Ba03491	ERF	2	IX	323	ap2 domain containing expressed	2.08E-52	71.65
PvERF023	Pavir.Ib03413	ERF	9	IX	256	ap2 domain containing expressed	1.06E-55	68.05
PvERF024	Pavir.Ia02224	ERF	9	IX	222	ap2 domain containing expressed	2.50E-42	83.95
PvERF025	Pavir.Ba03296	ERF	2	VI	243	pti6 -like	1.77E-22	67.05
PvERF026	Pavir.Ga01098	ERF	7	IX	322	ethylene responsive element binding factor 5	6.56E-73	75.10

Table 4-2 Continued.

Gene name	Phytozome identifier or unitranscript ID	Sub-family	Chromosome	group	Amino acid sequence length	NCBI BLASTP annotation	Minimum eValue	Mean Similarity (%)
PvERF027	Pavir.Ga01097	ERF	7	IX	348	ethylene responsive element binding factor 5	2.86E-73	74.25
PvERF028	Pavir.J23730	ERF	7	IX	230	ethylene responsive element binding factor 5	1.09E-52	76.35
PvERF029	Pavir.Gb01173	ERF	7	IX	332	ethylene responsive element binding factor 5	9.83E-87	69.30
PvERF030	Pavir.Aa01181	ERF	1	IX	326	ethylene responsive element binding factor 5	8.86E-94	73.75
PvERF031	Pavir.Ia00066	ERF	9	IX	226	ap2 domain containing expressed	5.69E-52	79.25
PvERF032	Pavir.Ia01823	ERF	9	IX	119	ap2 domain containing expressed	2.51E-42	80.90
PvERF033	Pavir.Fa00942	ERF	6	IX	133	ap2-related transcription factor	1.21E-14	70.38
PvERF034	Pavir.Da00220	ERF	4	VI	155	pti6 -like	1.33E-26	69.20
PvERF035	Pavir.Fa00010	ERF	6	IX	126	pathogenesis-related genes transcriptional activator pti5-like	1.16E-42	78.45
PvERF036	Pavir.Bb01308	ERF	2	IX	134	ap2 domain containing expressed	8.86E-48	79.85
PvERF037	Pavir.Ga01099	ERF	7	IX	298	ap2-related transcription factor	5.25E-112	77.25
PvERF038	Pavir.Aa01184	ERF	1	IX	299	ap2-related transcription factor	1.93E-108	78.00
PvERF039	Pavir.Ba02374	ERF	2	IX	258	ethylene response factor erf1	3.76E-46	68.05
PvERF040	Pavir.Ba02781	ERF	2	IX	325	ethylene response factor erf1	3.16E-45	70.65
PvERF041	Pavir.Ga02734	ERF	7	IX	172	ethylene response factor erf1	2.21E-26	55.80
PvERF042	Pavir.Ha01264	ERF	8	IX	312	ethylene response factor erf1	2.53E-40	71.55
PvERF043	Pavir.Ha01200	ERF	8	IX	275	ap2 domain containing expressed	4.03E-41	67.75
PvERF044	Pavir.Ha00307	ERF	8	IX	208	ap2 domain containing expressed	9.93E-39	64.95
PvERF045	Pavir.Ha00304	ERF	8	IX	192	ap2 domain containing expressed	6.52E-35	66.60
PvERF046	Pavir.J02402	ERF		X	330	ap2 domain containing expressed	1.07E-47	65.50
PvERF047	Pavir.Ea03713	ERF	5	X	355	ap2 domain containing expressed	9.20E-76	64.00
PvERF048	Pavir.Ca01358	ERF	3	VIII	233	ethylene-responsive element binding factor	3.48E-72	75.10
PvERF049	Pavir.Ca01410	AP2	3	AP2	410	ap2 dna-binding domain	0	73.85
PvERF050	Pavir.Ea03147	ERF	5	VIII	234	erebp transcription factor	1.86E-80	78.85

Table 4-2 Continued.

