July 2019

Brachypodium distachyon GNRF, SWAM1 and SWAM4 are transcriptional regulators of secondary cell wall biosynthesis

Sandra Romero-Gamboa

Brachypodium distachyon GNRF, *SWAM1* and *SWAM4* are transcriptional regulators of secondary cell wall biosynthesis

A Dissertation Presented

by

SANDRA P. ROMERO-GAMBOA

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

MAY 2019

Plant Biology Program

Brachypodium distachyon GNRF*, *SWAM1* and *SWAM4
are transcriptional regulators of secondary
cell wall biosynthesis

A Dissertation Presented
by
SANDRA P. ROMERO-GAMBOA

Approved as to style and content by:

_____
Samuel P. Hazen, Chair

_____
Ludmila Tyler, Committee Member

_____
Madelaine Bartlett, Committee Member

_____
John M. Lopes, Committee Member

_____
Elizabeth Vierling, Director
Plant Biology Graduate Program

DEDICATION

To my parents Ines Gamboa & Martin Romero, to my sisters Dylips, Luz & Rosita.

To my lovely little family, Daniel Carbajal-Quiroz & Gabriella Carbajal-Romero

ACKNOWLEDGMENTS

I would like to thank to my advisor, Samuel Hazen, for giving me the opportunity to be part of the Hazen Lab. I am grateful for his contributions in my academic education, his guidance in the scientific journey of graduate school, his valuable and constant encouragement, his financial support and for being a model of academic excellence, critical thinking and kindness.

I would like to thank my Dissertation Committee, Ludmila Tyler, Madelaine Bartlett and John M. Lopes, for their acceptance of being part of my committee, for their time and the prompt mentorship, and for helping me with my dissertation.

I am thankful to all Hazen Lab members who have been part of this journey. Thanks to my colleagues and friends Pubudu Handakumbura, Scott Lee, Ian McCahill, Ian Whitney, Michael Harrington & Anirudha Dixit. Special thanks to Joshua Coomey and Kirk Mackinnon for their collaboration on this dissertation.

I am thankful to all of the Plant Biology Program, especially to Susan Capistran, James Wright and Sarah Czerwonka. Thanks to Christopher Phillips and Daniel Jones for their collaboration at the Morrill greenhouse, as well as to all faculty and students.

I am thankful to Stephen Dellaporta and Maria Moreno for their friendship, help and support. I want to give a special acknowledgement to Christopher Fragaso and Xing Wu for their collaboration on this research. Thanks to Christopher Heffelfinger and Evan Buss for their support and friendship.

Finally, I would like to thank to my family and friends at UMass, Connecticut, Mexico and Colombia, for your love, prayers and support.

ABSTRACT

BRACHYPODIUM DISTACHYON GNRF, *SWAM1* AND *SWAM4* ARE TRANSCRIPTIONAL REGULATORS OF SECONDARY CELL WALL BIOSYNTHESIS

MAY 2019

SANDRA P. ROMERO-GAMBOA, B. SC UNIVERSIDAD DE LOS ANDES, COLOMBIA

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Samuel Hazen

Plant cell walls are complex structures that contain a matrix of cellulose, lignin and hemicellulose. The regulation of the biosynthesis of these components has been well-studied in the eudicot plant *Arabidopsis thaliana*, and a transcriptional network has been elucidated. Several NAC and MYB family transcription factors are key regulators of secondary cell wall biosynthesis, and their functional characterization provides significant insight into the complex underlying transcriptional network. Genetic and structural evidence suggests that genes controlling this process might be different between eudicots and monocots. Here, the model grass *Brachypodium distachyon* has been selected to characterize the function of *GNRF* (*GRASS NAC REPRESSOR OF FLOWERING*), *SWAM1* (*SECONDARY WALL ASSOCIATED MYB1*), and *SWAM4* in the regulation of secondary cell wall biosynthesis. Phylogenetic analysis identified that *GNRF* and *SWAM4* as the respectively *AtSND2* and *AtMYB61* transcription factors in *B. distachyon*. Co-expression analysis showed that both, *GNRF* and *SWAM4*, clustered with putative cell-wall-associated genes. Functional characterization was performed by using the overexpression plants *GNRF-OE* and *SWAM4-OE*; sodium azide mutant plants from a TILLING (Targeting Induced Local Lesion IN Genome) collection for *gnrf-1*, *gnrf-2*, *gnrf-3*, *gnrf-4, gnrf-5*, *swam4-1*, and *swam4-2*; a T-DNA insertional mutant plant, *gnrf-6*; and a dominant repressor plant, *SWAM4-DR. GNRF-OE* plants remained at juvenile stage and exhibited persistent vegetative growth, and some *gnrf* mutant plants were late flowering. *SWAM4-DR* plants were severely dwarfed. Stems of all genotypes were subjected to lignin quantification, cell wall

thickness measurements, Q-RT-PCR, and RNA-seq analysis. Cell wall and transcriptomic analysis revealed that *GNRF* is a repressor of *SWAM1*, a MYB activator of cell wall thickening, and represses genes encoding cellulose, lignin, and xylan biosynthetic enzymes. *GNRF* was found to function as a pleiotropic repressor of cell wall biosynthesis, flowering, and transport proteins. Protein-DNA interactions were revealed in yeast by yeast-one-hybrid assays; the *GNRF* binding site (CT/GTA/G/CA/TNNNNT/G/CAA/CA/T/GA/TA/T) was identified by DNA affinity purification sequencing (DAP-seq) assay. *SWAM4* is a putative regulator of cell wall biosynthetic genes (*CESA4*, *CESA7*, *CESA8*, *CAD*, and *COMT*), and other proteins associated with cell wall formation. Collectively, GNRF, SWAM4, and SWAM1 were characterized as secondary cell wall regulators in *B. distachyon*.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Plant cell walls

Plant cell walls are complex and dynamic structures that influence cell shape and differentiation, intercellular communication, and plant growth and development. Cell wall composition differs among plant species, developmental stage, and cell type (Farrokhi et al., 2006; Wolf et al., 2012). The primary cell wall is elastic and consists of a network of cellulose microfibrils embedded in a matrix of cross-linked hemicelluloses, pectins, and proteins. During secondary cell wall development, cellulose and hemicelluloses are deposited along with lignin, which confers structural rigidity. Cellulose consists of unbranched β-1,4 glucan chains that form long crystalline microfibrils. In most cases, cellulose microfibrils are highly organized and determine the direction of growth by mechanical anisotropy in a dynamic process in which expansins, wall-loosening factors, promote rearrangement of microfibrils (Cosgrove and Jarvis, 2012; Wolf et al., 2012). Hemicelluloses are synthesized by glycosyltransferases and bind to cellulose by non-covalent interactions to form a framework that is required for mechanical strength of cell walls (Zhong and Ye, 2015). While xyloglucans, xylans, mannans, and glucomannans are hemicelluloses found in terrestrial plants, beta-glucans are exclusively found in grasses (Scheller and Ulvskov, 2010). Lignin is a polyphenolic compound that cross-links with the other cell wall components and is formed from the radical polymerization of three distinct monolignols subunits: *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) (Hatfield and Vermerris, 2001). While complexes at the plasma membrane synthesize cellulose,

hemicelluloses are synthesized in the Golgi apparatus and transported in vesicles to the plasma membrane. The genes involved in the biosynthesis of the cell wall components have been extensively studied by genetic and biochemical approaches to identify the molecular elements and mechanisms related to plant cell wall formation (Zhong et al., 2018). *CELLULOSE SYNTHASE* (*CESA*) genes are part of a cellulose synthase complex. Three different CESA proteins are organized to form a rosette subunit and, subsequently, a hexameric synthase complex. *CELLULOSE SYNTHASE-LIKE* (*CSL*) genes are associated with the synthesis of the β-D-glycan backbone of hemicelluloses that are integrated into the cell walls by enzymes and binding mechanisms. After ceasing elongation, cell walls thicken in some cell types. Secondary cell wall deposition does not occur in all cells, but mainly in sclerenchyma and xylem cells that provide strength and support (Cosgrove, 2005; Cosgrove and Jarvis, 2012; Farrokhi et al., 2006; Nicol and Hofte, 1998; Wolf et al., 2012).

Industrial production of paper, lumber, textiles, films, and food thickeners among other products uses plant cell walls as raw material. The secondary cell wall constitutes the majority of plant biomass and a source of renewable energy in the biofuel industry (Kebrom et al., 2017). The use of a repertoire of functional integrated genomics, transcriptomics and proteomics techniques and imaging technologies has been used to identify genes and enzymes associated with cell wall biosynthesis, modification, and disassembly. These scientific advances have facilitated the study of plant biological processes and provided a better understanding of the molecular mechanisms related to secondary cell wall formation (Farrokhi et al., 2006; Nishitani and Demura, 2015). Genetic studies have also contributed to the understanding of plant cell walls as a system where the regulation of developmental processes occurs in response to both abiotic

and biotic factors (Bashline et al., 2014; Wolf et al., 2012). Wall construction and the regulation of cellulose, hemicellulose, and lignin biosynthesis have been investigated at the transcriptional level, and several transcription factors involved in this regulation have been identified (Handakumbura and Hazen, 2012; Hussey et al., 2013; Rao and Dixon, 2018; Zhang et al., 2018a). Understanding the molecular basis of plant cell wall transcriptional regulation undoubtedly generates knowledge of the molecular mechanism of its formation that can be used to maintain plant integrity, enhance biological properties in response to environmental changes, and to explore its usefulness in biotechnological applications.

1.2 Transcriptional regulators of secondary cell wall biosynthesis

Functional genomic analyses of the model eudicotyledonous plant *Arabidopsis thaliana* have identified several crucial transcription factors regulating secondary cell wall biosynthesis. Multiple studies have identified NAC (NAM, ATAF1,2, CUC2), MYB (Myeloblastosis), HB (Homeobox), and recently WRKY as families of proteins that contain key regulators of secondary cell wall biosynthesis (Hussey et al., 2013; Nakano et al., 2015; Rao and Dixon, 2018; Tian et al., 2007; Zhang et al., 2018a). The R2R3 MYB genes in *A. thaliana* were subject to study, and functional information was collected to isolate genes implicated in the regulation of the secondary cell wall within a transcriptional network (Stracke et al., 2001; Tian et al., 2007; Zhong and Ye, 2007). Functional genomic techniques were used to generate mutant plants that were used to identify and characterize genes related to cell wall formation. For instance, *AtMYB61* was found to be implicated in ectopic lignification by studying its function in the mutant plant *det3* in *A. thaliana* (Newman et al., 2004).

Interest in NAC protein structure and functionality increased (Olsen et al., 2005; Zhong and Ye, 2007), and functional genomic analysis with T-DNA lines and chimeric repressors was used to identify the NAC transcription factors *NAC SECONDARY WALL THICKENING PROMOTING FACTOR* (*NST1*) and *NST2* in the regulation of cell wall thickening (Mitsuda et al., 2005). Two other NAC proteins, VASCULAR-RELATED NAC-DOMAIN6 (VND6) and VND7 and were found to be transcriptional switches for xylem vessel formation (Kubo et al., 2005). By using the expression of chimeric repressors and double knockout plants, *NST1* and *NST3* (also called *SECONDARY WALL-ASSOCIATED NAC DOMAIN PROTEIN1, SND1*) were found to be key regulators of secondary wall thickening in interfascicular fibers and secondary xylem (Mitsuda et al., 2007; Zhong et al., 2006). NST1 and NST3 were found to have a redundant function (Zhong et al., 2007b). *SND1* was revealed to regulate *MYB46*, and subsequently, *MYB85* and *KNAT7* (*KNOTTED ARABIDOPSIS THALIANA7*) were highly upregulated when *MBY46* was overexpressed (Zhong et al., 2007a), suggesting a cascade and redundancy of transcriptional events. Evidently, a transcriptional network was elucidated, and *SND1* was defined as a master regulator with functional homology to *NST1, NST2, VND6*, and *VND7*. Secondary cell wall thickening in fibers was reduced by dominant repression of *SND2, SND3, MYB103, MYB85, MYB52, MYB54*, and *KNAT7.* Consistently, an increase in cell wall thickening was induced by overexpression of *SND2*, *SND3*, and *MYB103* (Zhong et al., 2008). RNA interference (RNAi) and T-DNA insertional techniques were used to identify more genes associated with this transcriptional network. Silencing of *MYB83* and *MYB46* by RNAi revealed their redundancy in the regulatory cascade of secondary cell wall in fibers and vessels (McCarthy et al., 2009). Genetic studies showed relevant and specific information about the direct regulation of cellulose synthase genes, lignin polymers, monolignols, hemicelluloses and genes related to

their biosynthetic pathways. For example, functional genomics studies of *MYB58* and *MYB68* revealed their association to lignin biosynthesis by the regulation of the laccase *LAC4* (Zhou et al., 2009), and the overexpression of the NAC protein *SND2* shown upregulation of the cellulose synthase gene *CESA8*; this line was used in transcriptome-wide gene expression profiling that revealed the upregulation of cellulose, xylan, mannan and lignin biosynthetic genes (Hussey et al., 2011b). Two proteins from *VND1* through *VND5* were functionally characterized as transcriptional regulators of secondary cell wall formation in vessels in *A. thaliana* (Zhou et al., 2014).

Transcriptional regulation requires protein-DNA interactions, and different methods have been used to identify them including DNA binding assays *in vitro* and *in vivo.* The regulatory mechanisms of gene expression in tracheary elements (TEs) led to the finding of an 11-bp cis-element TERE (TE-specific expression) to promote transcriptional activation (Pyo et al., 2007). Genome-wide analysis showed that *MYB46* directly regulates the expression of genes associated with secondary cell wall biosynthesis and also that *MYB83* binds to the same SMRE (secondary wall MYB responsive element) with the consensus site ACC(A/T)A(A/C)(T/C) (Zhong and Ye, 2012). Genome-wide analyses of target genes were also performed on the NAC genes *SND1*, *NST1*, *NST2*, *VND6*, and *VND7*, collectively grouped as secondary wall NACs (SWNs), and indicated their binding to an imperfect palindromic 19-bp SNBE (Secondary Wall NAC Binding Element: (T/A)NN(C/T)(T/C/G)TNNNNNNNA(A/C)GN(A/C/T)(A/T)) (McCarthy et al., 2011; Zhong et al., 2010). Transcription factor binding profiles and histone modifications have been identified by chromatin immunoprecipitation (ChIP) (Spencer et al., 2003). Currently, these types of techniques have been adapted to next-generation sequencing approaches to develop high-throughput methods and obtain large-scale datasets. (Kaufmann et al., 2010; Pepke et al.,

2009; Zhang et al., 2008). For instance, ChIP-seq and DAP-seq (DNA Affinity Purification Sequence) have been used as high-throughput transcription factor binding site discovery methods to find binding motifs (O'Malley et al., 2016). DAP-seq was used to identify a VND, NST, and SND (VNS) consensus element C(G/T)TNNNNNNNA(A/C)G in *A. thaliana* (Olins et al., 2018).

Genetic data from the transcription factors and target genes identified in previous studies of transcriptional regulation of secondary cell wall biosynthesis in *A. thaliana* were used to investigate orthologous genes in other species and subsequently discover components of the regulation of cell wall components. Among other eudicot plants, functional characterization of MYB proteins in eucalyptus trees led to the identification of *EgMYB1* in *Eucalyptus grandis* as highly expressed in differentiating xylem and containing an active repressor motif that negatively regulates cell wall formation (Legay et al., 2010). *EgMYB2* and the pine *PtMYB4* of *Pinus taeda* were found to bind the SMRE element and identified as transcriptional activators during wood formation (Zhong et al., 2013).

Plant cell wall composition and molecular dynamics are diverse within walls of different cells and plants from different species (Foster et al., 2010). The structural diversity of stems and wood in eudicot, monocot, and conifer plants indicates differences in cell composition and presumably denotes a divergence in the biosynthetic process of cell wall formation (Handakumbura and Hazen, 2012; Plomion et al., 2001). In grasses, the maize transcription factor *ZmMYB31* was identified as a repressor of genes involved in lignin biosynthesis. Plants mutant for *ZmMYB31* were dwarf, and genetic analysis showed that this transcription factor represses monolignol production possibly via an AC-type binding element that was found by ChIP (Fornale et al., 2010). The rice and maize NAC proteins *OsSWN* and *ZmSWNs*

(*SECONDARY WALL ASSOCIATED NACs*) were expressed in *A. thaliana*, resulting in ectopic deposition of cellulose, xylan, and lignin. Both proteins were found to activate cell wall formation and complement the double mutant *snd1/nst1*. *ZmMYB46* and *OsMYB46* were identified, as well, as positive regulators of cell wall synthesis and characterized as having the same regulatory functions as their orthologous genes in *A. thaliana* (Zhong et al., 2011). Similarly, *OsSND2* the ortholog of *AtSND2* in rice, was identified as a positive regulator of cell wall synthesis and as a putatively regulator of *OsMYB61a* and *OsMYB61b*, orthologous genes of *A. thaliana* activators of cell wall formation. Annotated coding sequence data of Arabidopsis, poplar, rice, maize, and switchgrass were used in a comparative genomic analysis to facilitate the study of transcription factors and establish genetic associations among grasses. Conservation among R2R3 MYB transcription factors was identified (Zhao and Bartley, 2014). Transcriptional co-regulatory network analysis of MYB transcription factors in rice has been performed to facilitate the understanding of the function of *OsMYB* transcriptional regulators within a regulatory network (Hirano et al., 2013b; Smita et al., 2015). These co-expression analyses of MYB proteins contribute to the identification of genes related to cell wall formation. In the grass *Brachypodium distachyon*, secondary cell wall biosynthetic genes have been characterized (Handakumbura et al., 2013; Trabucco et al., 2013), and co-expression analyses were generated to identify genes involved lignin biosynthesis (Sibout et al., 2017). *SWAM1*, a MYB transcription factor was identified as a positive regulator by functional genetic approaches (Handakumbura et al., 2018). Transcriptional regulation of secondary cell wall biosynthesis in grasses has gained interest and compared to eudicots, appears to be relatively conserved; nevertheless, gene regulation also seems to maintain unique features within grasses.

Secondary cell wall regulation was also studied in gymnosperm trees. Transcriptional

regulation studies in *Pinus taeda* revealed that the recombinant protein *PtMYB4* binds to AC elements, rich in adenosine and cytosine, of lignin biosynthetic genes. Additionally, lignin deposition was increased in tobacco plants overexpressing *PtMYB4* (Patzlaff et al., 2003a). *PtMYB1* was abundantly expressed in xylem cells in pine, and transient transcriptional activation demonstrated a possible function of *PtMYB1* as a transcriptional regulator by binding AC elements of *PAL2* (*PHENYLALANINE AMMONIA-LYASE 2*) in pine xylem (Patzlaff et al., 2003b). Ectopic deposition of lignin in plants overexpressing *PtMYB8* and upregulation of lignin biosynthetic genes in *PtMYB8* and *PtMYB1* led to the implication of these transcription factors in the regulation of lignin (Bomal et al., 2008)

Families of transcription factors have been studied to identify genetic and protein information about DNA binding motifs and functional annotations that reveal transcription factors associated with different biological processes. In *Populus* (poplar), a comprehensive expression analysis of NAC proteins and additional analysis by quantitative real time PCR (Q-RT-PCR) to confirm tissue-specific expression were performed (Hu et al., 2010). This study provided information about the divergence of these proteins and the occurrence of gene duplications. MYB transcription factors are associated with numerous functions in *A. thaliana,* and phylogenetic comparisons have revealed differences and associations with monocots and other eudicot plants (Yanhui et al., 2006). Functional genomics in *Populus, A. thaliana* and *Vitis vinifera* (grape) led to the identification of two transcriptions factors from the largest family of transcription factors WRKY, WRKY12 and WRK19, that function as positive and negative regulators of xylem development, respectively (Yang et al., 2016). Recent studies in switchgrass aim to use meta-analysis to find common features of the MYB, NAC and WRKY families of transcription factors (Rao et al., 2019).

Among the transcription factors involved in secondary cell wall biosynthesis in land plants identified in the different families of proteins, conservation of gene function seems to be associated with hierarchical position in the transcriptional regulatory network. While transcription factors identified as master switches appeared to share a conserved function, the function of other transcription factors that are regulated by these master switches seems to diverge (Zhang et al., 2018a). Among grasses, the transcriptional regulation seems to be conserved and specific to grasses compared to eudicots (Rao and Dixon, 2018).

The functional characterization of transcription factors of NAC, MYB, HB and WRKY families in *A. thaliana* has been the foundation to study the regulation of secondary cell wall biosynthesis in plants. Characterization of orthologous genes from these transcription factor families in other species has provided insight into the molecular mechanisms behind cell wall formation. This dissertation aims to functionally characterize *GNRF, SWAM1*, and *SWAM4* in the transcriptional regulation of secondary cell wall biosynthesis in the model grass *Brachypodium distachyon*.