Gene name	Phytozome identifier or unitranscript ID	Sub-family	Chromosome	group	Amino acid sequence length	NCBI BLASTP annotation	Minimum eValue	Mean Similarity (%)
PvERF051	Pavir.Ea00119	ERF	5	VI	369	pti6 -like	1.16E-126	64.65
PvERF052	Pavir.Db02131	ERF	4	VI	323	erebp transcription factor	9.30E-37	65.80
PvERF053	Pavir.Ba01557	ERF	2	VII	264	erebp transcription factor	3.61E-25	72.45
PvERF054	Pavir.Cb01365	ERF	3	VII	192	erebp transcription factor	2.42E-53	68.05
PvERF055	Pavir.Ha01720	ERF	8	X	296	ap2 domain containing expressed	8.57E-51	69.55
PvERF056	Pavir.Cb02005	ERF	3	X	420	ap2 domain containing expressed	5.11E-56	73.35
PvERF057	Pavir.Eb02324	ERF	5	VII	213	erebp transcription factor	3.91E-66	71.00
PvERF058	Pavir.Ba00136	ERF	2	VII	207	erebp transcription factor	7.74E-64	73.80
PvERF059	Pavir.Aa01729	ERF	1	VII	264	erebp-like protein	3.80E-26	74.85
PvERF060	Pavir.Aa01785	ERF	1	VII	294	ap2 domain containing expressed	1.58E-27	76.35
PvERF061	Pavir.J15316	ERF		VII	255	erebp-like protein	1.61E-35	72.20
PvERF062	Pavir.J36215	ERF		VIII	211	pti6 -like	5.64E-43	75.60
PvERF063	Pavir.Eb03165	ERF	5	VIII	260	erebp-like protein	1.65E-41	72.70
PvERF064	Pavir.J34881	ERF		X	318	ap2-related transcription factor	2.35E-50	79.85
PvERF065	Pavir.Fa00656	ERF	6	X	272	erebp-like protein	2.29E-99	74.30
PvERF066	Pavir.J24236	ERF		VI	242	erebp transcription factor	1.78E-24	65.25
PvERF067	Pavir.J22854	ERF		IX	126	ap2 domain containing expressed	3.56E-47	82.90
PvERF068	Pavir.Ea02288	ERF	5	VI	304	pti6 -like	7.74E-131	71.75
PvERF069	AP131STG74146	ERF		X	263	erebp transcription factor	1.83E-90	67.40
PvERF070	Pavir.Aa01903	ERF	1	VIII	250	frizzy panicle	1.45E-74	76.35
PvERF071	Pavir.Gb01913	ERF	7	VIII	271	frizzy panicle	4.78E-89	77.75
PvERF072	Pavir.Fb00452	ERF	6	VIII	182	frizzy panicle	6.77E-36	73.50
PvERF073	Pavir.Ba00192	ERF	2	VIII	302	frizzy panicle	1.48E-113	79.95
PvERF074	Pavir.Ia04256	ERF	9	VI	293	pti6 -like	5.94E-77	63.90

Table 4-2 Continued.

Gene name	Phytozome identifier or unitranscript ID	Sub-family	Chromosome	group	Amino acid sequence length	NCBI BLASTP annotation	Minimum eValue	Mean Similarity (%)
PvERF075	Pavir.Ia00165	ERF	9	VI	245	frizzy panicle	9.55E-58	65.05
PvERF076	Pavir.Aa01483	ERF	1	VIII	283	frizzy panicle	2.72E-30	75.50
PvERF077	Pavir.J21660	ERF		VIII	301	frizzy panicle	3.69E-44	79.55
PvERF078	Pavir.Da01882	DRE B	4	II	195	ap2 domain transcription factor-like	1.45E-48	67.75
PvERF079	Pavir.Aa00296	DRE B	1	II	216	ap2 domain transcription factor-like	4.04E-77	68.55
PvERF080	Pavir.Fa00741	DRE B	6	II	231	ap2 domain transcription factor-like	3.89E-69	71.15
PvERF081	Pavir.Ia04341	DRE B	9	IV	304	ap2 domain containing protein	2.49E-125	78.10
PvERF082	Pavir.Ca00939	ERF	3	IX	266	ap2-related transcription factor	1.17E-72	73.55
PvERF083	Pavir.Fa00470	ERF	6	XI	211	ap2 domain transcription factor-like	5.44E-14	55.95
PvERF084	Pavir.Fa00469	ERF	6	XI	185	ap2 domain transcription factor-like	9.79E-15	58.25
PvERF085	Pavir.Fb01907	ERF	6	XI	191	ap2 domain transcription factor-like	1.42E-13	53.20
PvERF086	Pavir.Aa02576	DRE B	1	III	249	dre binding factor 2	1.89E-51	70.10
PvERF087	Pavir.Ia04390	ERF	9	VIII	286	erebp transcription factor	3.21E-40	68.70
PvERF088	AP13CTG15320	ERF	5	VI-L	327	ap2 domain containing protein	4.48E-93	55.21
PvERF089	Pavir.Fa01379	ERF	6	VI-L	286	hypothetical protein	2.57E-17	80.50
PvERF090	Pavir.Ba03000	ERF	2	VI-L	290	ap2 domain-containing transcription factor-like protein	6.20E-68	56.00
PvERF091	Pavir.Ba02999	ERF	2	VI-L	279	erebp transcription factor	8.50E-100	63.70
PvERF092	Pavir.Ca01923	ERF	3	VIII	261	ap2 domain containing protein	3.11E-17	56.75
PvERF093	Pavir.Aa01947	Singleton	1	Singleton	222	unnamed protein product	2.57E-116	90.67
PvERF094	Pavir.Ia02531	DRE B	9	X	248	ap2 domain containing protein	1.09E-53	64.00
PvERF095	Pavir.Aa00501	DRE B	1	I	328	ap2-domain dre binding factor dbf1	1.67E-126	69.65
PvERF096	Pavir.Da01302	DRE B	4	I	344	ap2-domain dre binding factor dbf1	1.70E-94	68.50
PvERF097	Pavir.Ia00345	DRE B	9	I	353	ap2-domain dre binding factor dbf1	1.74E-83	63.40
PvERF098	Pavir.Ia04894	DRE B	9	I	199	ap2-domain dre binding factor dbf1	4.07E-97	76.10

Table 4-2 Continued.

Gene name	Phytozome identifier or unitranscript ID	Sub-family	Chromosome	group	Amino acid sequence length	NCBI BLASTP annotation	Minimum eValue	Mean Similarity (%)
PvERF099	Pavir.Ib02495	DREB	9	III	186	dre binding factor 2	5.49E-54	65.90
PvERF100	Pavir.J13316	DREB		I	255	ap2-domain dre binding factor dbf1	6.23E-83	71.30
PvERF101	Pavir.Fa01600	DREB	6	I	286	ap2-domain dre binding factor dbf1	2.02E-90	68.70
PvERF102	AP13CTG59014	DREB	2	I	207	ap2-domain dre binding factor dbf1	1.29E-83	78.70
PvERF103	Pavir.J17038	ERF		X	250	erebp transcription factor	1.65E-76	76.10
PvERF104	Pavir.Aa01242	DREB	1	I	314	ap2-domain dre binding factor dbf1	4.04E-123	78.55
PvERF105	Pavir.Ga00913	DREB	7	I	413	ap2-domain dre binding factor dbf1	3.02E-104	79.55
PvERF106	Pavir.J01332	DREB		I	284	ap2-domain dre binding factor dbf1	8.49E-126	78.95
PvERF107	Pavir.Gb01735	ERF	7	X	354	ap2-related transcription factor	4.62E-53	84.40
PvERF108	Pavir.J28880	ERF		VII	222	erebp transcription factor	1.62E-29	79.05
PvERF109	Pavir.Aa02595	ERF	1	X	247	ap2-related transcription factor	8.98E-48	77.40
PvERF110	Pavir.Ib02796	ERF	9	VII	295	ap2 domain containing expressed	1.23E-47	63.35
PvERF111	Pavir.Ia04158	ERF	9	VII	403	ap2 domain containing expressed	3.57E-73	50.00
PvERF112	Pavir.Ib00527	ERF	9	VII	297	erebp transcription factor	1.14E-46	57.20
PvERF113	Pavir.Ib00526	ERF	9	VII	319	erebp transcription factor	4.10E-88	54.90
PvERF114	Pavir.Db00700	ERF	4	VIII	178	enhancer of shoot regeneration esr1-like protein	8.94E-12	61.33
PvERF115	Pavir.Bb03337	ERF	2	VII	358	erebp transcription factor	1.54E-59	58.15
PvERF116	Pavir.Aa00281	ERF	1	VII	370	erebp transcription factor	6.98E-166	68.00
PvERF117	Pavir.Ia01229	ERF	9	VII	274	erebp transcription factor	1.05E-131	73.75
PvERF118	Pavir.Da01893	ERF	4	VII	256	erebp transcription factor	4.77E-76	61.05
PvERF119	Pavir.Bb02257	ERF	2	VII	397	erebp transcription factor	7.27E-165	64.30
PvERF120	Pavir.Fa00811	ERF	6	VII	350	erebp transcription factor	1.48E-35	68.75
PvERF121	Pavir.Fb02184	ERF	6	IX	132	pathogenesis-related genes transcriptional activator pti5-like	1.20E-41	76.35

Table 4-2 Continued.

Gene name	Phytozome identifier or unitranscript ID	Sub-family	Chromosome	group	Amino acid sequence length	NCBI BLASTP annotation	Minimum eValue	Mean Similarity (%)
PvERF122	Pavir.Ib00287	ERF	9	IX	125	ap2 domain containing expressed	5.93E-48	84.70
PvERF123	Pavir.Eb00249	DREB	5	III	242	dre binding factor 2	3.83E-74	73.00
PvERF124	Pavir.Aa00832	DREB	1	III	303	dre binding factor 2	4.06E-51	77.85
PvERF125	Pavir.Aa00833	DREB	1	III	289	dre binding factor 2	3.24E-52	78.20
PvERF126	Pavir.Ga01094	DREB	7	III	233	dre binding factor 2	2.83E-57	80.30
PvERF127	Pavir.Aa00834	DREB	1	III	292	dre binding factor 2	1.48E-66	71.40
PvERF128	Pavir.Ga01093	DREB	7	III	222	dre binding factor 2	3.42E-64	76.70
PvERF129	Pavir.Ia02208	DREB	9	III	276	dre binding factor 2	2.34E-83	78.70
PvERF130	Pavir.Ga01095	DREB	7	III	302	dre binding factor 2	9.35E-73	77.00
PvERF131	Pavir.Ia02495	DREB	9	I	274	ap2-domain dre binding factor dbf1	9.06E-72	68.85
PvERF132	Pavir.Ib00031	DREB	9	III	204	dre binding factor 2	2.59E-65	65.90
PvERF133	Pavir.Ab02683	DREB	1	III	206	dre binding factor 2	8.66E-83	70.80
PvERF134	Pavir.Ga00932	DREB	7	III	200	dre binding factor 2	1.10E-78	72.70
PvERF135	Pavir.Aa00912	DREB	1	III	249	transcription factor cbf1	1.06E-96	67.00
PvERF136	Pavir.Ga00931	DREB	7	III	209	transcription factor cbf1	1.42E-101	65.70
PvERF137	Pavir.J35991	DREB		III	224	transcription factor cbf1	1.21E-69	69.75
PvERF138	Pavir.Fa00114	DREB	6	III	225	crt dre binding factor	1.33E-63	62.10
PvERF139	AP13ISTG62835	DREB		III	258	crt dre binding factor	7.58E-60	60.85
PvERF140	Pavir.J27101	DREB		III	289	crt dre binding factor	4.70E-95	65.10
PvERF141	Pavir.Ga01110	DREB	7	III	179	dre binding factor 2	1.52E-72	71.10
PvERF142	Pavir.Ba01281	DREB	2	III	238	crt dre binding factor	6.19E-90	65.85
PvERF143	Pavir.Ba01280	DREB	2	III	236	crt dre binding factor	4.23E-100	67.25
PvERF144	Pavir.Ea03031	ERF	5	IX	224	ethylene responsive protein	1.48E-37	74.25
PvERF145	Pavir.J03893	DREB		III	235	crt dre binding factor	6.87E-93	65.75

Table 4-2 Continued.