CHAPTER 2

GRASS NAC REPRESSOR OF FLOWERING (*GNRF*) IS A TRANSCRIPTIONAL REPRESSOR OF CELL WALL BIOSYNTHESIS AND FORMS A NEGATIVE FEEDBACK LOOP WITH *SWAM1* TO REGULATE SECONDARY WALL THICKENING IN *BRACHYPODIUM DISTACHYON*

2.1 Introduction

Functional genomic analyses of the model plant *A. thaliana* have identified several key transcription factors regulating secondary cell wall biosynthesis. Gene expression analysis revealed that several NAC (NAM, ATAF1,2, CUC2) and R2R3 type MYB (MYELOBLASTOSIS) family proteins are key regulators of secondary cell wall biosynthesis in sclerenchyma cells. MYB transcription factors that have two highly conserved DNA-binding domains called R2R3 domains, have been studied and genetic analysis of these R2R3 MYB proteins has revealed the role of *MYB46* and *MYB83* as activators of secondary cell wall biosynthesis (Nakano et al., 2015; Zhong and Ye, 2012). Overexpression of *NAC SECONDARY WALL THICKENING PROMOTING FACTOR1* (*NST1*) and *NST2* induced ectopic thickening of secondary walls and these genes were also required for anther dehiscence (Mitsuda et al., 2005). By using the expression of chimeric repressors and double-knockout plants, *NST1* and *NST3* (also called *SECONDARY WALL-ASSOCIATED NAC DOMAIN PROTEIN1, SND1*) were found to be key regulators of secondary wall thickening in interfascicular fibers and xylem (Mitsuda et al., 2007). The NAC-domain transcription factors *VASCULAR-RELATED NAC-DOMAIN*, *VND6*, and *VND7* were associated with metaxylem and protoxylem vessel differentiation by a

microarray analysis (Kubo et al., 2005). A complex transcriptional network was described, and *SND1* was defined as a master regulator with functional homology to *NST1*, *NST2*, *VND6*, and *VND7*. Secondary cell wall thickening in fibers was reduced by dominant repression of *SND2, SND3*, *MYB103*, *MYB85*, *MYB52*, *MYB54*, and *KNAT7* (*KNOTTED ARABIDOPSIS THALIANA7*). Conversely, an increase in cell wall thickening was induced by overexpression of *SND2*, *SND3*, and *MYB103* (Zhong et al., 2008). *VND1* through *VND5* proteins were functionally characterized as transcriptional regulators of secondary cell wall formation in vessels in *A. thaliana* (Zhou et al., 2014). The study of NAC and MYB transcription factors in the regulation of secondary cell wall biosynthesis in *A. thaliana* has been the foundation to identify an intricate plant transcriptional network of cell wall formation.

Transcription factors regulate gene expression through their interaction with *cis*-elements localized in upstream regions and in the gene bodies their target genes. Ectopic expression of *VND6* and *VND7* suggested that these transcription factors regulate the expression of genes containing the 11-bp TERE motif [Tracheary Element Regulating *cis*-Element: CTT/(C)NAAA/(C)GCNA(T)] (Pyo et al., 2007). Genome-wide analysis of the NAC genes *SND1*, *NST1*, *NST2*, *VND6* and *VND7* collectively grouped the encoded proteins as secondary wall NACs (SWNs), which have been shown to interact with the 19-bp SNBE [Secondary Wall NAC Binding Element: (T/A)NN(C/T)(T/C/G)TNNNNNNNA(A/C)GN(A/C/T)(A/T)] (Zhong et al., 2010). A high-throughput transcription factor binding site discovery method was created to find binding motifs (O'Malley et al., 2016), and identified a VND, NST, and SND (VNS) consensus motif C(G/T)TNNNNNNNA(A/C)G (Olins et al., 2018). Multiple putative SNBEs were also found in the promoters of downstream targets of SND1, including *MYB83, MYB103, SND3*, and *KNAT7* (Zhong et al., 2010).

In *A. thaliana,* a genome-wide characterization of the biosynthesis of lignin during vascular development revealed the presence of well-conserved *cis*-regulatory AC-type promoter elements [ACC(A/T)A(A/C)(T/C)] in genes functioning in vascular lignification (Nakano et al., 2015; Raes et al., 2003; Zhong and Ye, 2012). Direct targets of *AtMYB46* and *AtMYB83* were identified by using an estrogen-inducible direct activation system. The target sequence was designated as the SMRE (secondary wall MYB responsive element), and the map of the consensus sequence was designated as ACC(A/T)A(A/C)(T/C). Genome-wide analysis showed that *MYB46* directly regulates the expression of genes associated with secondary cell wall biosynthesis and that *MYB83* binds to the same SMRE consensus (Zhong and Ye, 2012). The identification of functional genomic regulatory elements is essential to understanding gene regulation and to further establish regulatory networks.

Gene network analysis showed the co-expression of *SND* genes *(SND1, SND2,* and *SND3)* and cell-wall-regulating genes (Yao et al., 2012). *SND2* and *SND3*, distantly related to *SND1* (Hu et al., 2010), are two transcription factors that were found to be positive regulators of secondary cell wall thickness in *A. thaliana* and were included in the transcriptional network described (Zhong et al., 2008). *SND2* overexpression and analysis of fiber cross-sectional area in *Eucalyptus* trees demonstrated a significant increase in the cell wall thickness (Hussey et al., 2011a). Poplar *SND2*, an ortholog of *AtSND2* in *Populus trichocarpa,* was also found to positively regulate secondary cell wall thickening, as well as lignin and cellulose biosynthesis (Wang et al., 2013). On the contrary, overexpression of *Populus PopNAC154*, showed a reduction in height and increased in the proportion of xylem and phloem-cambial tissue (Grant et al., 2010). These divergent results were an indication of a diversified and poorly understood function, between *SND2* orthologous genes in woody and herbaceous plants.

Far less is known of transcriptional regulation of secondary cell wall thickening in grasses than in *A. thaliana*. Vascular bundle arrangements and cell wall composition vary between eudicots and monocots, which suggests a distinct transcriptional regulation (Handakumbura and Hazen, 2012; Vogel, 2008). A co-expression gene module study revealed that expression levels of two related *SND2* genes in rice were positively correlated with cell wall synthesis (Guo et al., 2014). One of these genes, *OsSND2* (*Os05g48850*), was identified as an activator of cellulose biosynthesis possibly by binding to MYB transcription factors (Ye et al., 2018). In general, *SND2* appears to be a cell wall activator, but its function may differ between *SND2* orthologs in woody and herbaceous plants, and further analysis is required to elucidate its function in grasses. This chapter describes a functional characterization of Bradi2g46197 *GNRF* (*GRASS REPRESSOR OF FLOWERING*) encoding a grass protein ortholog to SND2 in *B. distachyon*. Overexpression and mutant plants were studied with genetic and biochemical approaches to gain insight into the function of this putative transcription factor.

2.2 Materials and Methods

2.2.1 Phylogenetic analysis

Twenty-seven NAC proteins from *Arabidopsis thaliana*, *Arabidopsis lyrata, Capsella rubella*, *Glycine max* (soybean), *Citrus sinensis* (orange), *Solanum lycopersicum* (tomato), *Solanum pimpinellifolium* (currant tomato), *Vitus vinifera* (grape), *Theobroma cacao* (cacao), *Populus thrichocarpa* (poplar), *Brachypodium distachyon*, *Sorghum bicolor* (sorghum), *Oryza sativa* (rice), *Triticum aestivum* (wheat), *Hordeum vulgare* (barley), *Setaria viridis, Panicum virgatum* (switchgrass), *Ananas comosus* (pineapple), *Musa acuminata* (banana) and *Amborella*

trichopoda were selected for protein similarity to *GNRF* and economical interest in agriculture and biomass production. Protein sequences were downloaded from NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi), TAIR BLAST v2.2.8 (https://www.arabidopsis.org/Blast/), and Phytozome v12.1 (https://phytozome.jgi.doe.gov) servers. Protein sequences of GNRF and NAC homologous proteins were aligned using the MAFFT service for multiple sequence alignment with the iterative refinement method L-INS-I (Katoh et al., 2017). Sequences were selected to construct the phylogenetic tree using the neighbor-joining method, JTT substitution model, and bootstrap resampling value of 1000. An unrooted phylogenetic tree of GNRF homologs was visualized on the web application Phylo.io (http://phylo.io) (Robinson et al., 2016).

2.2.2 Plant material and growth conditions

B. distachyon inbred line Bd21-3 was used throughout (Vogel and Hill, 2008). Seeds from wild-type (WT) and mutant alleles describe below were imbibed for ten days in tubes with water at 4°C. Seeds were planted in potting mix (Sun Gro Sunshine #8 / Fafard 2 Mix, Burton, OH) combined with turface (Pro's Choice Sports Field Products, Chicago, IL) 3:1. The soil was treated with Gnatrol (Valent Bioscience Corporation, Libertyville, IL) before planting. Plants were grown in a growth chamber at long-day conditions of 20-hours light/4-hours dark at 26°C and 18°C respectively. Short-day grown plants were placed in a growth chamber for 10-hour light/14-hour dark at 26°C and 18°C, respectively. The growth chamber was equipped with a combination of 40W halogen bulbs and 215W fluorescent bulbs. Light intensity was 220 μmol m^{-2} sec^{-1} and relative humidity was 68%.

2.2.3 *GNRF* overexpression plants *(GNRF-OE)*

GNRF overexpression plants were previously created by cloning the full-length coding region of *GNRF* into a modified version of the Gateway-compatible pOL001 destination vector. This construct contains the maize ubiquitin protomer to obtain constitutive expression and the hygromycin resistance gene that functions as a selectable marker. *B. distachyon* callus was transformed with this construct via *Agrobacterium*-mediated transformation (Vogel et al., 2006; Vogel and Hill, 2008). The seed was planted, and leaf tissue was used to extract DNA as described (Handakumbura et al., 2013). *GNRF-OE* plants were identified by amplifying by PCR the hygromycin resistance gene under the following conditions: 95°C for 30s followed by 35 cycles of 95°C for 30s, 57°C for 45s, and 68°C for 60s, with a final extension step at 68°C for 5 min. Genotyping was confirmed by PCR amplification of a fragment from the junction between the *ZmUbi* promoter from the pOL001 vector and the *GNRF* gene. This amplification was performed using Taq DNA polymerase (New England Biolabs) under the following conditions: 95°C for 30s followed by 35 cycles of 95°C for 20s, 60°C for 30s, and 68°C for 40s, with a final extension step at 68°C for 5 min. Primers are listed in Table 2.1.

2.2.4 *gnrf-1*, *gnrf-2*, *gnrf-3*, *gnrf-4*, *gnrf-5* mutant allele identification

Five lines from a TILLING (Targeting Induced Local Lesion IN Genome) collection that were mutagenized with sodium azide (NaN_3) at INRA-Versailles, France were used in this study (Dalmais et al., 2013). All mutant lines were planted, genotyped and confirmed by sequencing to identify several mutations per line. Five mutant lines were selected including four individual

non-synonymous mutations and one with the introduction of a stop codon. Genomic DNA extraction from leaf tissue was performed as described (Handakumbura et al., 2013). PCR amplification was performed by using a Taq DNA polymerase kit (New England Biolabs, Ipswich, MA) and specific primers that amplified a fragment containing each mutation, which was confirmed by sequencing (Table 2.1). Briefly, a 619 bp fragment was amplified by PCR under the following conditions: 95°C for 30s followed by 40 cycles of 95°C for 20s, 55°C for 30s, and 68°C for 40s, with a final extension step at 68°C for 5 min. Amplicons were gel-purified and extracted (Zymoclean Kit, Zymo research, Irvine, CA) and sent to Macrogen USA, Boston, MA, for sequencing. In addition, PCR amplicons of *gnrf-1* mutant samples were genotyped by BaeGI (New England Biolabs) digestion of the fragments. *gnrf-1* PCR products were precipitated and cleaned with 20 μg/μl of glycogen (ThermoFisher, Waltham, MA), sodium acetate pH 5.2 and 100% ethanol at -20 °C overnight; samples were then washed with 70% ethanol and resuspended in 1XTE buffer (10mMTris-HCl containing 1mM EDTA). Purified amplicons were digested with BaeGI at 37°C for 2 hours. While WT-type allele digestion resulted in two bands of 145 bp and 474 bp, the *gnrf-1* allele remained undigested. Homozygosity fo the *gnrf-3* allele was identified by amplifying a 578 bp fragment by PCR under the following conditions: 95°C for 30s followed by 40 cycles of 95°C for 20s, 60°C for 30s, and 68°C for 40s, with a final extension step at 68°C for 5 min. Amplicons were gel-purified, extracted and sent for sequencing as described above.

2.2.5 *gnrf-6* mutant allele identification

Seed from the line JJ5517, with a *B. distachyon* Bd21-3 background, for the *gnrf-6* allele was obtained from the Joint Genome Institute (JGI) Brachypodium T-DNA insertion collection.

This line was created with the pJJ2LBA vector that contains transcriptional enhancers for activation tagging within the T-DNA sequence (Bragg et al., 2012). *gnrf-6* contains a T-DNA insertion within the *GNRF* 5'UTR (318 bp upstream of the translation start site). To identify plants positive for the insertion, leaf tissue was used for genomic DNA extraction and genotyping by PCR amplification as described in Bragg et al. 2012. Briefly, gene-specific primers flanking the insertion site were used in combination with a primer known to bind the left border of the T-DNA region to diagnostically amplify, or fail to amplify, the insertion region depending on the presence of the T-DNA. Amplification was performed with a Taq DNA polymerase kit (New England BioLabs) under the following conditions: 95°C for 30s followed by 40 cycles of 95°C for 30s, 55°C for 30s, and 68°C for 30s, with a final extension step at 68°C for 5 min. Primers are listed in Table 2.1

2.2.6 Plant phenotyping

Flowering time was determined when the inflorescence emerged from the flag leaf, and data were collected from 6-11 plants per genotype. Above-ground biomass was measured at senescence from 3-7 plants per genotype. Plant height data were collected from 3-4 plants per genotype at the senesced stage. Error bars correspond to means ± SEM (Standard Error of the Mean). Two-tailed Student's *t*-tests were performed, and significance was set to $p<0.05$ and $p<0.01$.

2.2.7 Cross sections and histochemical analysis of stem lignification

The first internode of the tallest stem of the plants at complete senescence was sectioned using a vibrotome Leica VT 1200S (Leica Biosystems, Buffalo Gove, IL). Stems were imbibed in 5.4% agarose and cut with one side of a double-edge razor blade. The 55μm stem cross sections were obtained with a speed of 5-50, and frequency 8. Cross sections were stained with phloroglucinol-ethanol 100mg/5ml for 2 mins and HCl-dH_2O 1:1 as previously described (Matos et al., 2013). The images were captured in a Nikon Eclipse E200MV microscope (Nikon, Melville, NY) and a PixeLINK 3 MP camera (PixeLINK, Ottawa, Canada). The images were visualized using PixeLINK uSCOPE software (PixeLINK, Ottawa, Canada), captured using 4x, 10x and 20x microscope objective lenses, and processed with Adobe Photoshop CC 20.0.1 (Adobe Systems, San Jose, CA). Areas and perimeters of cross sections were measured for the stained sections and analyzed in Image J 1.50i (Wayne Rasband NIH, USA).

2.2.8 Interfascicular fiber and xylem vessel wall thickness measurements

Stained images from cross sections were used to perform cell wall thickness measurements. The images captured with the 20x objective of the Nikon Eclipse E200MV microscope were used. The images were analyzed in ImageJ. Cell wall thickness measurements were made from the interfascicular region that contains mostly sclerenchyma fibers, and from xylem vessels in the vascular bundle area. Lines from across the adjacent cell wall were drawn by using the line tool. Three independent cross sections were analyzed with a total of 45 thickness measurements per plant as previously described (Handakumbura et al., 2018).

2.2.9 Acetyl bromide soluble lignin measurement

Stem tissue from 3-5 plants per genotype at complete senescence was pulverized and used for a quantitative assessment of the lignin content as described (Foster et al., 2010; Handakumbura et al., 2018). Briefly, 1.0-1.5 mg of stem pulverized stem tissue was rinsed in a 2 ml flask tube and rinsed with 250 µl of acetone. After evaporation of the acetone, 100 µl of freshly made acetyl bromide solution (25% v/v acetyl bromide in glacial acetic acid) was added, and the samples were heated at 50°C for 2 h under the hood. Samples were heated for an additional hour and vortexed every 15 minutes. Tubes were cooled on ice, and 400 µl of 2M sodium hydroxide and 70 µl of freshly prepared 0.5M hydroxylamine hydrochloride were added. Tubes were vortexed and filled with glacial acetic acid to a final volume of 2 ml. Samples were mixed, and 200 µl was used to measure absorbance at 280 nm in a UV-transparent 96-well plate on a microplate reader (SpectraMax M5). Percentage of acetyl bromide soluble lignin (%ABSL) was calculated using the grass coefficient (17.75) for Brachypodium with the formula described (Foster et al., 2010).

2.2.10 Transient gene expression in rice protoplasts

Two-week-old rice seedlings grown in the dark on Murashige and Skoog (MS) sucrose were finely cut and incubated in enzyme solution to make protoplasts. pENTR-D TOPO:*GNRF* and pENTR-D TOPO:*gnrf-1* constructs were created by cloning full length cDNA *GNRF* and *gnrf-1* and then used to clone these coding sequences downstream of the 35S promoter into the destination vector p2GW7. Constructs were co-transformed with a *Ubi:GUS* vector as an internal transformation control and four replicates were used. Each protoplast transformation was carried

out with $5x10^5$ cells of rice protoplast cells and 6ug of each construct. After transformation, RNA was extracted using the Zymo Quick RNA prep kit, and 500 ng of RNA were used to synthesize cDNA using BioRadiScrip reverse transcription supermix. Next, the cDNA was diluted 5-fold and used to perform Q-RT-PCR analysis. Rice Ubiquitin5 (OsUbi5) and coil-coil Protein 55 are stably expressed in different rice tissues and were used as reference genes for qPCR. Four technical replicates were performed and two-tailed t-test $p<0.05$ was used for statistical analysis. In collaboration with the laboratory of Laura Bartley, rice protoplast preparation and transient gene expression were carried out at the University of Oklahoma, Norman.

2.2.11 RNA extraction

For gene expression analysis by Q-RT-PCR, RNA was extracted from plants grown in short days with a 10:14 hours light:dark cycle. Leaf and stem samples were taken at Zeitgeber Time (ZT) 14 h from 3-6 plants at the same vegetative stage into 1.5 ml microcentrifuge tubes containing two metal beads. Samples were flash-frozen in liquid nitrogen and store at -80°C until processing. Total RNA was extracted using the RNeasy Plant mini kit (Qiagen, Gaithersburg, MD) and treated with RNase-Free DNase (Qiagen). RNA samples were resuspended in 30 µl DEPC-treated water.

For RNAseq transcriptome profiling, stem tissue samples from four developmentally comparable plants each of WT*, GNRF-OE*, *gnrf-1*, *gnrf-2*, *gnrf-3*, *gnrf-4* and *gnrf-6* were individually taken once the inflorescence had fully emerged, stage 5-9 on the BBCH-scale (Hong

et al., 2011). Samples were taken in the middle of the dark cycle at ZT 22 h and placed in 1.5 ml microcentrifuge tubes containing two metal beads. Samples were flash-frozen in liquid nitrogen and stored at -80°C until processing. Frozen tissue was ground in a ball mill grinder and shaken for 1 min at maximum frequency; tubes were kept frozen in liquid nitrogen and shaken for an additional minute. Total RNA was extracted using the RNeasy Plant mini kit (Qiagen) and treated with RNase-Free DNase (Qiagen). RNA samples were resuspended in 50 µl DEPC-treated water, and RNA concentration was estimated by using a Qubit RNA BR assay kit (Life Technologies, Waltham, MA) in a Qubit 2.0 fluorometer (Life Technologies) following the manufacturer's protocol.