Gene name	Phytozome identifier or unitranscript ID	Sub-family	Chromosome	group	Amino acid sequence length	NCBI BLASTP annotation	Minimum eValue	Mean Similarity (%)
PvERF146	Pavir.Ba01283	DRE B	2	III	227	crt dre binding factor	4.05E-61	64.25
PvERF147	Pavir.J26137	DRE B		III	197	dre binding factor 2	9.74E-74	74.55
PvERF148	Pavir.J34361	DRE B		II	188	ap2 domain containing protein	5.55E-58	77.85
PvERF149	AP13ISTG61813	DRE B		II	214	ap2 domain containing protein	1.27E-72	74.80
PvERF150	Pavir.Ia03856	DRE B	9	II	242	ap2 domain transcription factor-like	3.34E-68	64.40
PvERF151	Pavir.Ia01487	ERF	9	X	194	ap2 domain-containing transcription factor-like	4.30E-14	62.00
PvERF152	Pavir.J02555	DRE B		II	236	dre binding factor 2	9.27E-87	66.80
PvERF153	Pavir.J26231	DRE B		II	225	dre binding factor 2	1.36E-61	66.65
PvERF154	Pavir.J06813	DRE B		II	189	ap2 domain transcription factor-like	1.58E-37	60.25
PvERF155	Pavir.Ca01725	DRE B	3	IV	339	ap2 domain containing protein	1.08E-87	73.15
PvERF156	Pavir.Ea00429	DRE B	5	IV	257	ap2 domain containing protein	1.91E-122	75.80
PvERF157	Pavir.J29572	DRE B		IV	252	ap2 domain containing protein	7.14E-31	69.55
PvERF158	Pavir.J25464	DRE B		IV	376	ap2 domain containing protein	2.07E-71	63.70
PvERF159	Pavir.J32663	DRE B		III	213	ap2 domain transcription factor-like	8.81E-31	65.45
PvERF160	Pavir.Ea03550	AP2	5	AP2	666	ap2 domain containing protein	0	67.75
PvERF161	Pavir.Da00839	DRE B	4	III	214	dre binding factor 2	7.07E-41	78.90
PvERF162	Pavir.Da01034	ERF	4	IX	124	ap2 domain containing expressed	2.58E-45	76.35
PvERF163	Pavir.Ba02199	DRE B	2	I	240	ap2-domain dre binding factor dbf1	7.21E-84	75.55
PvERF164	Pavir.Ba00410	ERF	2	VII	324	erebp transcription factor	1.50E-66	55.15
PvERF165	Pavir.J31711	ERF		XI	179	ap2 domain containing expressed	9.48E-12	57.60
PvERF166	Pavir.Ea00084	ERF	5	VII	193	ethylene-responsive transcription factor rap2-3-like		0.00

Table 4-2 Continued.

Gene name	Phytozome identifier or unitranscript ID	Sub-family	Chromosome	group	Amino acid sequence length	NCBI BLASTP annotation	Minnum eValue	Mean Similarity (%)
PvERF167	Pavir.Ib03119	ERF	9	IX	138	ap2 domain containing expressed transcription factor	2.06E-43	80.60
PvERF168	Pavir.Ib00323	DREB	9	III	127	rcbf2	5.24E-14	86.00
PvERF169	Pavir.J08112	ERF		IX	240	ethylene responsive element binding factor 5	2.19E-30	64.80
PvERF170	Pavir.J19621	ERF		VI-L	288	ap2 domain-containing transcription factor-like protein	4.47E-35	68.00
PvERF171	Pavir.J17768	ERF		IX	172	ap2 domain containing expressed	1.52E-27	67.60
PvERF172	Pavir.Hb01232	ERF	8	IX	311	ap2 domain containing expressed	1.06E-39	68.00
PvERF173	Pavir.Hb01145	ERF	8	IX	234	ethylene response factor erf1	3.57E-25	53.90
PvERF174	Pavir.J01079	ERF		IX	206	ethylene response factor erf1	4.65E-22	49.33
PvERF175	Pavir.Bb00652	ERF	2	IX	349	ap2 domain containing expressed	1.89E-51	67.85
PvERF176	Pavir.Bb03080	ERF	2	VII	135	ethylene-responsive element binding protein 2	9.77E-13	75.00
PvERF177	Pavir.J37767	AP2		AP2	182	ap2 domain containing expressed	1.52E-63	86.35
PvERF178	Pavir.J26710	ERF		VI	302	pti6 -like	1.68E-25	65.25
PvERF179	Pavir.J04335	ERF		X	362	ap2 domain containing expressed	1.17E-61	73.05
PvERF180	Pavir.J32493	ERF		IX	246	ap2 domain containing expressed	1.16E-34	70.55
PvERF181	Pavir.Ca01826	RAV	3	RAV	343	ap2 domain containing protein	1.48E-162	70.00
PvERF182	Pavir.Ea02092	RAV	5	RAV	416	ap2 domain containing protein	0	70.70
PvERF183	Pavir.Ca01165	RAV	3	RAV	223	ap2 domain containing protein	8.78E-80	72.05
PvERF184	Pavir.Ea00549	RAV	5	RAV	363	ap2 domain containing protein	8.14E-152	71.35
PvERF185	Pavir.Ea00485	RAV	5	RAV	322	ap2 domain containing protein	1.05E-114	72.05
PvERF186	Pavir.Ea03213	AP2	5	AP2	421	ap2 erbp transcription factor wrinkled1	1.93E-93	70.40
PvERF187	Pavir.J19507	AP2		AP3	456	ap2 dna-binding domain	7.27E-157	78.65
PvERF188	Pavir.Bb02213	AP2	2	AP2	387	ap2-erbp transcription factor	0	78.05

Table 4-2 Continued.

Gene name	Phytozome identifier or unitranscript ID	Sub-family	Chromosome	group	Amino acid sequence length	NCBI BLASTP annotation	Minmum eValue	Mean Similarity (%)
PvERF189	AP13CTG03306	AP2		AP2	442	aintegumenta-like protein	1.75E-145	76.95
PvERF190	AP13ISTG54847	AP2	6	AP2	429	ap2-erebp transcription factor	6.89E-123	73.40
PvERF191	Pavir.J02216	AP2		AP2	404	ap2 dna-binding domain	3.66E-168	79.05
PvERF192	Pavir.Ia03990	AP2	9	AP2	627	ac103891_21 ovule development protein	0	83.45
PvERF193	Pavir.Ga00050	AP2	7	AP2	473	aintegumenta-like protein	0	85.50
PvERF194	Pavir.Da00344	AP2	4	AP2	466	aintegumenta-like protein	0	83.65
PvERF195	Pavir.Hb01059	AP2	8	AP2	553	aintegumenta-like protein	0	73.45
PvERF196	Pavir.Ia03819	AP2	9	AP2	459	aintegumenta-like protein	4.74E-118	77.95
PvERF197	Pavir.Ba03911	AP2	2	AP2	622	ap2 erebp transcription factor	0	73.35
PvERF198	Pavir.Ia00425	AP2	9	AP2	651	ap2 erebp transcription factor	0	75.10
PvERF199	Pavir.Ia04350	AP2	9	AP2	516	ap2 domain containing expressed	0	82.45
PvERF200	Pavir.Aa01355	AP2	1	AP2	665	ap2 erebp transcription factor-like protein	0	79.35
PvERF201	Pavir.Ca00106	AP2	3	AP2	289	ovule development protein aintegumenta-like	3.63E-141	80.40
PvERF202	Pavir.Aa00475	AP2	1	AP2	374	ant-like protein	6.47E-156	75.15
PvERF203	Pavir.Da00466	AP2	4	AP2	387	transcription factor ap2d22	1.74E-137	79.25
PvERF204	Pavir.J23348	AP2		AP2	486	transcription factor ap2d23-like	0	76.90
PvERF205	Pavir.J23520	AP2		AP2	429	apetala2-like protein	0	75.95
PvERF206	Pavir.Ba03394	AP2	2	AP2	466	indeterminate spikelet 1	4.88E-177	74.45
PvERF207	Pavir.Gb00344	AP2	7	AP2	378	transcription factor ap2d23-like	9.39E-148	77.85

Table 4-3 Summary of the AP2/ERF superfamily gene members found in various plant species. The switchgrass (*Panicum virgatum*) data are from this study. Note that switchgrass is the only polyploidy species listed below.

Family	Subfamily	Group	<i>Panicum virgatum</i>	<i>Oryza sativa</i> ^a	<i>Arabidopsis thaliana</i> ^a	<i>Populus trichocarpa</i> ^b
AP2			25	29	18	26
ERF	DREB	I	12	9	10	5
		II	11	15	15	20
		III	27	26	23	35
		IV	5	6	9	6
		Total	55	56	57	66
	ERF	V	10	8	5	10
		VI	9	6	8	11
		VI-L	7	3	4	4
		VII	17	15	5	6
		VIII	25	13	15	17
		IX	37	18	17	42
		X	12	13	8	9
		Xb-L	-	-	3	4
		XI	4	7	-	-
		Total	121	76	65	103
RAV			5	5	6	6
Singleton			1	1	1	1
Total			207	174	147	202

^aNakano et al., 2006; ^bZhuang et al., 2008

Table 4-4 Chromosomal distribution of AP2/ERF superfamily in switchgrass.

Chromosome	DREB	ERF	RAV	AP2	Soloist	Total
1	9	11	-	2	1	23
2	5	19	-	3	-	27
3	2	5	2	2	-	11
4	3	8	-	2	-	13
5	2	9	3	2	-	16
6	3	14	-	1	-	18
7	7	11	-	2	-	20
8	-	7	-	1	-	8
9	9	17	-	4	-	30
Unmapped	15	20	-	6	-	41
Total	55	121	5	25	1	207

Table 4-5 Lists of conserved motifs discovered using the MEME Suite version 4.7.0. The height of a letter in the LOGO indicates its relative frequency at the given position.