2.2.12 cDNA synthesis and qPCR analysis

To synthesize cDNA, 250ng of total RNA were used with the SuperScript III First-Strand Synthesis SuperMix for Q-RT-PCR (Life Technologies). The cDNA was diluted four-fold with RNase-free water (Qiagen) and used for quantitative PCR using the qPCR Quantifast SYBR Green kit (Qiagen). Primers were designed with the QuantiPrime tool (Arvidsson et al., 2008) and were synthesized by FisherScientific (FisherScientific, Madison, CT). Primers are listed in Table 2.1. Primer efficiency was tested by constructing standard curves from serial dilutions of a pool of the cDNA used in the experiment. The qPCR reactions were made in triplicate using 1µl of diluted cDNA within a final volume of 20 µl. The qPCR amplification was performed in an Eppendorf Realplex2 Mastercycler under the following conditions: 95°C for 2 min, 40 cycles of 95°C for 15s, 60°C for 15s, and 68°C for 20s, followed by a melting curve analysis. Transcript

levels were determined by normalizing threshold cycle (Ct) values from target genes to the reference gene *BdACTIN7* (Hong et al., 2008*)*.

2.2.13 Library construction and sequencing

RNA quality assessment of RNA samples was performed in an Agilent 2100 Bioanalyzer (Agilent Technologies, Lexington, MA). RNA samples with an RNA integrity number above 7.0 were selected for library construction and sequencing. Seventeen mRNA libraries were created by poly-A selection and sequenced on the Illumina NovaSeq 6000 system -flow cell type S2 2 x 100bp. Library preparation and sequencing were performed at the Yale Center for Genome Analysis (YCGA Yale-West campus, Orange, CT) using the Illumina platform.

2.2.14 RNA-seq analysis

RNA-seq data were stored and processed on a Farnam server from the Yale Center for Research Computing. An assessment of the quality of the RNAseq data was performed using FastaQC (Babraham Bioinformatics). Reads were trimmed to remove adaptors and overrepresented sequences using the trimmomatic bioinformatic tool (Bolger et al., 2014). Reads were mapped to the *Brachypodium distachyon* Bd21 reference genome from Phytozome v12 (https://phytozome.jgi.doe.gov), and the annotation file was obtained from the same source. Sequencing alignment was performed by the alignment system STAR/HISAT (Dobin et al., 2013; Kim et al., 2015), and multiplex aliment was carried out with RSEM to identify transcript abundance for multiple genes (Li and Dewey, 2011). Normalization of read counts to infer

differential expression between the samples for sequence counting was performed with the R/Bioconductor package, DESeq (Anders and Huber, 2010). Genes with a p-value < 0.05 were identified to be significantly differentially expressed. Principal component analysis (PCA), scatter plots, heatmaps and hierarchical analysis diagrams were created with the programming software R (version 3.5.2), and a more stringent cut-off value was used, i.e. p-value <0.005.

2.2.15 Yeast-one-hybrid assays

2.2.15.1 Protein-DNA binding to *CESA4* Promoter and 5'UTR

The full-length coding region of *GNRF* (1005bp) was used to create a pDEST22 *GNRF* construct using the pDEST22-GAL4-AD (Invitrogen, Carlsbad, CA) destination vector. This construct was included in a sub-library collection of transcription factors suitable to test protein-DNA binding in yeast-one-hybrid assays. The pDEST22 *gnrf-1* construct was created by PCR amplification of the full-length coding region of *gnrf-1* using cDNA of *gnrf-1* mutant plants as DNA template. Primers for vector construction are listed in Table 2.1. pDEST22 *GNRF* and pDEST22 *gnrf-1* were used as pray proteins for the yeast-one-hybrid assay that tested GNRF-protein and *CESA4*-promoter binding affinity. The pDEST22 Myb-like DNA binding protein (*Bradi4g06317*) and the pDEST22 empty vector from the sub-library collection were used as negative and normalization controls respectively.

Three overlapping fragments of the *CESA4* Promoter (1,035 bp upstream of the translation start site) were cloned into pENTR/D-TOPO (Invitrogen, Life Technologies) and recombined independently into a Gateway-compatible pLUC destination vector. The pLUC construct contains a luciferase reporter and is suitable for quantitative yeast-one-hybrid screen as described

(Bonaldi et al., 2017; Pruneda-Paz et al., 2014). Primers for vector construction are listed in Table 2.1. The constructs containing the *CESA4* Promoter and 5'UTR region fragment 1 (-394 to -1) and *CESA4* Promoter fragment 2 (-747 to -323), were linearized with StuI restriction enzyme (New England Biolabs). The construct containing *CESA4* Promoter fragment 3 (-1035 to -1) was linearized with ApaI restriction enzyme (New England Biolabs). All constructs were transformed into the yeast strain YM4271. Yeast cultures were plated on SD-U medium and incubated at 30°C as described (Deplancke et al., 2004; Deplancke et al., 2006; Walhout and Vidal, 2001). Yeast baits were tested for self-luminescence activity, and non-self-active yeast colonies were used to perform the yeast-one-hybrid screen. After transformation with the prey vectors (pDEST22 *GNRF*, pDEST22 *gnrf-1*, pDEST22 *Myb-like*, and pDEST22 empty vector) colonies were grown in SD-TU selective medium (Yeast minimal media, uracil and tryptophan) at 30°C for three days. As substrate, 80 µl of native coelenterazine (Sigma-Aldrich) were added to 20 µl of the liquid yeast culture. Luminescence (relative luciferase units -RLU) was normalized by optical density (OD 600nm) and was measured in a microplate reader SpectraMax M5 (Molecular Devices, Downingtown, PA). RLU measurements were performed three times, every 5 minutes from three independent colonies per construct grown in triplicate. Fold change was presented as relative to the empty vector control. Error bars correspond to means ± SEM (Standard Error of the Mean). Two-tailed Student's *t*-tests were performed, and significance was set to $p<0.05$ and $p<0.01$.

2.2.15.2 Protein-DNA binding to *GNRF* Promoter and 5'UTR

A full-length transcription factor collection of 103 cDNAs from *B. distachyon* was cloned into pDEST22 (GAL4-AD prey vector, Invitrogen) using the Gateway cloning system (Table 2.2). This collection includes the sub-library collection mentioned briefly above that contains 27 cDNAs similarity in gene expression and protein homology to known cell wall regulators in *A. thaliana*, and 76 cDNAs encoding proteins predicted to bind DNA from a comprehensive Transcription Factor ORF Library (TORFL). Three overlapping fragments of the *GNRF* 5'UTR region (from -1,034 to -1 upstream of the translation start site) designated as fragment 1 (-304 to -1), fragment 2 (-648 to -204) and fragment 3 (-1,034 to -540) were cloned into pENTR/D-TOPO (Invitrogen, Life technologies) and subsequently recombined independently into a Gateway-compatible pLUC destination vector as described above. All constructs were linearized with StuI restriction enzyme (New England Biolabs) before yeast transformation. Non-self-active yeast colonies were transformed with all 103 constructs of the sub-library and TORFL transcription factor collections. Luciferase activity was measured in three independent yeast-one-hybrid assays, and RLU data were taken in triplicate. Four standard deviations above the average were used to identify significance.

Four additional overlapping fragments of *GNRF* promoter and 5'UTR region (from 2,255 bp to 1,035 bp upstream of the translation start site) were cloned into the pLUC destination vector. The construct containing the *GNRF* promoter region fragment 5 (-1,657 to -1,228bp) was linearized by the restriction enzyme ApaI (New England Biolabs). The constructs containing *GNRF* promoter fragment 7 (-2,255 to -1,905bp) and fragment 6 (-2,003 to -1,527bp) and *GNRF* 5'UTR region fragment 4 (-1,227 to -820bp) were linearized by restriction digestion with StuI (New England Biolabs). Yeast-one-hybrid assays were conducted as described above, and all

seven *GNRF* reporter constructs, including the three constructs used previously, were used as yeast baits. Non-self-active colonies were selected for transformation with the yeast prey constructs: pDEST22:*GNRF*, pDEST22:*SWAM1*, pDEST22:*SWAM4* and pDEST22-empty vector constructs from the sub-library collection. RLU data were measured three times, every 5 minutes from three independent colonies per construct grown in triplicate. RLU data were normalized by optical density, and fold change was presented as relative to the empty vector control. Error bars correspond to means ± SEM (Standard Error of the Mean). Two-tailed Student's *t*-tests were performed, and significance was set to $p<0.05$ and $p<0.01$.

2.2.15.3 Protein-DNA binding to *SWAM1* Promoter and 5'UTR

Three overlapping fragments of the *SWAM1* promoter (from -1320 bp to -330 bp upstream of the translation start site) and one fragment of the *SWAM1* 5'UTR region (-329 to -1) were cloned into the pLUC destination vector as described above. All constructs containing the designated fragment 1 (-329 to -1 bp), fragment 2 (-734 to -330 bp), fragment 3 (-1040 to -631 bp), and fragment 4 (-1320 to -919 bp) were linearized by restriction digestion with StuI (New England Biolabs) and transformed in yeast. After the non-self-activation test, yeast colonies were transformed with pDEST22:*GNRF*, pDEST22:*SWAM1*, pDEST22: *SWAM4* and pDEST22-empty vector constructs from the sub-library collection. Fold change relative to the vector was calculated as described above. Error bars correspond to means ± SEM (Standard Error of the Mean). Two-tailed Student's *t*-tests were performed, and significance was set to $p<0.05$ and $p<0.01$.

2.2.16 DNA Affinity Purification Sequence (DAP-seq)

DAP-seq was performed as described (Handakumbura et al., 2018), and the motif logo was generated by loading the consensus matrix into R and Bioconductor SeqLogo (version 2.10) as described (Olins et al., 2018).

2.3 Results

2.3.1 *GNRF* is the closest ortholog to *AtSND2*

GNRF belongs to the family of NAC transcription factors that comprises proteins containing the NAC domain. NAC proteins including *A. thaliana* SND1, SND2, SND3, NST1, NST2, and vascular-specific VND6 and VND7 have been identified as key transcriptional regulators of cell wall biosynthesis within a transcriptional network (Zhong et al., 2006; Zhong et al., 2008; Zhong et al., 2007b). To determine the phylogenetic relationship between GNRF and other NAC proteins, a phylogenetic tree was constructed using amino acid sequences of homologous proteins from selected eudicots and monocots (Figure 2.1). Protein sequences were aligned by using the MAFFT server and visualized with the Phylo.io application. GNRF is a grass ortholog to the eudicot *AtSND2* proteins from *A. lyrata* and *C. rubella*. *AtSND1*and *AtSND3*, secondary wall-associated proteins were removed from the tree due to a more distant relationship with GNRF compared to the protein sequences selected. NAC proteins from the eudicots *G. max* (soybean), *C*. *sinensis* (orange), *S*. *lycopersicum* (tomato), *S*. *pimpinellifolium* (currant tomato), *V*. *vinifera* (grape), *T*. *cacao* (cacao) and *P*. *trichocarpa* (poplar wood) were arranged in the eudicot clade. From the monocot plants of the subfamily Pooideae, *H. vulgare* (barley) and *T. aestivum* (wheat) have one ortholog gene closely related to *GNRF*. Three

monocot plant species *O. sativa* (rice), *S. bicolor* (sorghum), and *S. viridis* (green foxtail) were identified as having two copies of orthologous protein to GNRF, and each copy is arranged in one of two distinct grass clades. *P. virgatum* (switchgrass) has four ortholog proteins of GNRF, and they were arranged as pair of proteins in each grass clade. From the order Poales and Zingiberales, represented by *A. comosus* (pineapple) and *M. acuminata* (banana), respectively, one copy of the orthologous protein was found. These proteins were separated from the grass clusters described above and grouped as non-grass monocots.

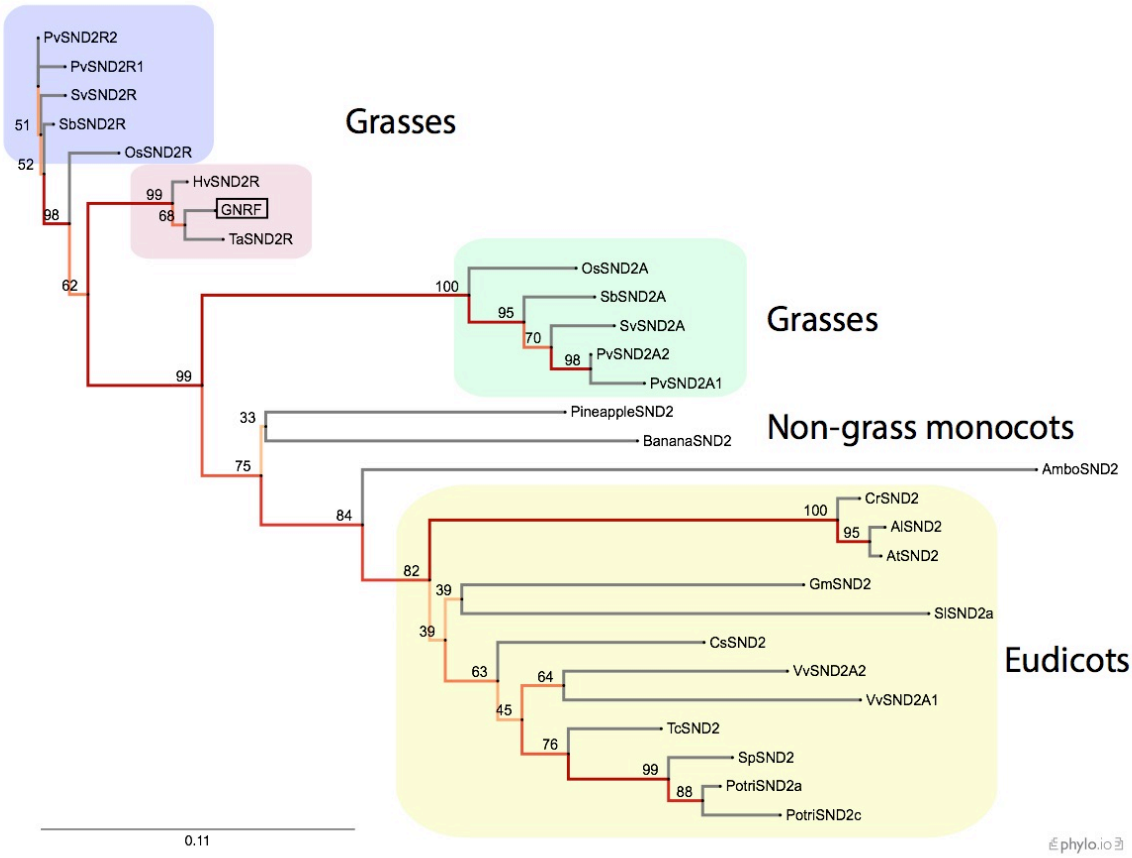


Figure 2. 1 **Phylogeny of *GNRF* (Bradi2g462197).** SND2 phylogeny illustrating amino acid sequence similarity between proteins from the eudicots *Arabidopsis thaliana*, *Arabidopsis lyrata*, *Capsella rubella*, *Glycine max*, *Citrus sinensis*, *Solanum lycopersicum*, *Solanum pimpinellifolium*, *Vitus vinifera*, *Theobroma cacao*, and *Populus trichocarpa* and the monocots *Brachypodium distachyon*, *Sorghum bicolor*, *Oryza sativa*, *Triticum aestivum*, *Hordeum vulgare*, *Setaria viridis*, *Panicum virgatum*, *Ananas comosus*, and *Musa acuminata* and the basal flowering plant *Amborella trichopoda*. Amino acid sequences were aligned with the MAFFT server for multiple sequence alignment. An iterative refinement method L-INS-I**,** neighbor-joining method and bootstrap resampling value of 1000 were applied. The Phylo.io application was used for visualization. Color scheme of the branches corresponds to the bootstrap support numbers that indicate similarity. Three grass clades (purple, pink and green), one eudicot clade (yellow) and two non-grass monocot proteins are indicated.

2.3.2 *GNRF* is relatively abundant in stem tissue, and it is co-expressed with other cell-wall-associated genes

GNRF was selected for functional characterization based on amino acid homology to the known cell wall regulator *AtSND2*. Overexpression of *AtSND2* increased secondary wall thickening in fibers and induced the expression of cellulose biosynthetic genes. On the contrary, a dominant repressor version of *AtSND2* caused significant reduction in the thickness of fiber cell walls in *A. thaliana* (Hussey et al., 2011a; Zhong et al., 2008). Data from a whole-genome tiling microarray reported previously (Handakumbura et al., 2013) were used to determine *GNRF* transcript abundance in leaf, root, and stem tissues. *GNRF* gene expression in the stem was approximately nine-fold and three-fold higher than leaf and root, respectively (Figure 2.2A). Similar to *GNRF*, *SWAM1* (SECONDARY WALL ASSOCIATED MYB, Bradi2g47590), a positive regulator of secondary cell wall thickening in *B. distachyon* (Handakumbura et al., 2018), showed a relative transcript abundance in stems six and three-fold greater than in leaf and root, respectively (Figure 2.2B). Consistent with this, the stem is the organ where secondary cell wall formation for the vascular system and fiber development occurs in specialized tissues such interfascicular fibers and xylary fibers (Matos et al., 2013). Additionally, ectopic deposition of secondary cell walls and reduced wall-thickening in stems have been reported due to perturbations of transcriptional regulators of cellulose, xylan and lignin biosynthetic genes (Rao and Dixon, 2018; Zhong et al., 2018).

NAC and MYB transcription factors have been identified as part of the regulatory structure within a transcriptional network associated with secondary cell wall formation (Nakano

et al., 2015). Comparative genomic analysis in *A. thaliana* revealed that SND (Secondary NAC Domain) and MYB proteins were co-regulated to modulate cell wall biosynthetic genes in a network (Yao et al., 2012). To elucidate if *GNRF* and *SWAM1* function as part of a transcriptional network in *B. distachyon*, the PlaNet database (http://aranet.mpimp-golm.mpg.de) was used to generate a co-expression network. *GNRF* (Bradi2g46197) and *SWAM1* (Bradi2g47590) were identified in the same co-expression network along with cell-wall-related genes (Figure 2.2C, Table 2.1).

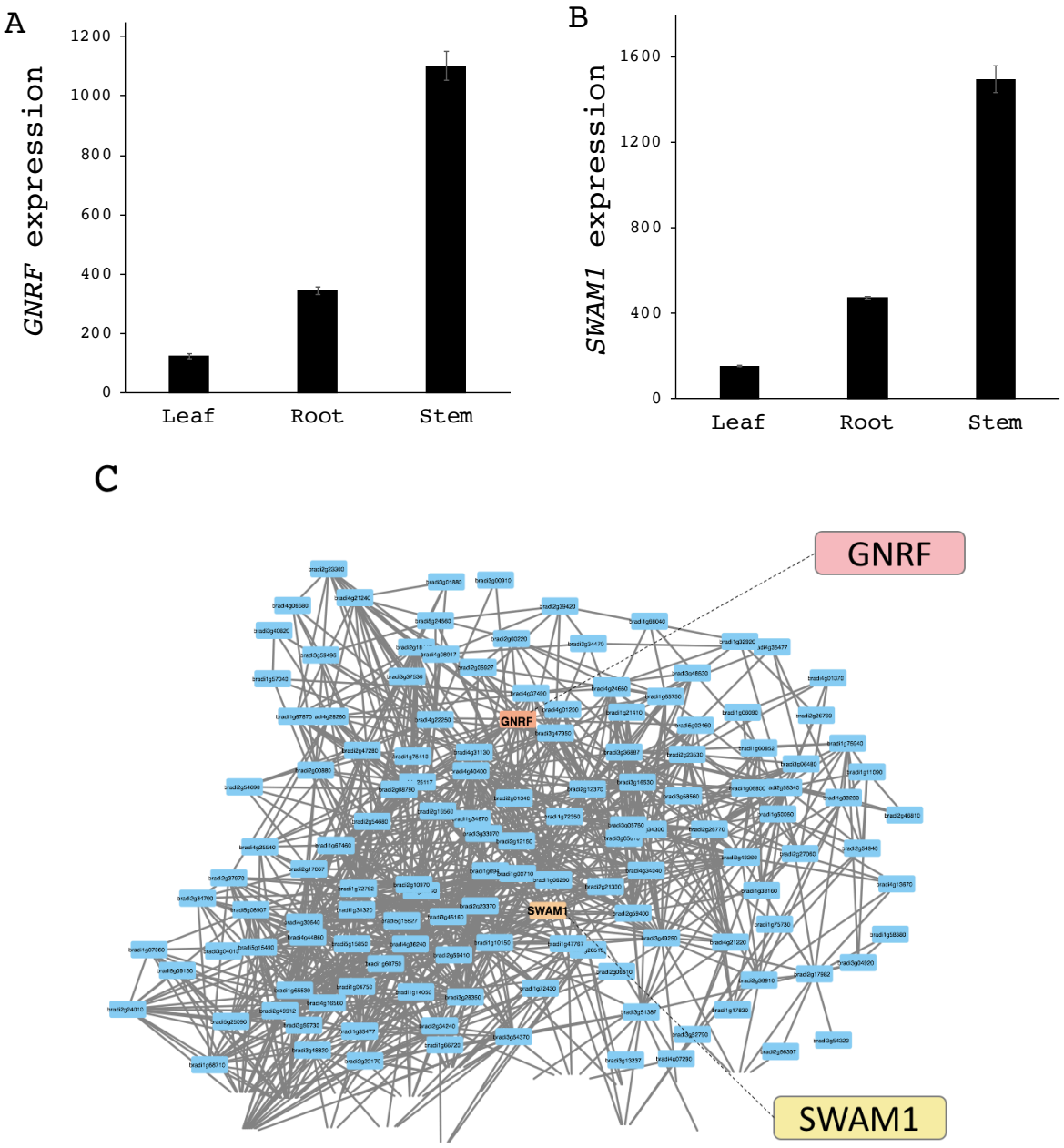


Figure 2. 2 **Relative expression and co-expression analysis of *GNRF* and *SWAM1*.** Relative transcript abundance of *GNRF* (A) and *SWAM1* (B) in leaf, root and stem tissue in a whole-genome tiling microarray. Error bars indicated SEM (Standard Error of the Mean) of three biological replicates. (C) PlaNet co-expression neighborhood of *GNRF* and SWAM1.