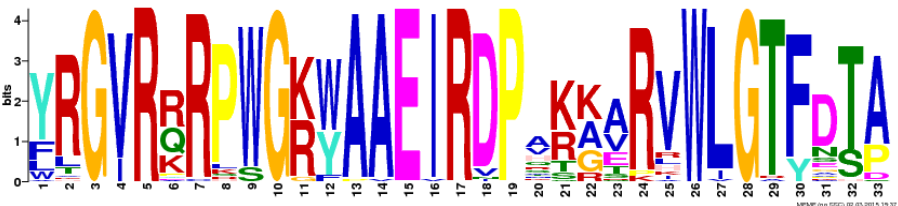
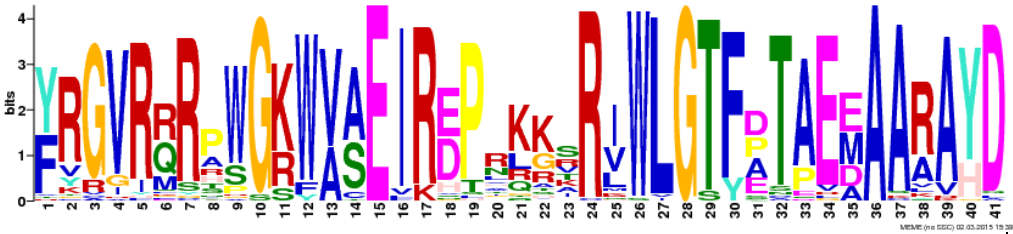
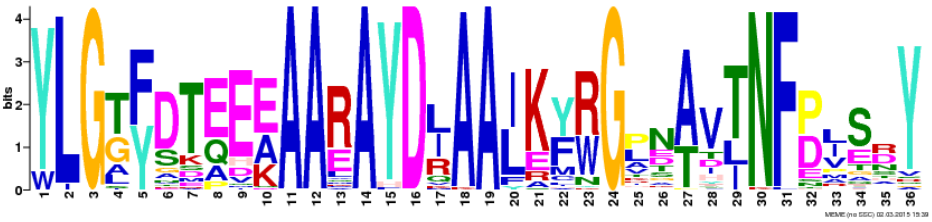
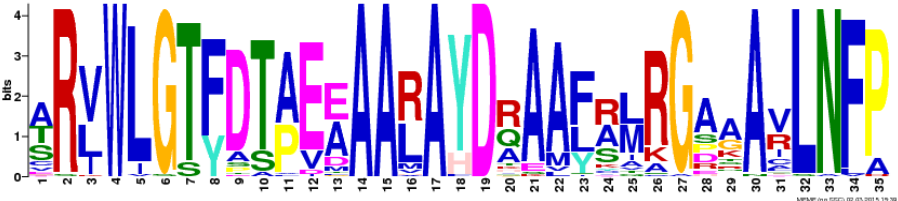
Motif	Logo	E value
M1		1.1e-1189
M2		6.1e-1323
M3		5.0e-1057
M4		3.7e-1048
M5		3.7e-733

Table 4-5 Continued.

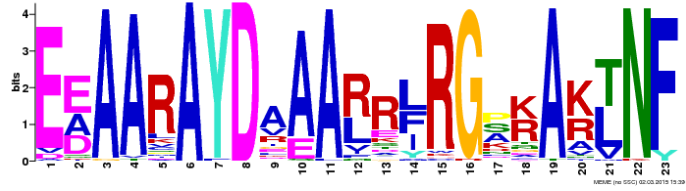
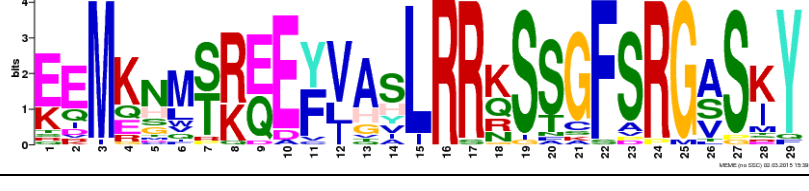
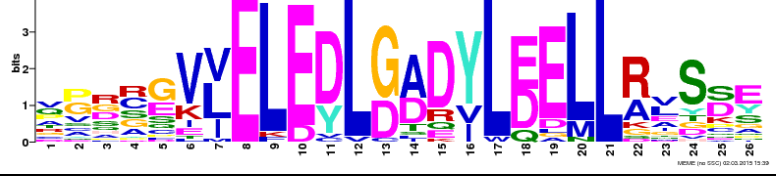
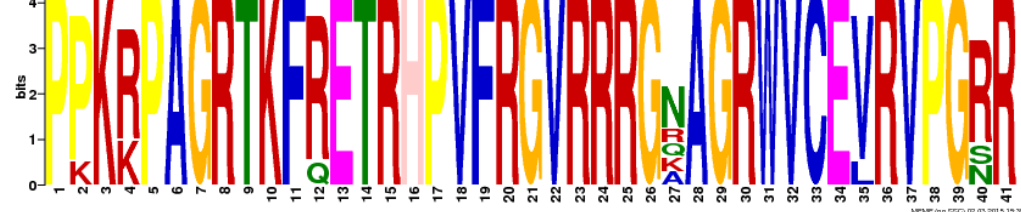
Motif	Logo	E value
M6		5.3e-564
M7		1.2e-492
M8		1.9e-349
M9		7.3e-278
M10		6.3e-193
M11		4.0e-158

Table 4-5 Continued.

Motif	Logo	E value
M12		7.8e-147
M13		1.3e-107
M14		2.5e-095
M15		1.1e-092
M16		8.0e-071
M17		1.9e-069
M18		7.5e-067

Table 4-5 Continued.

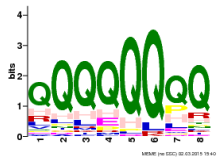
Motif	Logo	E value
M19		1.3e-064
M20		2.9e-063
M21		1.4e-060
M22		5.3e-058
M23		9.8e-054
M24		1.4e-053
M25		1.4e-050

Table 4-6 Morphology and biomass yields of transgenic switchgrass lines overexpressing *PvERF001* and non-transgenic control (WT) plants.