2.3.3 *GNRF* overexpression resulted in persistent vegetative growth.

To investigate the function of *GNRF*, *GNRF*-overexpressing plants were created by cloning the full-length coding region of *GNRF* downstream of the strong *ZmUbi* promoter in pOL001(Vogel et al., 2006). Plant transformation was performed into *B. distachyon* via the *Agrobacterium tumefaciens* calli method (Vogel and Hill, 2008). *GNRF*-overexpressed (*GNRF-OE*) plants remained vegetative until senescence and exhibited a bushy phenotype (Figure 2.3A-B, Figure 2.5). While *GNRF-OE* plants, were significantly shorter than control plants, the number of stem internodes was three-fold higher than the number of internodes in control plants (Figure 2.3C-E). When the inflorescence emerged from the flag leaf in wildtype plants *GNRF-OE* plants were short and had few leaves (Figure 2.5A); however, *GNRF-OE* continued growing and dramatically increasing its total plant biomass (Figure 2.6C). *GNRF-OE* stems were developed, but they remained short; hence, these plants were significantly shorter than wildtype (Figure 2.6B). *GNRF-OE* remained vegetative after 120 days after germination (DAG) (Figure 2.5B). The flowering time was dramatically increased (Figure 2.6A); nonetheless, a few plants ultimately flowered, and the genotype was able to propagate.

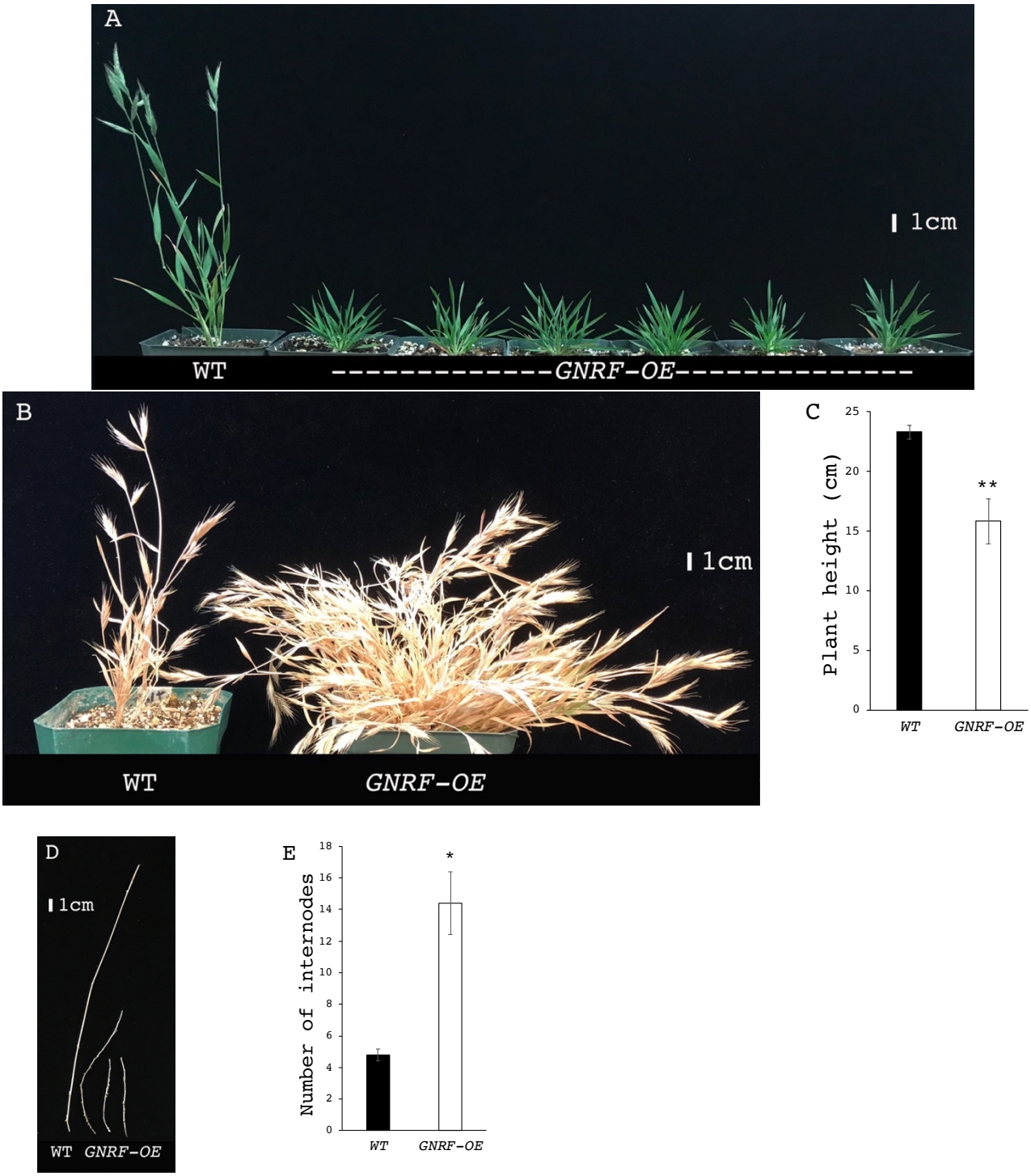


Figure 2. 3 **Overexpression of *GNRF* results in persistent vegetative growth**. (A) While WT plants were at reproductive stage, *GNRF-OE* plants remained at the juvenile stage. The image was captured 26 days after germination. (B-E) WT and bushy *GNRF-OE* plants >150 days after germination (DAG) at senesced stage. (C) Plant height of WT and *GNRF-OE* plants at fully mature stage. (D) Nodes and internodes of stems. (E) Number of internodes. Data are presented as means ± SEM. n=4. Two-tailed Student's t-tests were performed *p<0.05, **p<0.01.

2.3.4 Some *gnrf* mutant plants were late-flowering.

Five homozygous lines from a TILLING (Targeting Induced Local Lesion IN Genome) collection mutagenized with sodium azide (NaN_3) (Dalmais et al., 2013) were genotyped to identify point mutations in *GNRF*. The *gnrf-1* allele was identified as having a non-synonymous mutation that replaces a proline (P) with a leucine (L) in position 50 of the protein. The *gnrf-2*, *gnrf-3*, and *gnrf-5* mutant alleles presented non-synonymous mutations that replace an alanine (A) with a threonine (T) in the amino acid positions 12, 39, and 261, respectively. The *gnrf-4* allele replaces a glutamine (Q) with a TAG stop codon in the amino acid position 57 of GNRF protein (Figure 2.4). Only homozygous lines were used for further analysis.

The mutant allele *gnrf-6* was identified as carrying a T-DNA insertion 317 bp upstream of the translation start site and within the *GNRF* 5'UTR (Figure 2.4). *gnrf-6* homozygous and heterozygous lines were isolated, but only the homozygous line was used to evaluate *GNRF* function (Figure 2.4).

None of the *gnrf* mutations caused early flowering. On the contrary, *gnrf-2*, *gnrf-5* and *gnrf-6* were later flowering (Figure 2.6A). Plant height was significantly reduced in *gnrf-1*, *gnrf-4*, *gnrf-5*, and *gnrf*-6, and only *gnrf-2* was significantly taller than the wild type (Figure 2.6B). Total biomass was higher in *gnrf-1*, *gnrf-2*, *gnrf-3*, and *gnrf-6* (Figure 2.6C). Overall, all of the *gnrf* mutants showed overlapping changes in their phenotype related to the rate of growth, amount of growth and time to flower.

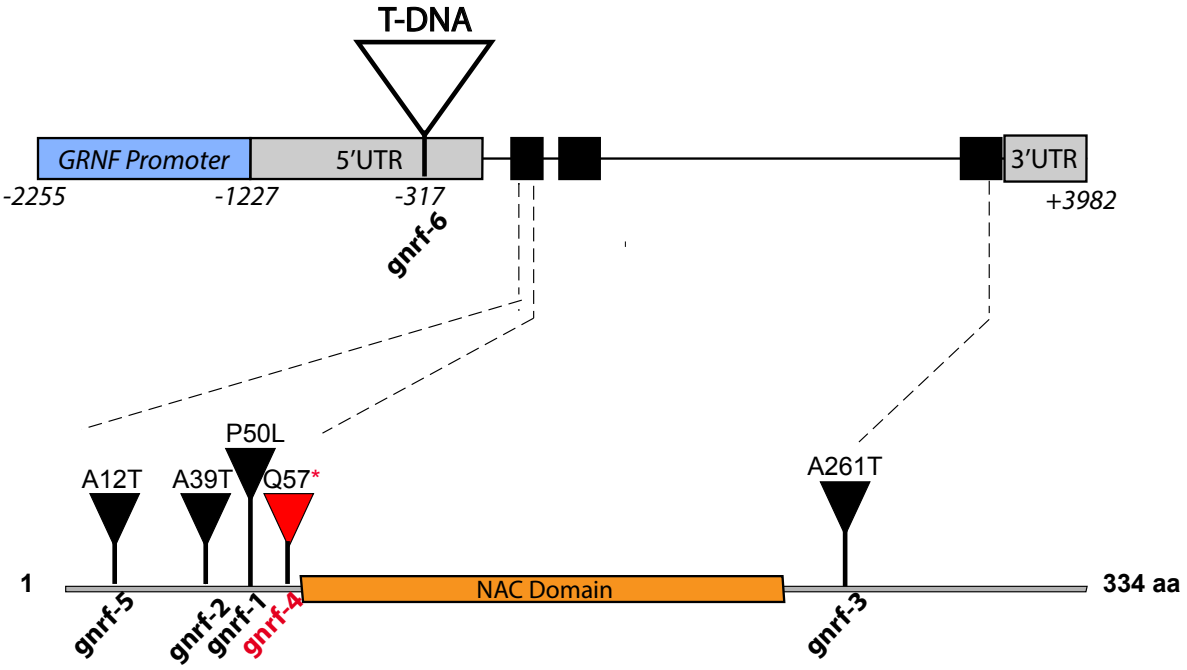


Figure 2. 4 ***gnrf* mutant alleles**. (Top) Diagram showing the position of the T-DNA insertion located -317 bp upstream of the translation start site in the 5'UTR region. Black boxes represent exons, and the white triangle a T-DNA insertion. (Bottom) Diagram of GNRF protein with the location of the *gnrf* mutations. Dashed lines show the position of the mutations within the GNRF coding region. Black triangles represent the position of each non-synonymous mutation. A, alanine; T, threonine; P, proline; Q, glutamine; * stop codon; aa, amino acids.

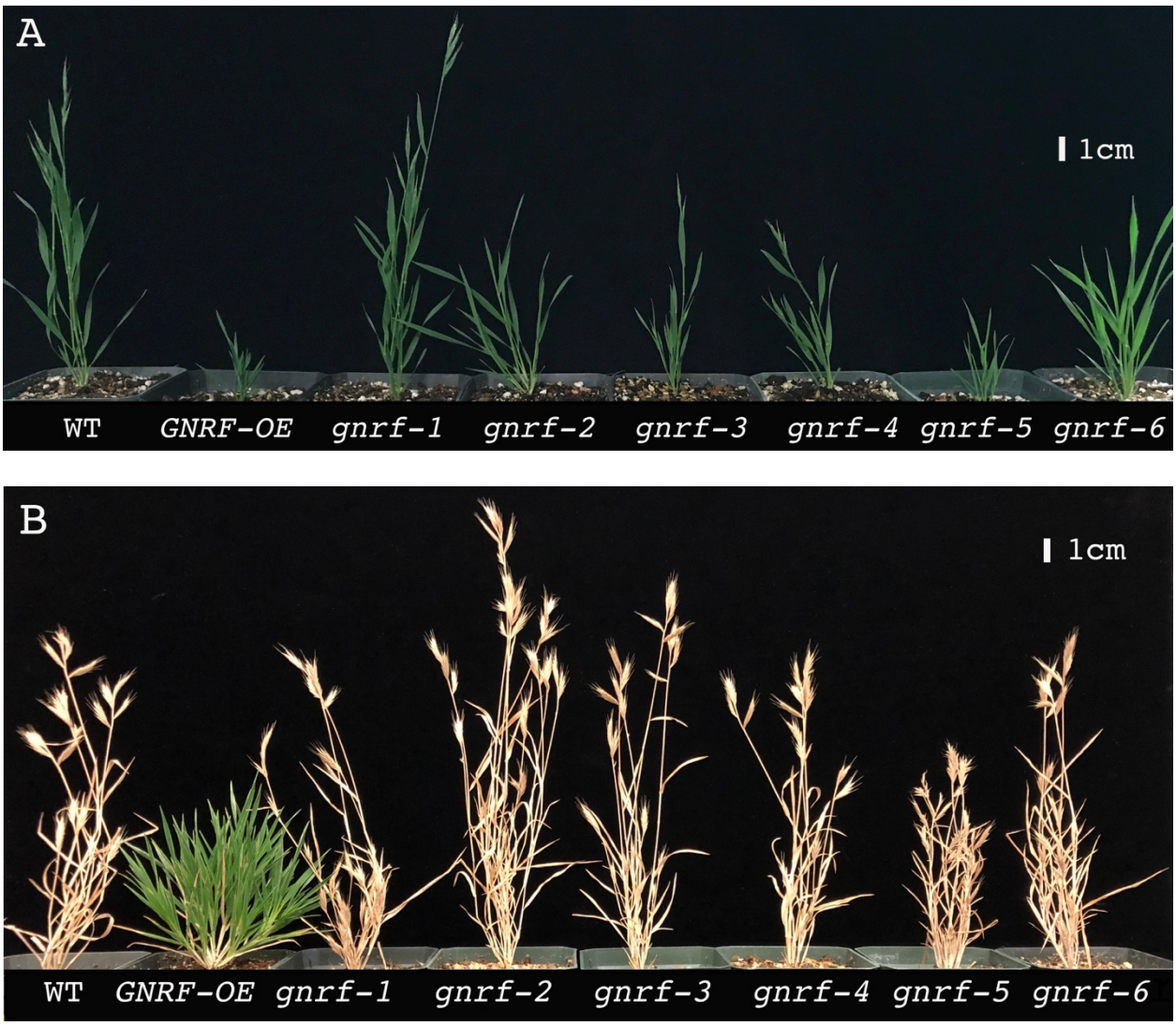


Figure 2. 5 **Plant phenotypes of WT, *GNRF* overexpression and *gnrf* mutant plants.** (A) Plants are from the juvenile to early reproductive stage. The image was captured 26 DAG. (B) Plant were photographed when completely senesced, with the exception of GNRF-OE that remains in a vegetative stage. The image was captured 120 DAG.

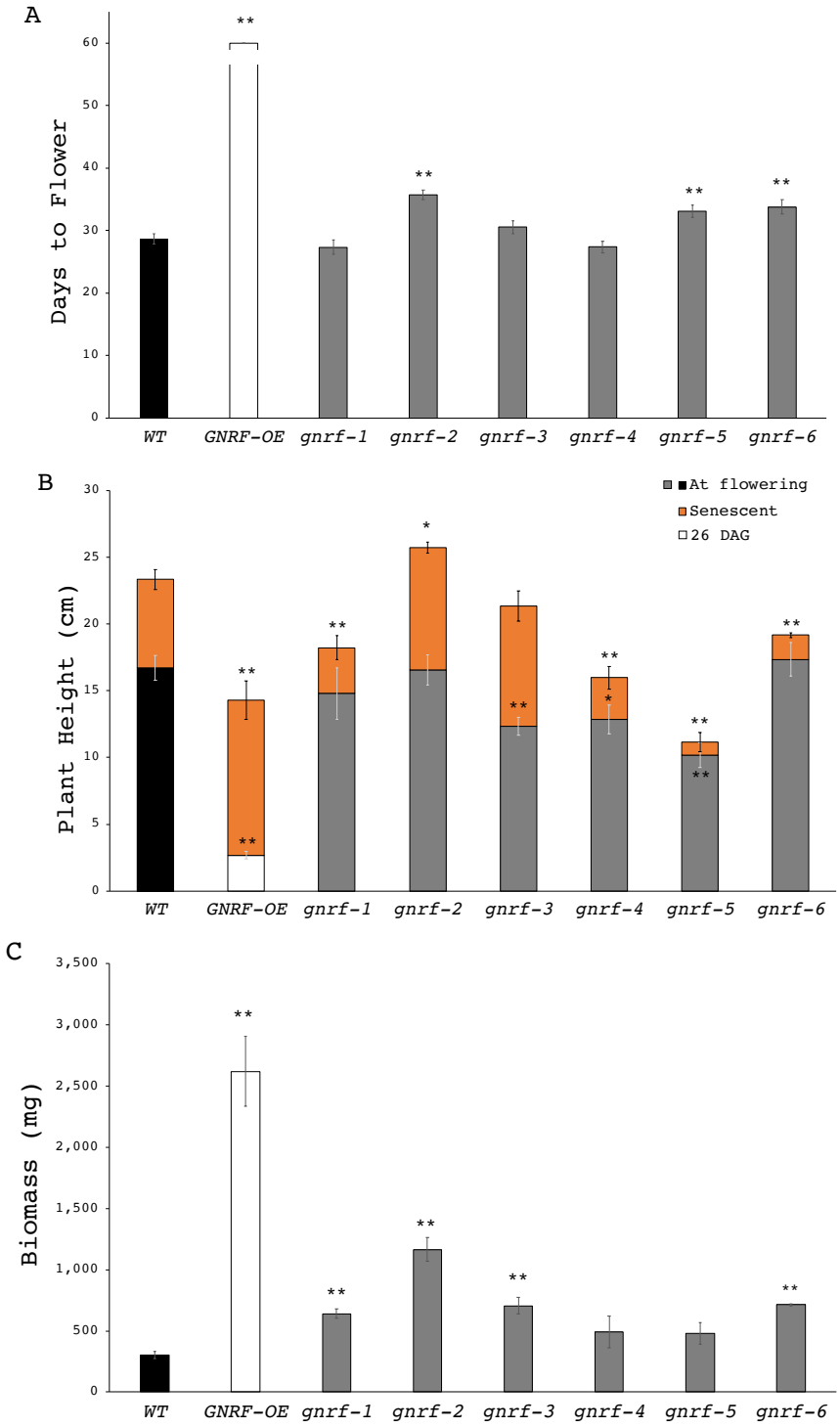


Figure 2. 6 **Whole plant phenotypes.** (A) Days taken to begin of heading, when the inflorescence emerges from the flag leaf. n=5-11 per genotype. *GNRF-OE* plants flowered after fertilizer application at >90 DAG. (B) Plant height at 23 DAG (white), at flowering (black, grey) and at senescent stage (orange) n=3-7. (C) Total above-biomass at complete senescence, n=3-4 plants per genotype. Means ± SEM are shown. Two-tailed Student's t-tests were performed *p<0.05, **p<0.01.

2.3.5 *GNRF* is a repressor of lignin biosynthesis

To identify a relationship between *GNRF* and cell wall formation like its ortholog At*SND2* (Hussey et al., 2011a; Zhong et al., 2008) and its co-expressed gene *SWAM1* (Handakumbura et al. 2018), *GNRF-OE* and *gnrf* mutants fully senesced stems of *GNRF-OE* and *gnrf* mutants were sectioned and stained with phloroglucinol- HCl (Figure 2.7). The qualitative estimation of lignin deposition by the stain showed a reduction of lignin by the lighter staining of the walls of interfascicular fibers and vascular bundles of *GNRF-OE* compared to the WT (Figure 2.7D-F). No visible reduction of lignin was found in vascular bundles of *gnrf-1*, *gnrf-2*, *gnrf-3*, and *gnrf-6*; however, all *gnrf* mutants showed a lighter staining pattern in walls of interfascicular fibers (Figure 2.7G-X). *gnrf-4* and *gnrf-5* showed the lightest staining among the mutants in both vascular bundles and interfascicular fibers, and *gnrf-5* exhibited an atypical stem shape (Figure 2.7S-U). Stem tissue of all genotypes at complete senescence was pulverized and used for lignin quantification with the acetyl bromide soluble lignin (ABSL) method. Consistent with the staining results, lignin content was significantly reduced in *GNRF-OE* plants as measured by the ABSL method. The *gnrf* mutants seemed to have a reduction in total lignin, except for *gnrf-6*; however, none were significantly different. To further investigate the function of *GNRF* in cell shape and cell wall formation, the area and perimeter of stems were measured. *GNRF-OE*, *gnrf-4*, and *gnrf-5* had a smaller area (Figure 2.8A); *gnrf-4* and *gnrf-5* showed a reduction in the perimeter as well (Figure 2.8B). Measurements of cell wall thickness revealed that *GNRF-OE* had significantly thinner walls of interfascicular fibers (Figure 2.8C). Xylem wall thickness was reduced in all genotypes but was only significantly different from the WT in *gnrf-5* plants.

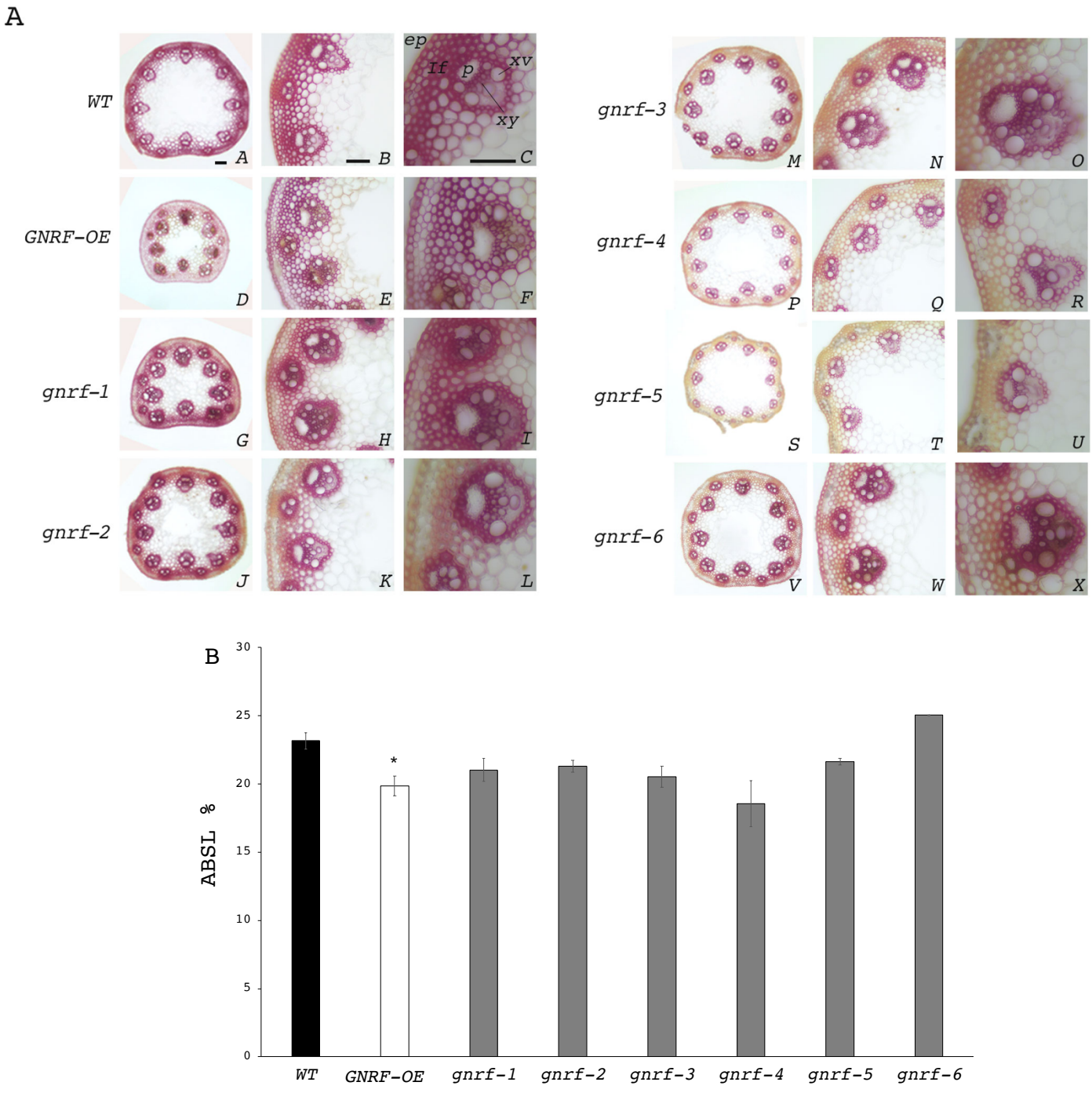


Figure 2. 7 **Histo-chemical and quantitative analysis for lignin composition in stems.** (A) Cross sections of the first internode of fully mature plants of WT(A-C), *GNRF-OE* (D-F), *gnrf-1* (G-I), *gnrf-2*(J-L), *gnrf-3* (M-O), *gnrf-4* (P-R), *gnrf-5* (S-U), and *gnrf-6* (V-X) stained with phloroglucinol-HCl. The images were captured using 4x, 10X and 20x objectives (left to right). ep, epidermis; if, interfascicular fibers; p, phloem; xv, xylem vessels; xy, xylem. Scale bar = 0.1 mm. (B) Acetyl bromide soluble lignin content of completely senesced stem tissue was compared between WT and *GNRF-OE* and *gnrf* mutants. Three plants per genotype were analyzed and samples were tested in triplicate. Means ± SEM are shown. Two-tailed Student's t-tests were performed *$p<0.05$

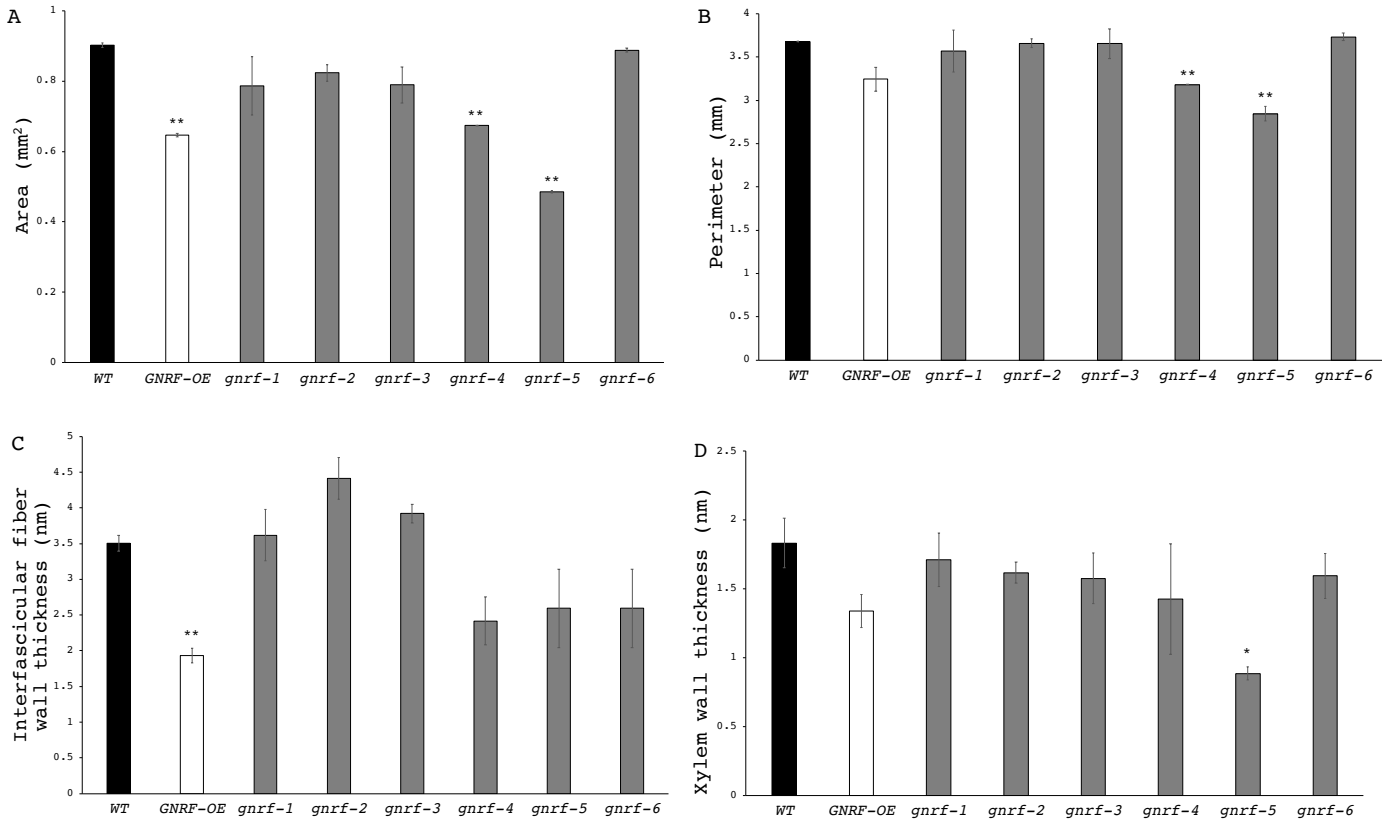


Figure 2. 8 **Stem and cell wall measurements of cross sections from the first internode of fully senesced plants**. Area (A) and perimeter (B) of WT, *gnrf-1*, *gnrf-2*, *gnrf-3, gnrf-4*, *gnrf-5*, *gnrf-6* plants are shown. n=3 for each genotype. Wall thickness of interfascicular fibers (C) and xylem cells (D) of three independent sections per genotype. n=15. Means ± SEM. Two-tailed Student's t-tests were performed *p<0.05, **p<0.01

2.3.6 *GNRF* is a repressor of cellulose synthase genes in rice protoplasts.

Overexpression of *AtSND2* increased secondary wall thickness and induced the expression of cellulose biosynthetic genes (Zhong et al., 2008). *OsSND2A* (Os05g48850), grouped in the green cluster (Figure 2.1), were also shown to increase the relative expression of rice cellulose biosynthetic genes *OsCESA4*, *OsCESA7*, and *OsCESA9* and cellulose content when *OsSND2* was overexpressed (Ye et al., 2018). To investigate the function of *GNRF* in

cellulose biosynthesis, expression of secondary wall cellulose synthases *OsCESA4*, *OsCESA7* and *OsCESA9* was measured in rice protoplasts overexpressing *GNRF* and *gnrf-1*. Contrary to the positive regulation by *AtSND2* and *OsSND2A* of cellulose biosynthetic genes, *GNRF* resulted in a significant reduction of *OsCESA4* and *OsCESA9* expression (Figure 2.9). Additionally, the non-synonymous mutation of *GNRF* coding region in g*nrf-1* interfered with the repression of cellulose genes in rice.

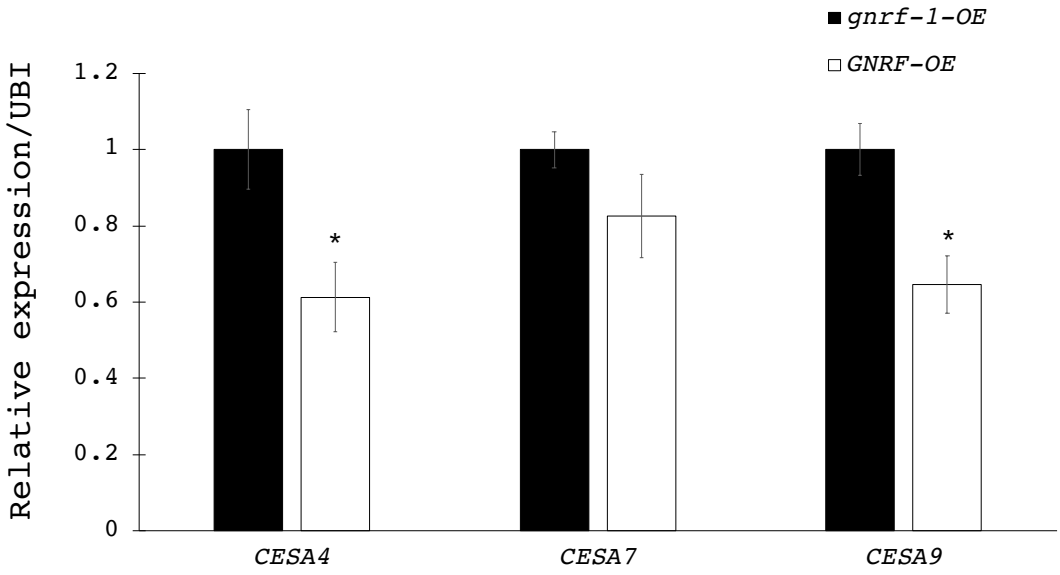


Figure 2. 9 **Relative expression of *CESA* genes in rice protoplasts overexpressing *GNRF* and *gnrf-1*.** Rice protoplasts were transformed with *35S:GNRF* and *35S:gnrf-1* constructs. Co-transformation with *Ubi:GUS* construct was used for an internal control. Q-RT-PCR was performed for four independent transformations and four technical replicates. Data were normalized to ubiquitin transcript expression. Means ± SEM are shown. Two-tailed Student's t-tests were performed *$p<0.05$

2.3.7 *GNRF* is a repressor of *SWAM1* and cell wall biosynthetic genes.

SWAM1 is a positive regulator of secondary cell wall biosynthesis (Handakumbura et al., 2018) and is co-expressed with *GNRF*. To investigate the *GNRF* and *SWAM1* genetic relationship, quantitative real-time PCR was used to measure transcript abundance of *GNRF* and

SWAM1 in WT and *GNRF-OE* stems. While *GNRF* transcript abundance was four-fold higher in *GNRF-OE* as expected, *SWAM1* transcript abundance was significantly reduced (Figure 2.10).

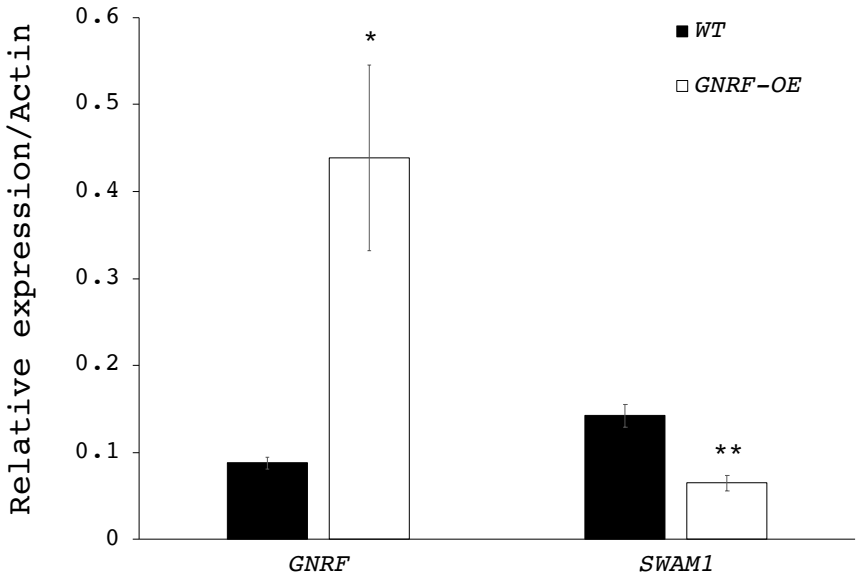


Figure 2. 10 **Transcript abundance of *GNRF* and *SWAM1* in WT and *GNRF-OE* plants.** RNA samples from plants grown under diurnal temperature and light cycles (LDHC; 20°C with 20 h light: 4 h dark) were used to perform Q-RT-PCR analysis. Data were normalized to expression of the actin housekeeping gene, and six individuals were analyzed in triplicate. Means ± SEM. Two-tailed Student's t-tests were performed *$p<0.05$, **$p<0.01$.

To identify whether *SWAM1* gene expression differences were a result of the overexpression of *GNRF* or from comparing reproductive and juvenile stages, WT, *GNRF-OE*, *gnrf-1*, and *gnrf-2* plants were grown in short days where all of the plants were in a vegetative phase (Figure 2.11). *GNRF* transcript abundance was thirteen-fold greater in *GNRF-OE* and significantly higher in *gnrf-2* compared to the WT. (Figure 2.12A). *SWAM1* transcript abundance was significantly reduced in *GNRF-OE* and significantly more abundant in *gnrf-1* and *gnrf-2* mutant plants compared to the WT (Figure 2.12B).

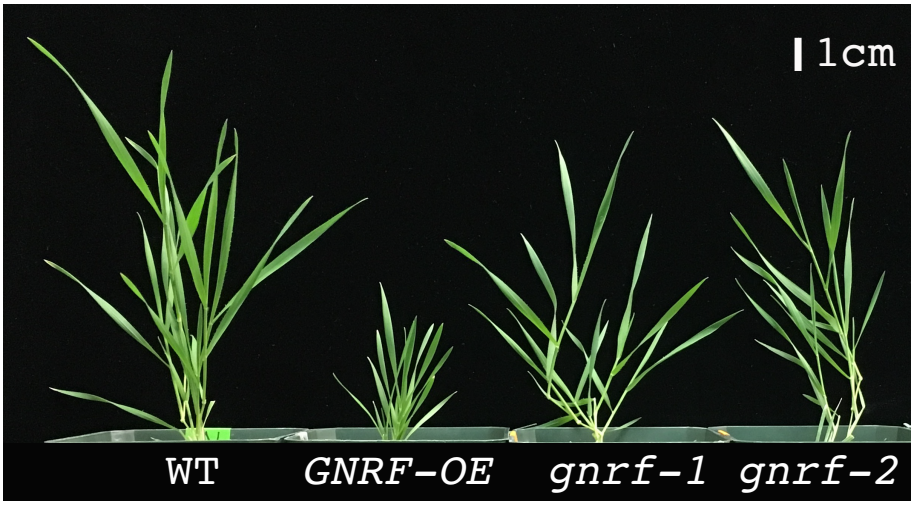


Figure 2. 11 **Plant phenotypes of WT, *GNRF-OE*, *gnrf-1* and *gnrf-2* mutant plants grown to developmentally equivalent stages.** Plants were grown at short-day cycle of 10-hours light/14-hours dark at 26°C and 18°C, respectively. The image was captured when plants were 30 days old.