Lines	Tiller height (cm)	Tiller number	Fresh weight (g)	Dry weight (g)	plant diameter (cm)	Fresh/dry weight ratio
1	98.9±2.0 ^b	13.3±1.5 ^{bc}	40.5±1.7 ^{cd}	12.8±0.3 ^c	1.38±0.06 ^b	3.15±0.07 ^a
2	105.8±2.9 ^{ab}	12.3±2.6 ^c	45.8±11.3 ^{bcd}	15.1±3.9 ^{bc}	1.36±0.06 ^{bc}	3.05±0.04 ^a
3	115.3±1.2 ^a	17.7±1.9 ^{ab}	70.2±9.9 ^a	21.9±3.3 ^a	1.54±0.03 ^a	3.21±0.08 ^a
7	115.7±3.6 ^a	21.0±1.6 ^a	66.7±3.6 ^{ab}	17.7±5.5 ^{ab}	1.23±0.06 ^{cd}	3.03±0.17 ^a
8	101.3±2.1 ^b	15.3±0.9 ^{abc}	42.9±1.8 ^{cd}	13.3±0.9 ^c	1.20±0.02 ^d	3.23±0.14 ^a
9	109.7±3.1 ^{ab}	17.0±0.8 ^{abc}	61.4±3.6 ^{abc}	16.1±5.2 ^{ab}	1.44±0.05 ^{ab}	3.00±0.11 ^a
WT	85.8±2.8 ^c	15.3±1.5 ^{abc}	34.6±4.7 ^d	10.7±1.6 ^c	1.05±0.03 ^e	3.27±0.08 ^a

Tiller height estimates were determined for each plant by taking the mean of the 5 tallest tillers within each biological replicate. The fresh and dry biomass measurements were obtained from aboveground plant biomass harvested at similar growth stages. Values are means of 3 biological replicates ± standard errors (n=3). Values represented by different letters are significantly different at $P \leq 0.05$ as tested by LSD method with SAS software (SAS Institute Inc.).

Table 4-7 Sugar release by enzymatic hydrolysis in transgenic switchgrass overexpressing *PvERF001* and non-transgenic control (WT) lines.

Transgenic lines	Glucose release (g/g CWR)	Xylose release (g/g CWR)	Total sugar release (g/g CWR)
1	0.214±0.009 ^d	0.175±0.003 ^c	0.389±0.011 ^c
2	0.238±0.008 ^{bc}	0.182±0.014 ^{abc}	0.420±0.018 ^b
3	0.234±0.003 ^{bcd}	0.192±0.008 ^a	0.427±0.006 ^{ab}
7	0.247±0.003 ^{ab}	0.176±0.009 ^{bc}	0.423±0.007 ^b
8	0.261±0.003 ^a	0.188±0.005 ^{ab}	0.448±0.012 ^a
9	0.227±0.020 ^{bcd}	0.188±0.007 ^{ab}	0.415±0.024 ^b
WT	0.225±0.007 ^{cd}	0.181±0.007 ^{abc}	0.405±0.003 ^{bc}

All data are means ± standard error (n=3). Values represented by different letters are significantly different at $P \leq 0.05$ as tested by LSD method with SAS software (SAS Institute Inc.). CWR, cell wall residues.

Table 4-8 List of the locus names and/or GenBank accession numbers of the rice and *Arabidopsis* sequences used in Figure 4-10.

Gene name	MSU Locus name/accession number	Species
OsERF012	LOC_Os02g10760.1	<i>Oryza sativa</i>
OsERF023	LOC_Os02g55380.1	<i>Oryza sativa</i>
OsERF045	LOC_Os04g56150.1	<i>Oryza sativa</i>
OsERF057	LOC_Os06g40150.1	<i>Oryza sativa</i>
OsERF072	LOC_Os07g10410.1	<i>Oryza sativa</i>
OsERF075	LOC_Os07g38750.1	<i>Oryza sativa</i>
OsERF100	LOC_Os06g08340.1	<i>Oryza sativa</i>
OsERF159	LOC_Os12g39330.1	<i>Oryza sativa</i>
AtERF001	At1g15360	<i>Arabidopsis thaliana</i>
AtERF002	At5g19790	<i>Arabidopsis thaliana</i>
AtERF003	At5g25190	<i>Arabidopsis thaliana</i>
AtERF004	At5g11190	<i>Arabidopsis thaliana</i>
AtERF005	At5g25390	<i>Arabidopsis thaliana</i>

Chapter 5: Conclusions

The need for clean renewable energy is tremendous due to exasperating economic and environmental snags as a result of reliance on non-renewable energy from fossil fuels. Lignocellulosic bioenergy feedstocks from perennial C4 grasses provide a potentially viable and sustainable alternative source for production of biofuel. Lignocellulosic biomass is composed mainly of cell wall polymers such as cellulose, hemicellulose and lignin that constitute the major part of organic biomass on the planet. Bioethanol is produced from the sugars embedded in these cell wall polymers except lignin. Extraction of the sugars from the cell wall carbohydrates is expensive due to the resistance of these cell wall carbohydrates to enzymatic hydrolysis. This resistance known as biomass recalcitrance is primarily imparted by the presence of lignin. Ensuring economic viability of biofuel production from switchgrass lignocellulosic biomass requires modification of the biomass composition for easy extraction of sugar as well as securing continuous supply of abundant and quality biomass. Clear understanding of the biology of switchgrass biomass formation, plant interaction with the environment as well as regulatory mechanisms behind the biosynthesis of lignin polymers is essential to utilize genetic engineering approaches for the development of elite lines with ideal traits.