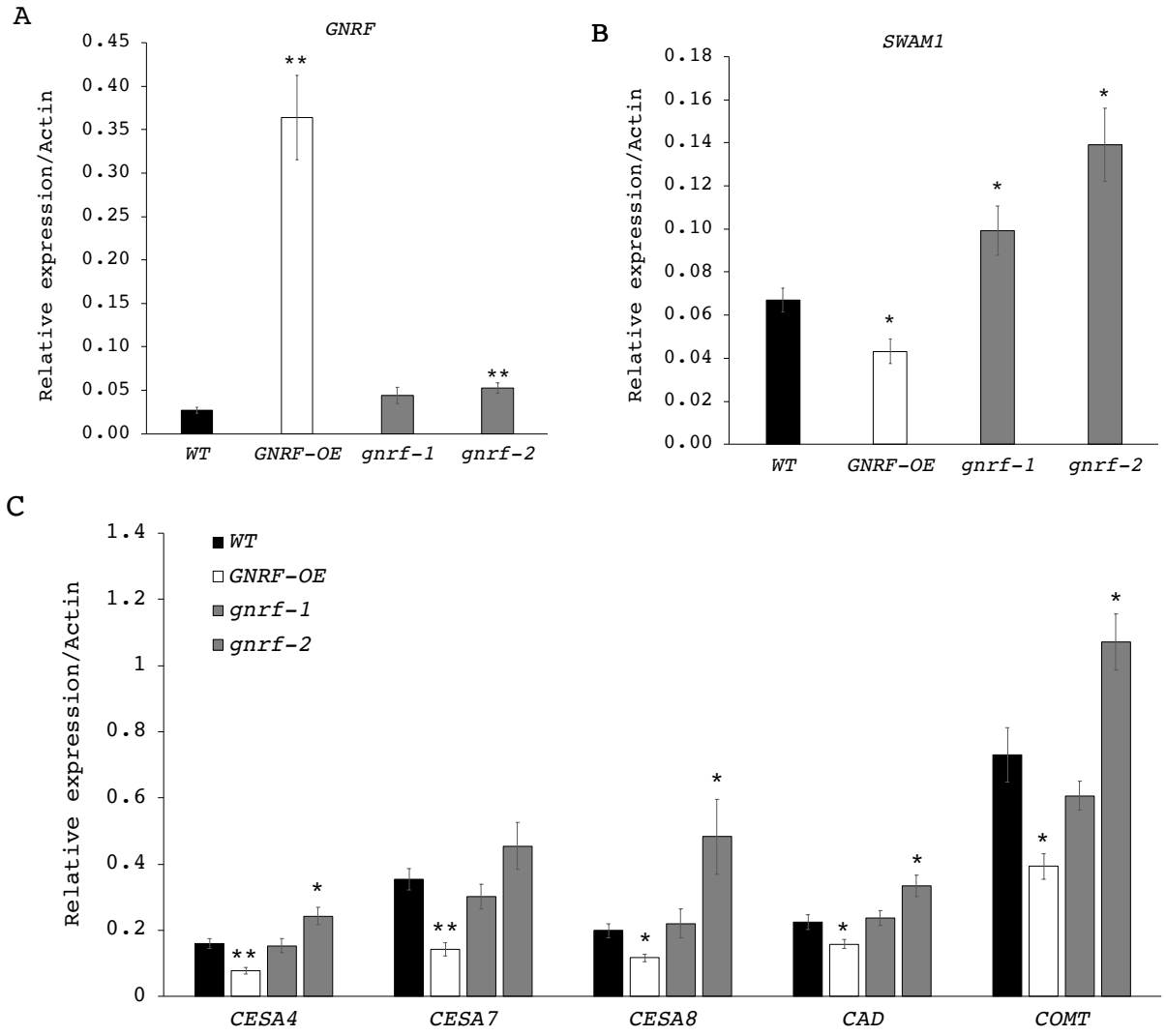


Figure 2. 12 **Transcript abundance of *GNRF, SWAM1* and cell wall biosynthetic genes.** Q-RT-PCR was performed in WT, *GNRF-OE*, *gnrf-1*, and *gnrf-2* plants grown under short days (developmentally equivalent stages). Gene expression levels of *GNRF* (A), *SWAM1* (B), and cell wall biosynthetic genes, *CESA4*, *CESA7*, *CESA8*, *CAD* and *COMT*, were measured (C). Data were normalized to expression of the actin housekeeping gene, and five to three individuals per genotype were analyzed in triplicate. Means ± SEM are shown. Two-tailed Student's t-tests were performed *$p<0.05$, **$p<0.01$.

SWAM1 regulates expression of the cellulose biosynthetic genes *CESA4*, *CESA7*, *CESA8*, the lignin biosynthetic genes *CAD1* and *COMT6*, and the hemicellulose biosynthetic gene *GT47-1* in *B. distachyon* (Handakumbura et al., 2018). To explore the role of *GNRF* in secondary cell wall biosynthesis in *B. distachyon*, gene expression analysis data were obtained by Q-RT-PCR.

Stem samples of WT, *GNRF-OE*, *gnrf-1* and *gnrf-2* plants grown to an equivalent developmental stage were used to test relative expression of *CESA4*, *CESA7*, *CESA8*, *CAD1*, and *COMT6*. All genes were significantly downregulated in *GNRF-OE* and upregulated in *gnrf-2* compared to the WT. No effect was observed in the *gnrf-1* mutant.

2.3.8 GNRF protein binds to the *CESA4* promoter in yeast.

To identify if *GNRF* is a putative direct regulator of the cellulose synthase gene *CESA4*, a protein-DNA interaction analysis with GNRF protein, gnrf-1 protein and three overlapping fragments of *CESA4* promoter was performed by using a luciferase-compatible yeast-one-hybrid assay. GNRF protein bound to all three fragments of the *CESA4* promoter. Nine-fold and two-fold change increases of luciferase activity relative to empty vector control were observed in yeast binding assays of *CESA4* promoter fragment 3 and GNRF and gnrf-1, respectively. Between a three- and two-fold change was obtained with GNRF and gnrf-1 in fragments 1 and 2 (Figure 2.13).

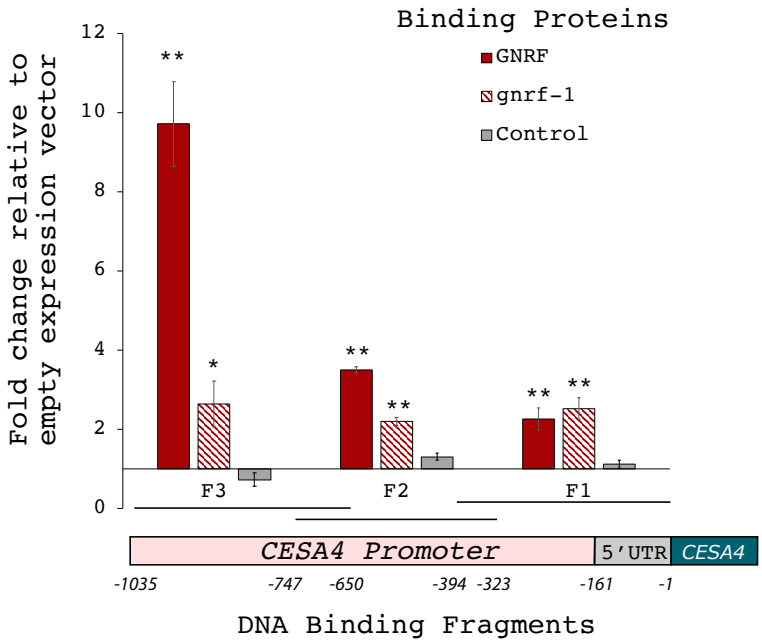


Figure 2. 13 **GNRF binding to *CESA4* promoter and 5'UTR regions in yeast.** (Top) Protein-DNA interaction of GNRF and three overlapping fragments of the *CESA4* promoter (-1035 to -1). Fragment 3 includes the 5'UTR region (-161 to -1). Binding was measured by quantitative yeast-one-hybrid assay using a luciferase reporter construct. (Bottom) Schematic representation of *CESA4* promoter and 5'UTR overlapping fragments. Relative luciferase units (RLU) were measured from three independent yeast colonies grown in triplicate and treated with the substrate (coelenterazine). Luminescence was measured five minutes apart at three times in the plate reader. Fold change was presented as relative to empty vector control. Means ± SEM. Two-tailed Student's t-tests were performed *$p<0.05$.

2.3.9 SWAM1 and SWAM4 proteins, among other proteins, bind to the *GNRF* promoter and the *GNRF* 5'UTR region in a yeast-one-hybrid assay.

To find putative regulators of *GNRF* and to investigate its association with other transcription factors, a luciferase-compatible yeast-one-hybrid assay was used to test protein-DNA interaction. One-hundred and three transcription factor proteins and three overlapping DNA fragments containing 1,036 bp upstream from the translation start site of *GNRF*, were tested for binding. These DNA fragments corresponded to part of the partial fraction of the

GNRF 5'UTR. Nineteen transcription factors bound significantly to these fragments in yeast, including six MYBs, three bHLHs, and two Homeobox proteins (Figure 2.14A). Based on these results, SWAM1, SWAM4 and GNRF proteins were selected to test protein-DNA interaction with seven overlapping fragments of the *GNRF* promoter (-2255 to -1228) and the complete *GNRF* 5'UTR region (-1227 to -1). All three proteins bound to the *GNRF* promoter and 5'UTR fragments 3, 4 and 5 significantly; SWAM4 and SWAM1 bound fragment 2, and SWAM1 bound to fragment 7 (Figure 2.14B). Additional yeast-one-hybrid assays were conducted to elucidate protein-DNA interactions between SWAM1, SWAM4 and GNRF proteins and *SWAM4* and *SWAM1* promoters and 5'UTR regions. All proteins bound to the 5'UTR region of *SWAM1*, significantly and only SWAM1 bound its promoter region in fragment 2 (Figure 2.15A). Similarly, all three of these proteins bound the 5'UTR region of *SWAM4* and fragment 2 of *SWAM4* promoter (Figure 2.15B).

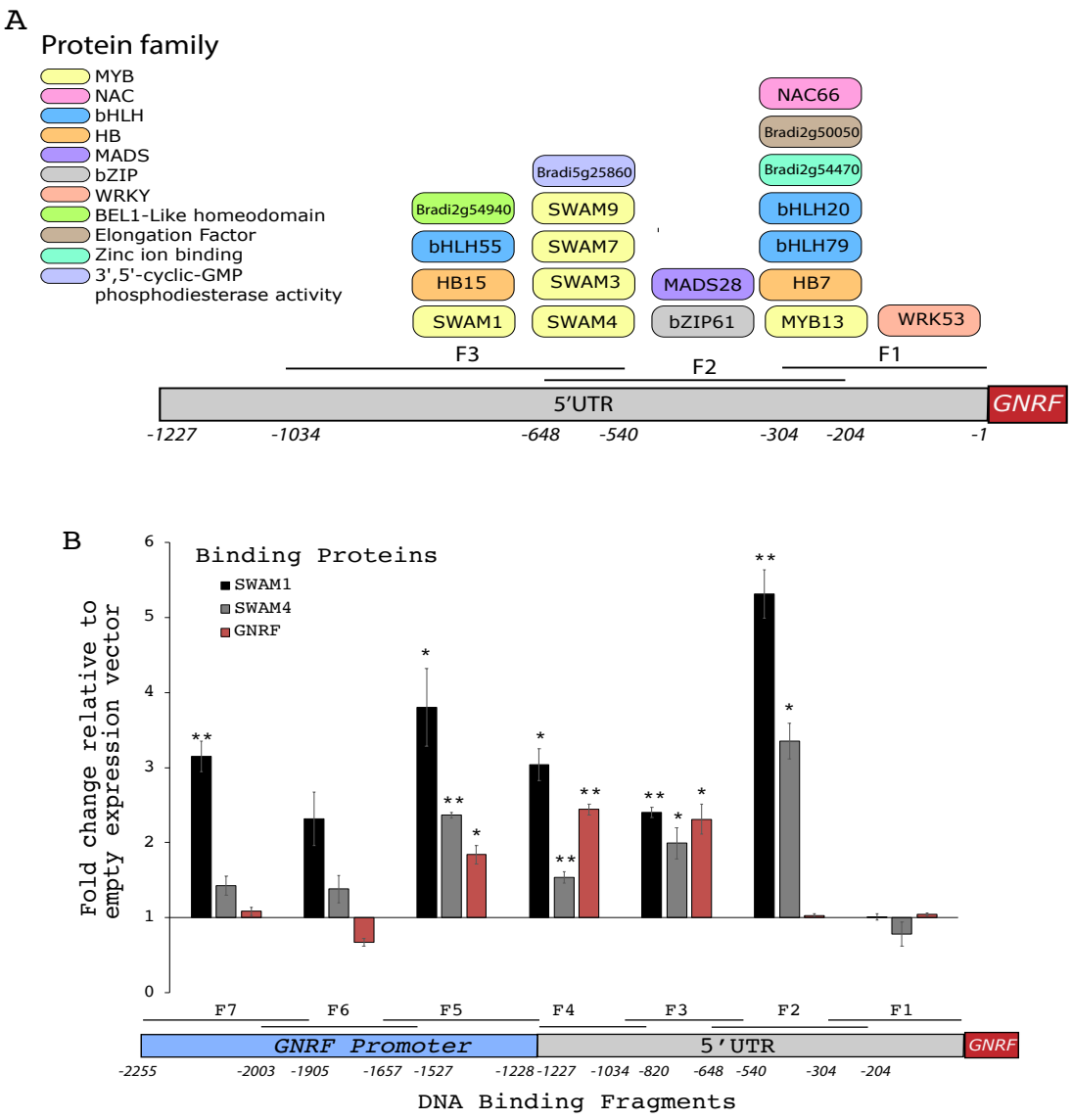


Figure 2. 14 **Transcription factor binding to the *GNRF* promoter region and 5'UTR region in yeast.** Protein-DNA interaction was measured by a quantitative yeast-one-hybrid assay using a luciferase reporter construct. (A) Shown is a schematic representation of nineteen proteins found to bind to three overlapping fragments (F1-F3) of a region of the 5'UTR of *GNRF* (-1034 to -1). Different protein families are represented by different colors. Luciferase activity was measured in three independent yeast-one-hybrid assays and RLU were normalized by optical density (OD 600 nm) in triplicate. Four standard deviations above the average were used to identify interaction. (B) SWAM1, SWAM4 and GNRF proteins bound to four overlapping fragments (F1-F4) of the 5'UTR region of *GNRF* (-1227 to -1) and three overlapping fragments (F5-F7) of the *GNRF* promoter region (-2255 to -1228). A schematic representation of the fragments is shown at the bottom. RLU were measured from three independent yeast colonies grown in triplicate and treated with the substrate (coelenterazine). Luminescence was measured five minutes apart at three times in the plate reader. Fold change was presented as relative to empty vector control. Means ± SEM are shown. Two-tailed Student's t-tests were performed *p<0.05, **p<0.01.

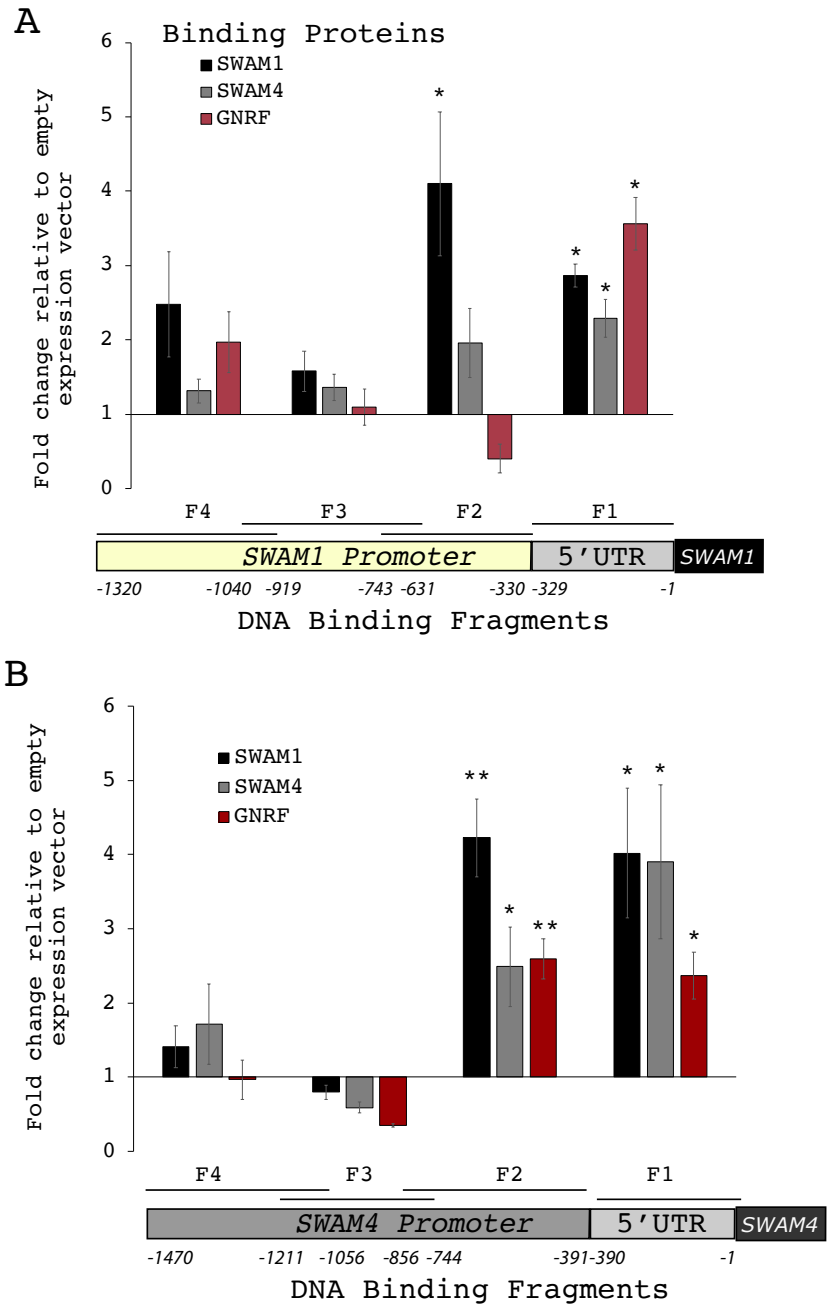


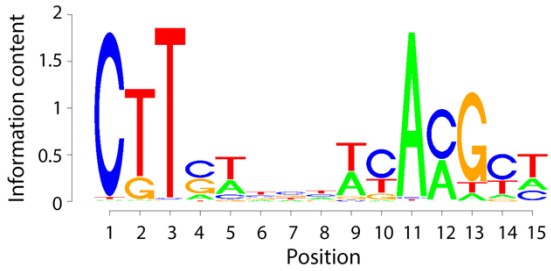
Figure 2. 15 **Protein-DNA binding to *SWAM1* and *SWAM4* promoter regions and 5'UTR region in yeast.** (A-B) Top) Protein-DNA interaction measured by quantitative yeast-one-hybrid assay using a luciferase reporter construct. SWAM1, SWAM4 and GNRF were used as prey proteins (Bottom) Schematic representation of *SWAM1* (A) and *SWAM4* (B) promoter and 5'UTR binding fragments. (A) 5'UTR region of *SWAM1* F1(-329 to -1) and the *SWAM1* promoter region F2-F4 (-330 to -1320) (B) 5'UTR region of *SWAM4* F1(-390 to -1) and the *SWAM1* promoter region F2-F4 (-391 to -1470). RLU were measured from three independent yeast colonies grown in triplicate and treated with the substrate (coelenterazine). Luminescence was measured five minutes apart at three times in the plate reader. Fold change was presented as relative to empty vector control. Means ± SEM are shown. Two-tailed Student's t-tests were performed *p<0.05.

2.3.10 GNRF binding site was identified by DNA affinity purification sequencing (DAP-seq).

Transcription factors can regulate gene expression through their interaction with *cis*-elements localized in the promoter regions of their target genes. Functional genomic and genome-wide analysis of the NAC genes *SND1*, *NST1*, *NST2*, *VND6*, and *VND7*, collectively grouped as secondary wall NACs (SWNs), has shown that the corresponding proteins interact with the 11-bp TERE motif (Tracheary Element Regulating *cis*-Element) and the 19-bp SNBE (Secondary Wall NAC Binding Element) in genes functioning in cell wall formation (Nakano et al., 2015; Pyo et al., 2007; Raes et al., 2003; Zhong and Ye, 2012). DNA affinity purification sequencing (DAP-seq) was used to identify a core motif VNS (VND/NST/SND) element C(G/T)TNNNNNNNA(A/C)G across all NAC proteins (Olins et al., 2018). Similarly, MYB binding elements of transcription factors regulating cell wall biosynthesis were identified as AC-type promoter elements (Handakumbura et al., 2018; Prouse and Campbell, 2012; Zhong and Ye, 2012). A MYB binding element ACC(A/T)A(A/C)(T/C) motif was designated as the SMRE (secondary wall MYB responsive element) (Zhong and Ye, 2012). The DAP-seq GNRF consensus binding site was (CT/GTA/G/CA/TNNNNT/G/CAA/CA/T/GA/TA/T, Figure 2.16A). NAC, including GNRF, binding sites and MYB binding elements (Figure 2.16B) were used to identify putative binding sites along the promoter and 5'UTR sequences of *GNRF*, *SWAM1*, *SWAM4* and the cell wall biosynthetic genes *CESA4*, *CESA7*, *CESA8*, *CAD1* and *COMT6*. A summary of the corresponding MYB and NAC binding sites and data related to the protein-DNA binding in yeast-one hybrid assay obtained previously was constructed (Figure 2.17). Yeast-one-hybrid and DAP-seq results showed that GNRF, SWAM1, and SWAM4 bound the 5'UTR and the adjacent promoter region of the transcription factor genes *GNRF*, *SWAM1* and *SWAM4*.

While NAC binding sites were located across the promoters of cellulose genes, MYB binding sites were found in the 5'UTR and within the promoter. Only two SNBE sites were found in the promoter of *CAD* and none in the promoter of *COMT*. While other NAC binding sites were located in the promoter of *COMT* distant from the 5'UTR, the MYB binding sites were located near the 5'UTR of *COMT*.

A



B

Protein Family	Binding Element	Sequence	References
NAC Binding Elements	SNBE	WNNYBTNNNNNNNAMGNHW	2011, Zong et al
	VNS	CKTNNNNNNNAMGHH	2018, Olins et al
	GNRF-BS	CKTVHNNNNBAMDHH	
MYB Binding Elements	SMRE	ACCWAMY	1989, Lois et al 1995, Hatton et al 2003, Raes et al 2012, Zhong & Ye 2013, Prouse & Campbell 2018, Handakumbura et al
	AtMYB61-BS	YAMCHAMH	2016, O'Malley et al

Figure 2. 16 **NAC and MYB binding elements.** (A) GNRF binding motif identified by the DAP-seq method. (B) Summary of identified binding elements of NAC and MYB protein families with their corresponding sequences.

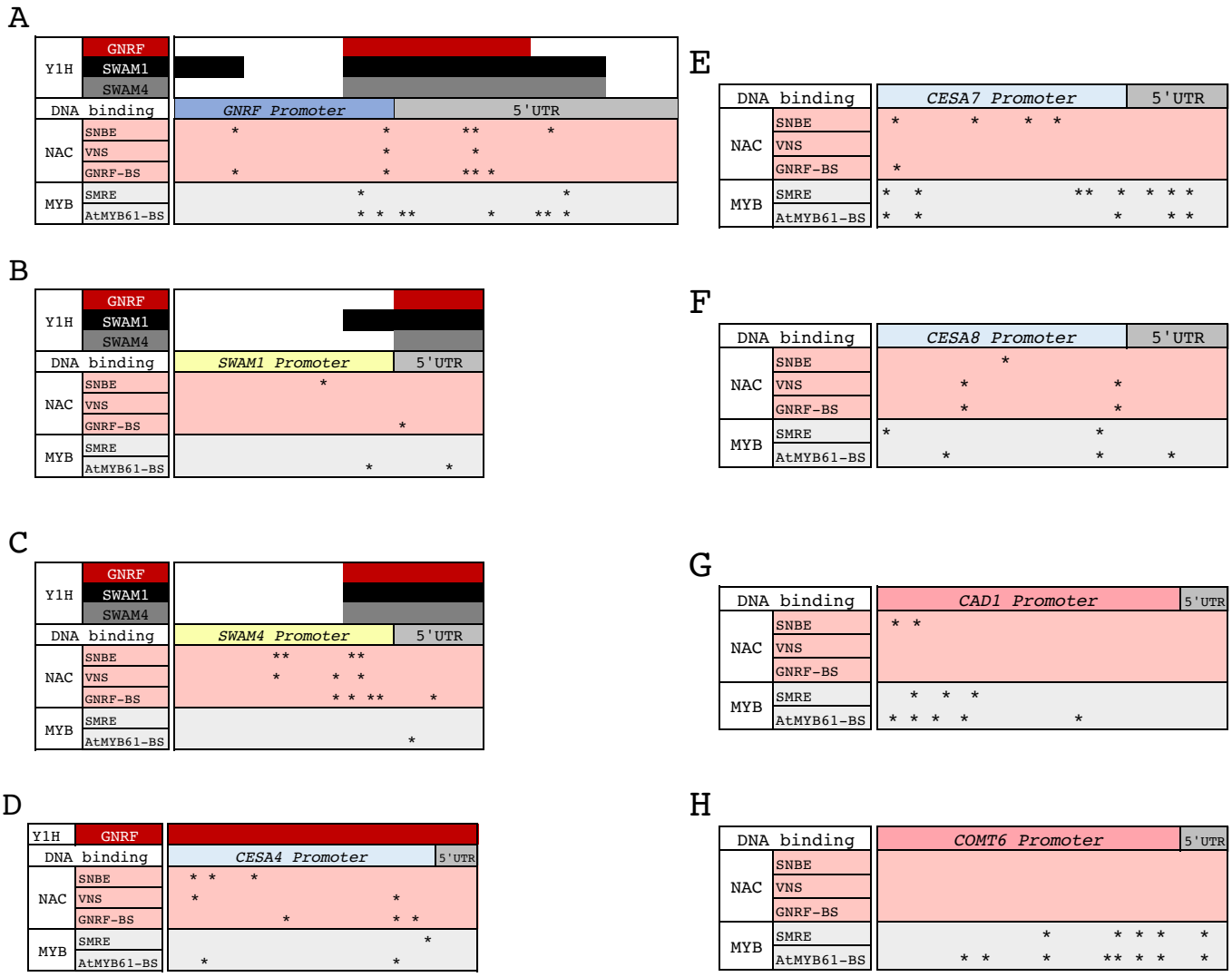


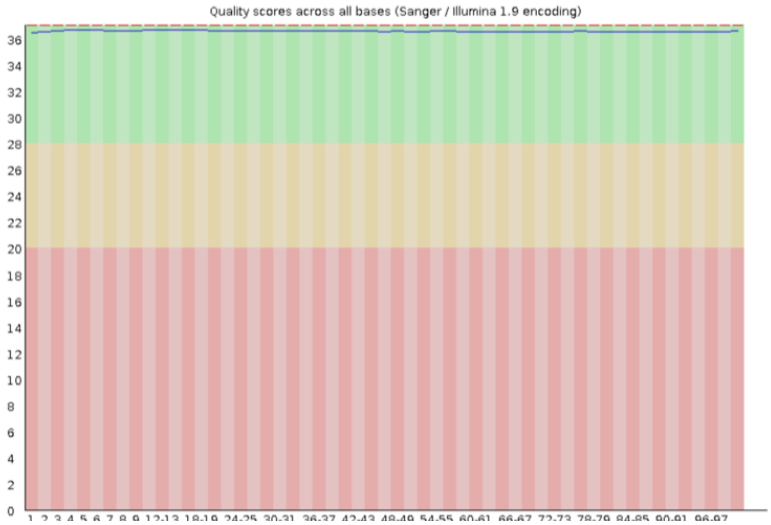
Figure 2. 17 **Summary of Protein-DNA binding analysis.** Diagram that displays the data obtained from the GNRF, SWAM1, SWAM4 protein binding to the corresponding DNA binding sequence in the yeast-one-hybrid assay. Significant interaction is illustrated by a color bar assigned to each transcription factor. NAC binding elements: SNBE (Secondary Wall NAC Binding Element), VNS (VND, NST, and SND consensus binding motif) and GNRF-BS (Binding Site). MYB binding elements: SMRE (Secondary Wall Myb Responsive Element) and AtMYB61-BS. All sequence elements identified in promoters and 5'UTR of *GNRF* (A), *SWAM1* (B), *SWAM4* (C), *CESA4* (D), *CESA7* (E), *CESA8* (F), *CAD* (G) and *COMT* (H) are indicated with asterisks (*)

2.3.11 *GNRF* is a repressor of genes associated with cell wall formation, transport, and flowering

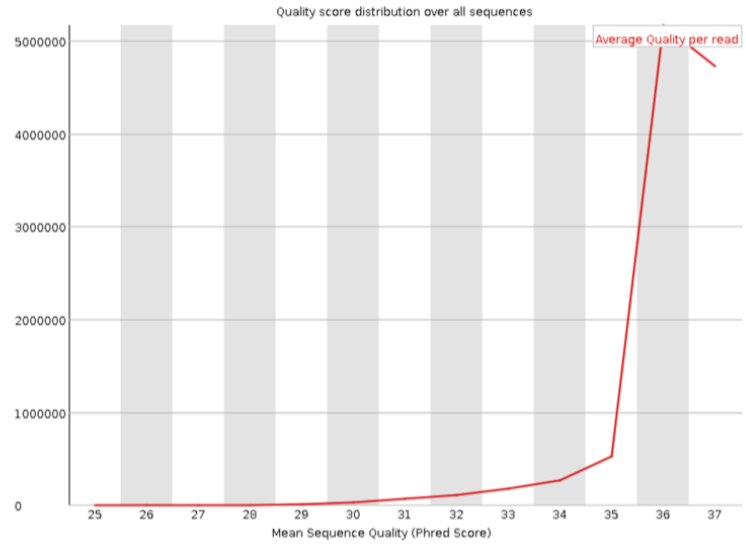
Transcriptome analysis was performed using RNA-seq of RNA from stems of WT, *gnrf-1*, *gnrf-2*, *gnrf-3*, *gnrf-*, and *gnrf-6*. Replicates of all genotypes were collected in the middle of the dark cycle (Table 2.4). Total RNA was extracted, selected by polyA sites, converted to cDNA in a library preparation and sequenced with the Illumina NovaSeq 6000 system-flow cell type S2 paired-end sequencing x100 bp. An assessment of the quality of the data was performed by FastaQC, generating a report of acceptable scores of sequence quality, sequence content, adapter content and overrepresented sequences. Results before and after trimming are shown (Figure 2.18).

Basic Statistics

Measure	Value
Filename	SR24_S6_L001_R1_001_P.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	11130382
Sequences flagged as poor quality	0
Sequence length	36-101
%GC	51



Per sequence quality scores



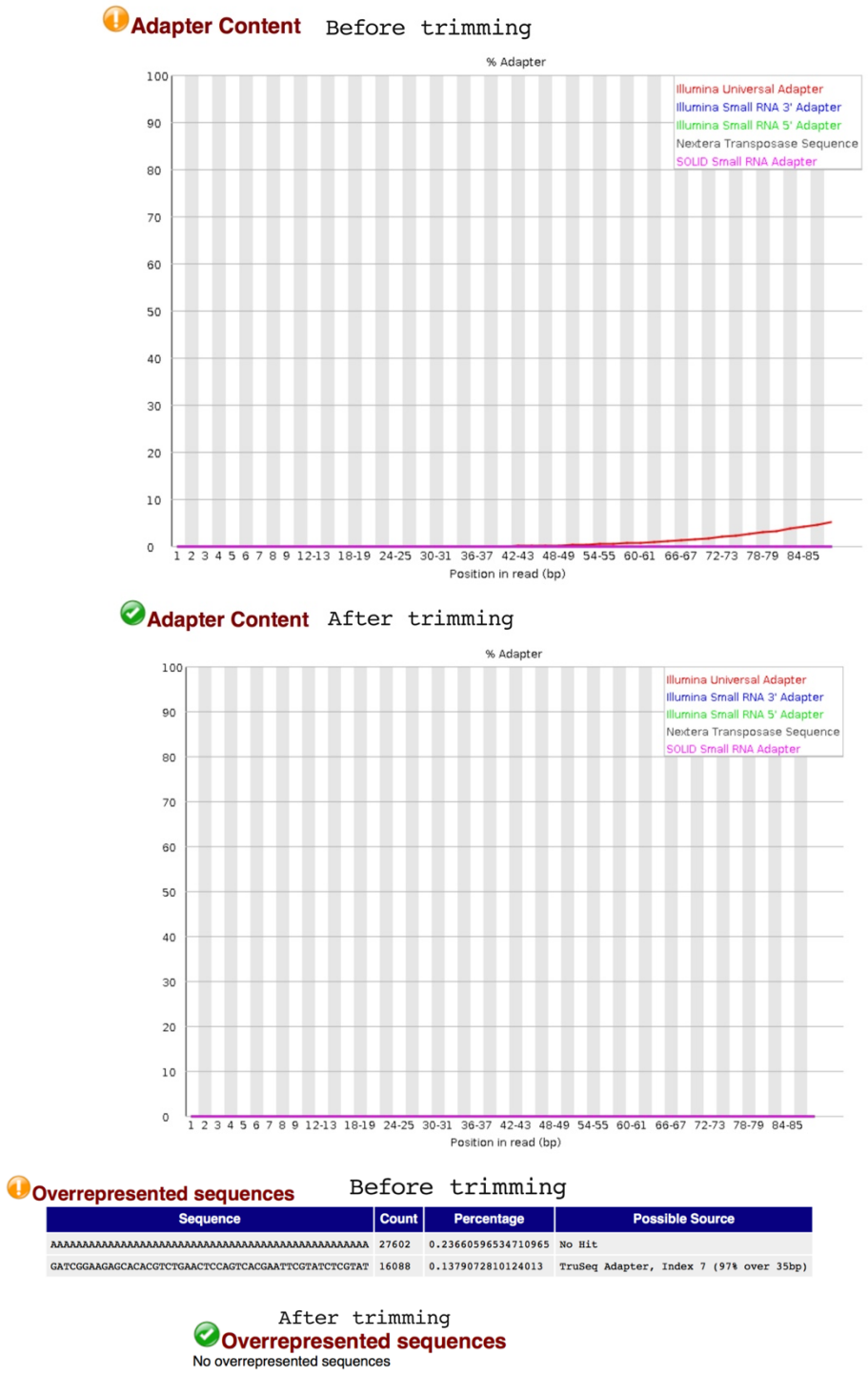


Figure 2. 18 **Summary of FastQC Report.** Sample SR24 gnrf-1 was selected as an example to evaluate the quality of raw sequence data. Per base sequence quality and quality scores graphs shown after trimming. Adapter content and overrepresented sequences before and after trimming are indicated.

Heatmaps and hierarchical analysis, PCA, and scatter plots were created comparing each group of replicates of individual genotypes with WT replicates. A stringent p-value <0.005 was used to generate these diagrams, but a less stringent p-value <0.05* and 0.01** was used to determine significant differential expression in further analysis. PCA of the transcriptome dynamics of all samples revealed that the *GNRF-OE* samples were similar to each other and very different from WT and *gnrf* mutants (Figures 2.19, 2.20A). PCA showed that *gnrf-2* and *gnrf-6* mutants also had high correlation within replicates (Figure 2.20C, F). *gnrf-1*, *gnrf-3*, and *gnrf-4* showed some variation within the replicates and the variables (Figures 2.20B, D, E). The heatmaps displayed the difference of genes upregulated (yellow) and downregulated (red) within replicates and the hierarchical cluster analysis was used to determine particular patterns in gene expression. Similar to PCA results, a heatmap showed a striking difference in gene expression between *GNRF-OE* and WT replicates (Figure 2.21A). *gnrf-2* replicates also showed a high similarity among replicates and a very different pattern compared to WT (Figure 2.21C). The gene expression of one replicate of *gnrf-6* (Figure 2.21F) did not show a similar pattern within the *gnrf-6* replicates, and it was clustered in the WT dendrogram; nevertheless, the other replicates showed a distinct gene expression pattern compare to WT replicates. Even though *gnrf-1*, *gnrf-3*, and *gnrf-4* showed a less differentiated gene expression pattern, the hierarchical analysis grouped the replicates with the corresponding clusters. (Figure 2.21B, D, E). Scatter plots were created by plotting the log2 fold change expression and the significance. Genes above and below 0 were identified as upregulated and downregulated respectively. Genes with a significant threshold p<0.005 were plotted in red. *GNRF-OE* samples showed a higher number of differentially expressed genes followed by *gnrf-2* and *gnrf-6.* The same proportion of genes were

found upregulated and downregulated in these genotypes (Figure 2.22A, C, F). Less number of genes deferentially expressed were found in *gnrf-1*, *gnrf-3*, and *gnrf-4* (Figures 2.22B, D, E).

A list of genes including genes from the *GNRF* co-expression analysis, genes that were downregulated in a microarray assay using *GNRF-OE* stem samples in a previous experiment (data not shown), and genes putatively associated with cell wall biosynthesis based on annotation and homology was created. Genes differentially expressed in the RNA-seq data from this subset list and additional genes that were significantly upregulated were included in a list of genes differentially regulated by *GNRF* (Table 2.5). Fifty-two out of the 96 genes in the co-expression of *GNRF* (Table 2.3) were differentially regulated by *GNRF*.

Transcription factors, flowering related genes, laccases, peroxidases, transporter and cell wall associated genes including cellulose lignin, hemicellulose biosynthetic genes were regulated by *GNRF*. *GNRF* expression in *GNRF-OE* plants was upregulated by 2.3-fold. Conversely, *GNRF* was downregulated by 9.02-fold in *gnrf-6* mutants. None of the other mutants showed a significant change in the *GNRF* gene expression. Interestingly, *SWAM1* was significantly downregulated in *GNRF-OE* plants and upregulated in *gnrf-2*, *gnrf-3*, and *gnrf-4* mutants. *SWAM4* was upregulated in *GNRF-OE* and *gnrf-2* (Figure 2.23A). Overall, *GNRF* appears to repress the expression of putative cell wall regulators (Figure 2.23B) and cell wall biosynthetic genes including *CESA4*, *CESA7*, *CESA8*, *CAD*, and *COMT* (Figure 2.24)

Several putative flowering time genes including *MAD5*, *VRN1*, *FTL2*, *FTL12*, *FUL2*, and *MADS*1, were dramatically repressed (from -14.47 to -11.39) in *GNRF-OE* plants. *VRN1* and *FTL2* were significantly upregulated in *gnrf-2* (Figure 2.25). Glycosyltransferases, peroxidases, laccases, and transport proteins were also repressed in *GNRF-OE* and upregulated in different

gnrf mutant plants (Figure 2.26). In general, *GNRF* appears to function as a pleiotropic repressor of cell wall, flowering, and transport associated proteins (Table 2.5).

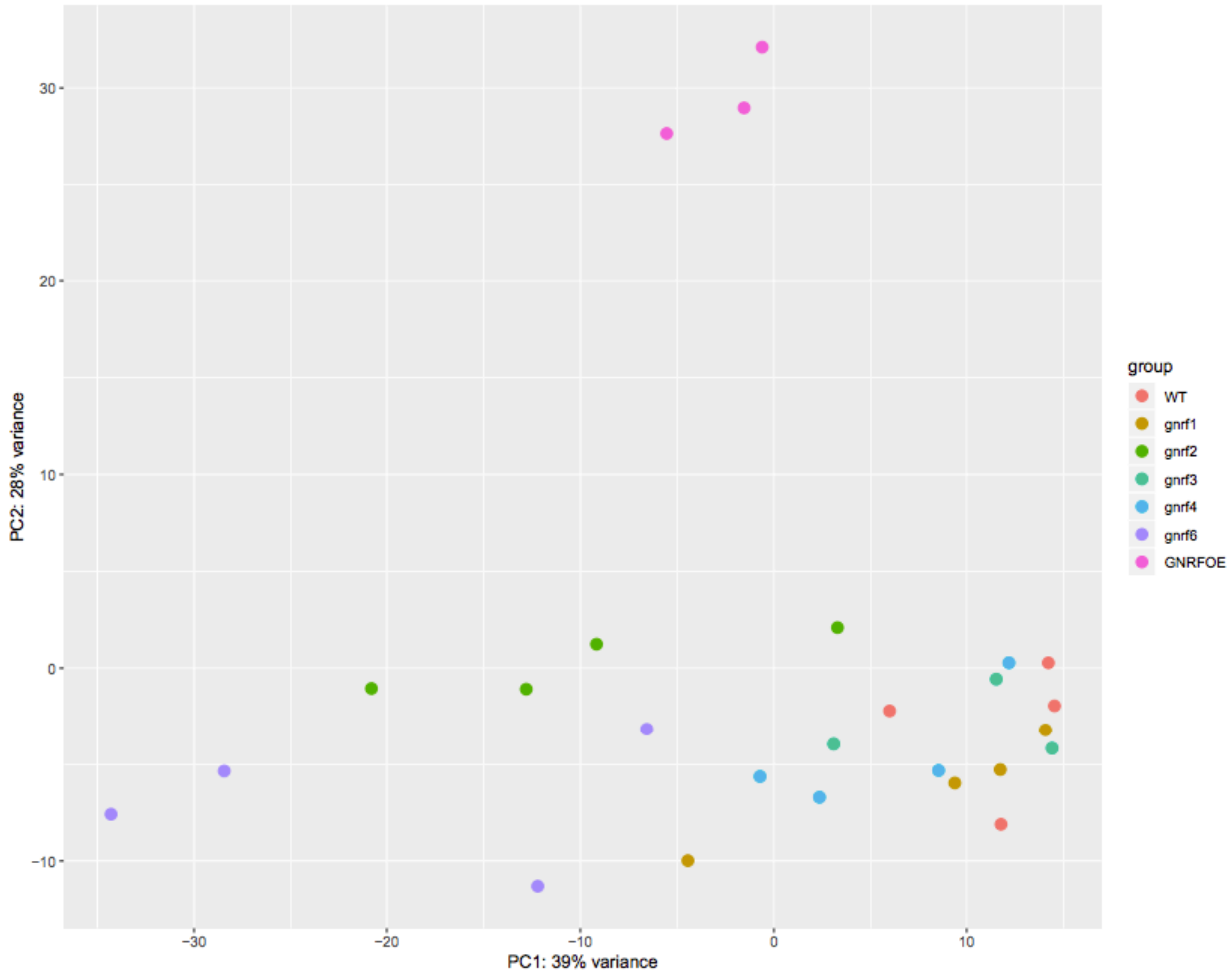


Figure 2. 19 **Genetic distribution of WT, *GNRF-OE*, *gnrf-1*, *gnrf-2*, *gnrf-3*, *gnrf-4* and *gnrf-6* individual plants.** Principal Component Analysis (PCA) illustrating the distribution of individual samples along the two PC.