Genetic engineering of lignin biosynthesis by direct targeting of individual lignin biosynthetic genes or its transcriptional repressors has been shown to improve saccharification efficiency of lignocellulosic biomass. Here, genetic manipulation of GA catabolic pathway by overexpression of switchgrass *GA2ox* genes that catalyze the deactivation of bioactive GA or its precursors through 2 β -hydroxylation reaction to modify plant architecture and reduce biomass recalcitrance was attempted. Three C₂₀ switchgrass *GA2ox* genes were identified showing differential regulation patterns among tissues including roots, seedlings and reproductive parts. Overexpression of two (*PvGA2ox5* and *PvGA2ox9*) of the 3 C₂₀ switchgrass *GA2ox* genes in switchgrass displayed characteristic GA-deficient phenotypes with modified plant architecture. The observed GA-deficient phenotypes were rescued when bioactive GA was supplied via foliar application. Increased tillering with minor tradeoff in biomass yield and slightly reduced lignin content and the syringyl/guaiacyl lignin monomer ratio was observed in semi-dwarf lines, which was accompanied by the reduced expression of most lignin biosynthetic genes. A modest

increase in the level of glucose release by up to 15% was observed in the semi-dwarf lines. Results suggest that overexpression of *GA2ox* genes in switchgrass could be used to reduce biomass recalcitrance and modify plant architecture to provide protection against soil erosion, lodging, and weed colonization on marginal lands.

Overexpression of transcriptional repressors of the lignin biosynthesis pathway can be utilized to reduce biomass recalcitrance. Switchgrass gene coding for Knotted1-like TF (*PvKN1*), a putative transcriptional repressor of lignin biosynthesis pathway, was ectopically overexpressed in switchgrass resulting in altered growth phenotypes predominantly at early growth stages. Moreover, transgenic lines exhibited reduced expression of most lignin biosynthetic genes with modest reduction in lignin content. Based on the results, it appears that *PvKN1* may regulate lignin biosynthesis perhaps via its role in the regulation of GA signalling pathway as suggested by the reduced expression of GA biosynthetic gene (*GA20ox*) concurrently with increased expression of GA catabolic gene (*GA2ox*). An increase in the level of glucose release by up to 16% was observed in the transgenic vs the non-transgenic control lines. Therefore, overexpression of *PvKN1* in bioenergy feedstocks may as well provide an alternative strategy for improving biomass characteristics for biofuel.

To realize cost-efficient biofuel production from lignocellulosic biomass, on top of optimizing the biomass quality it is critical to enhance its biomass productivity under adverse environmental conditions in order to maintain continuous supply. Genes in the AP2/ERF TF superfamily that are known to be involved in the regulation of various growth and developmental programs including stress responses would perhaps provide ideal candidates to engineer bioenergy feedstocks for such traits. Genome-wide investigation of this TF superfamily identified a total of 207 genes belonging to 4 subfamilies containing 25 AP2, 121 ERF, 55 DREB and 5 RAV genes. The majority of genes in the AP2/ERF families likely participate in the transcriptional regulation of various biosynthetic processes and responses to environmental stimuli. The expression analysis of many of these genes showed diversified pattern during various stages of development from seed germination to seed maturation highlighting their putative role in plant growth and development. Quite remarkably, several members of ERF and DREB subfamilies were found to be highly expressed in actively lignifying tissues such as vascular bundles implicating potential

role in the regulation of the factors in cell wall biosynthesis pathway. Ectopic overexpression of an ERF subfamily gene (*PvERF001*) in switchgrass showed enhanced biomass productivity and modest increase in sugar release efficiency in transgenic lines signifying the potential of these TFs in the genetic engineering of various pathways to improve biomass characteristics of lignocellulosic feedstocks for biofuel. In conclusion, these results provide vital resources for mining genes that may be manipulated to impart switchgrass tolerance to various environmental stress factors as well as improve plant productivity and biomass quality.

Vita

Wegi Aberra Wuddineh was born on March 18, 1981 in Adaba, Oromia, Ethiopia to parents, Megersa Aberra and Tiki Degefa and raised by grandparents Aberra Wuddineh and Belayenesh Jima. He completed junior and secondary education at Adaba junior and secondary high school and graduated in 1999. He joined then Alemaya now Haromaya University in Dire-Dawa, Ethiopia and graduating with a Bachelor of Science degree in Plant Sciences major in 2003. Subsequently, he worked briefly on teaching position at Mizan-Teferi technical and vocational education and training college before joining Ethiopian Agricultural research institute to work as a junior researcher in breeding oilseed crops at Werer research center, Afar, Ethiopia. He joined Ben-Gurion University of the Negev (BGU), Israel as a graduate student in 2006, working on the development of seawater-grown cash-crop halophytes. He graduated with Masters of Science degree in Desert Studies with specialization in Dry Land Biotechnology in 2008. He joined Dr. Stewart's laboratory, Department of Plant Sciences at the University of Tennessee-Knoxville in fall of 2009 for graduate study for PhD. His research focused on improvement of switchgrass biomass characteristics for biofuel. Wegi is planning to pursue career in industry or academia.