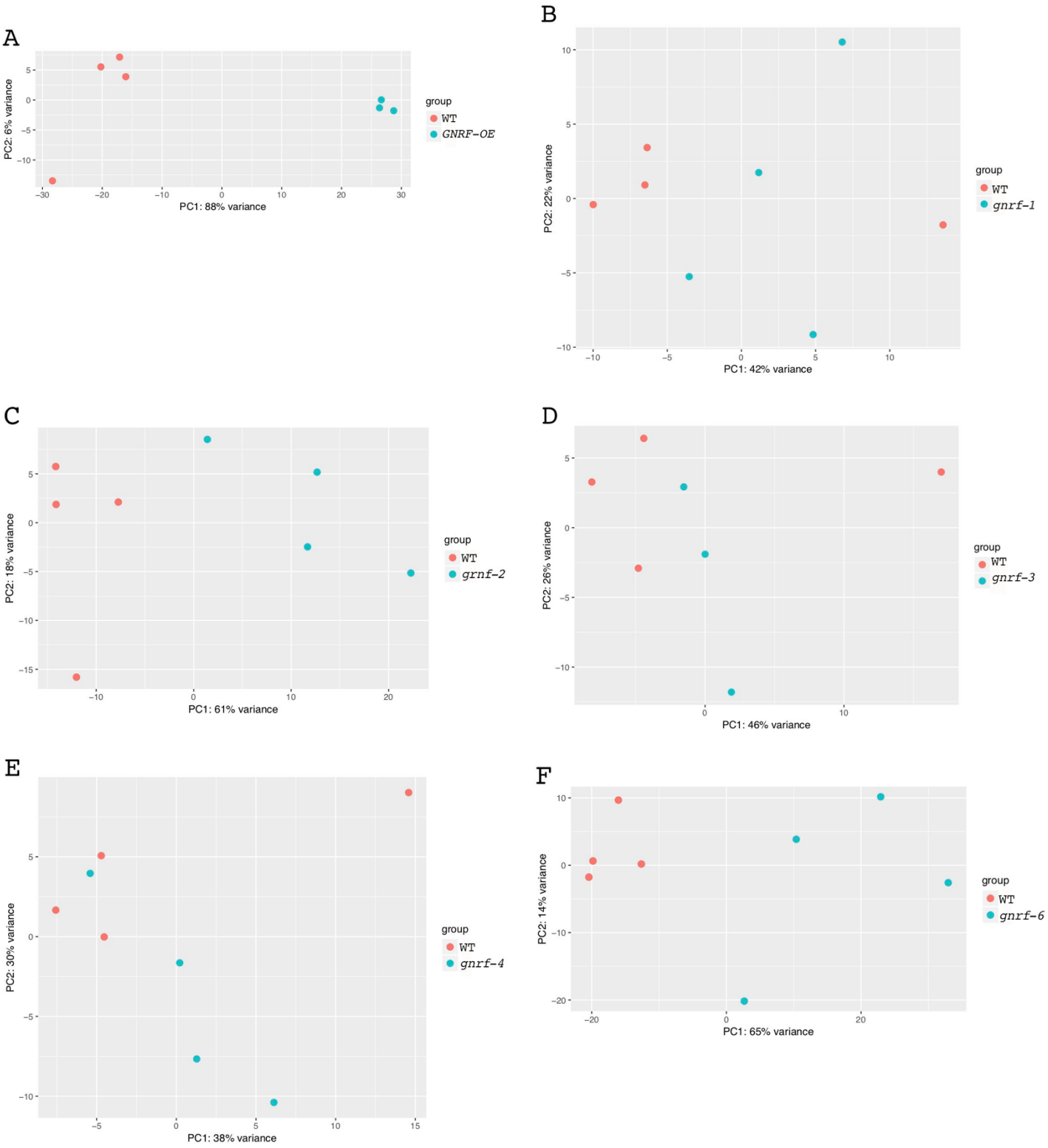


Figure 2. 20 **Total variation and distribution of WT and *GNRF* mutants. Principal Component Analyses (PCA).** Percentages of variance are indicated.

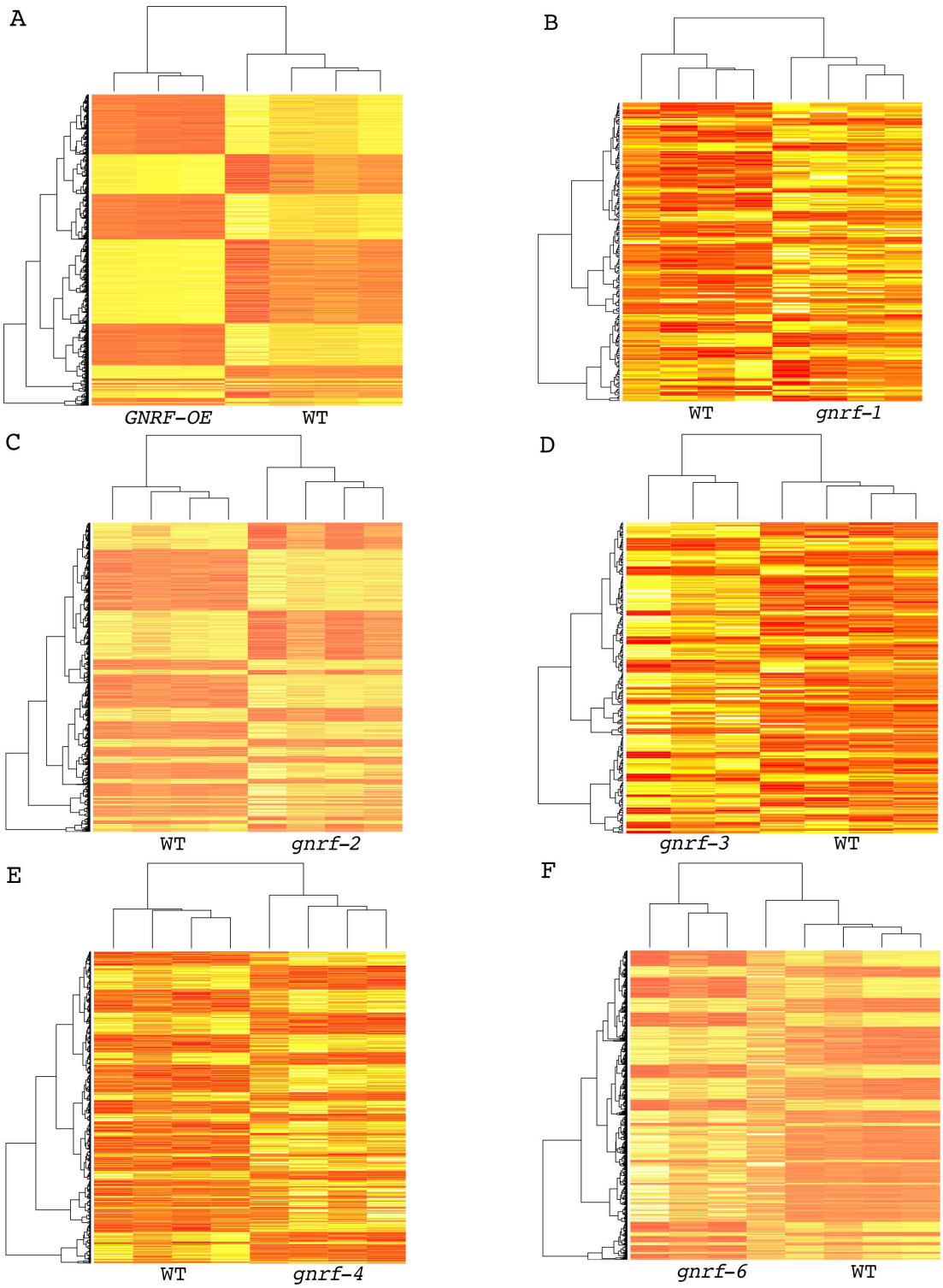


Figure 2. 21 **Differential gene expression of WT and *GNRF* mutants. Heatmaps and hierarchical analyses.** Upregulation is indicated in yellow and downregulation in red.

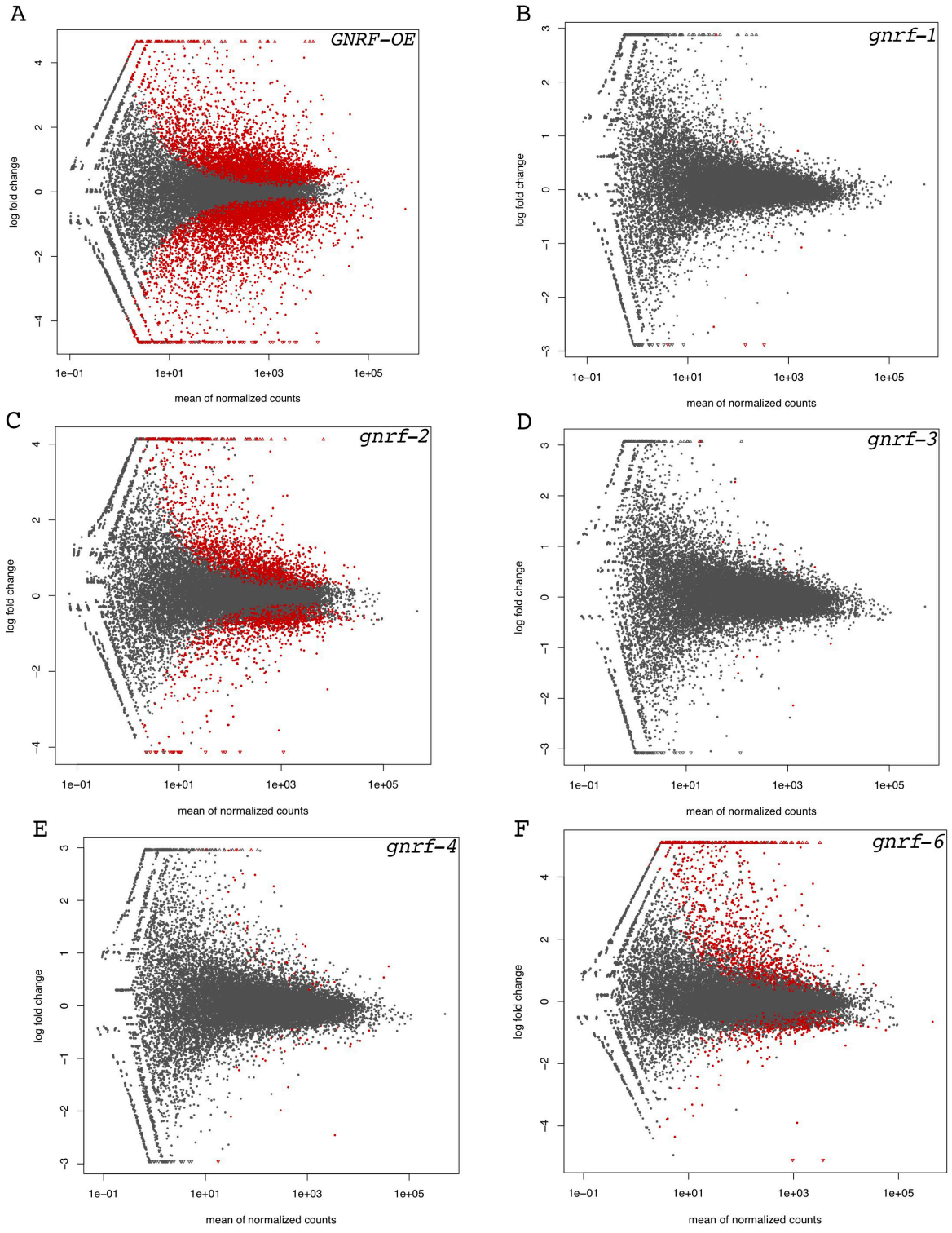


Figure 2. 22 **Genes differentially expressed (Log2 Fold change) of *GNRF* mutants relative to WT**. **Scatter plots**. Upregulation (above 0) and downregulation (below 0) are illustrated. Significance threshold p<0.005 (red dots).

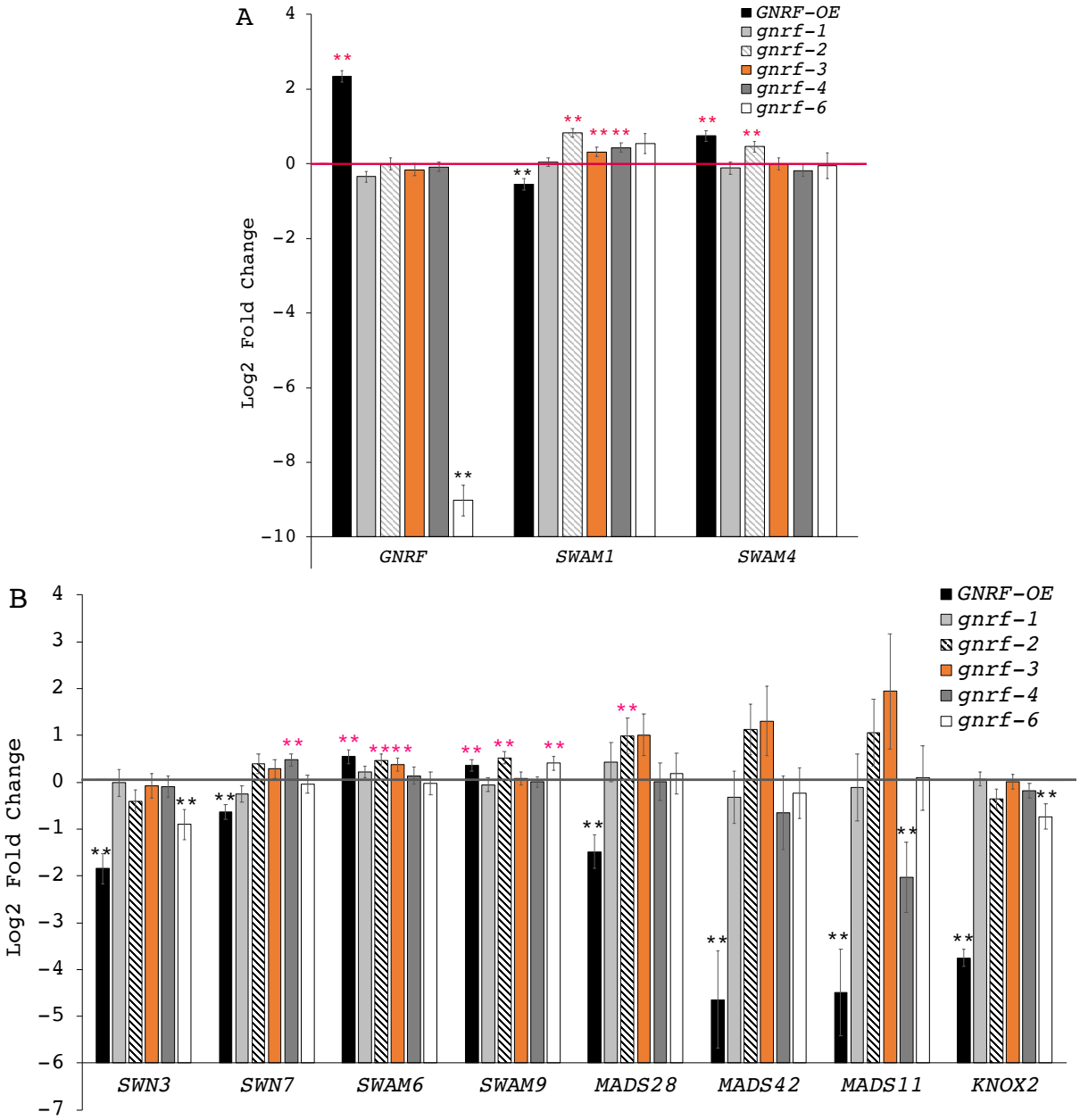


Figure 2. 23 **Transcription factors that are differentially expressed in *GNRF* mutants relative to WT.** (A) Log2 Fold change (LFC) of *GNRF*, *SWAM1* and *SWAM4* (B) Log2 Fold change of transcription factors with homologues genes in other species associated with cell wall formation. Log2 Fold change ± log fold change standard error (lfcSE). Positive and negative LFC values are above and under the threshold (0) respectively. *p<0.05, **p<0.01.

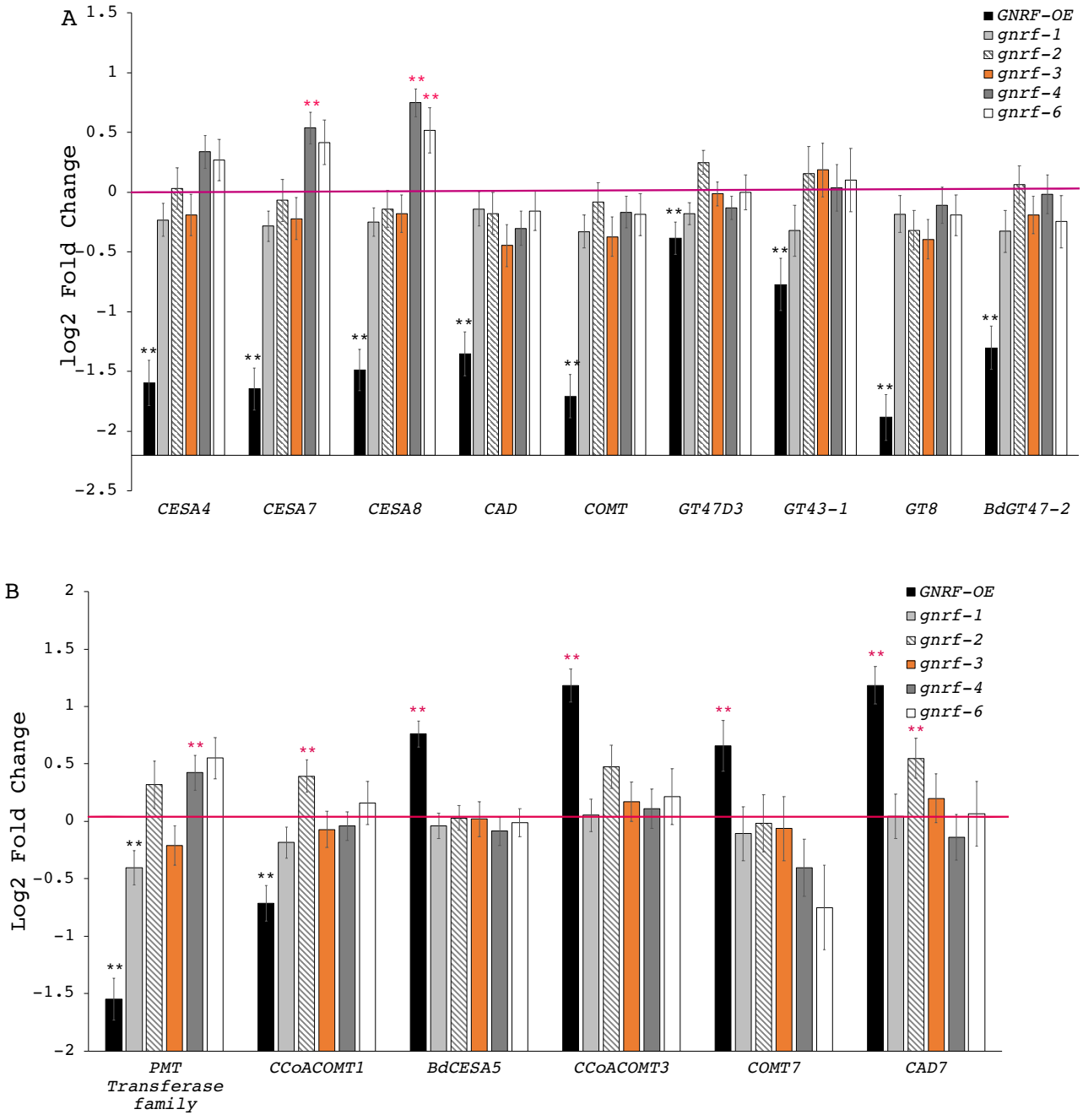


Figure 2. 24 **Cell wall biosynthetic genes that are differentially expressed in *GNRF* mutants relative to WT.** (A) Log2 Fold change (LFC) of cellulose (*CESA4*, *CESA7*, *CESA8*), lignin (*CAD1*, *COMT6*), and hemicellulose biosynthetic genes (*GT47D3*, *GT43-1*, *GT8*, *GT47-2*). (B) Cell wall associated genes. Log2 Fold change ± log fold change standard error (lfcSE). Positive and negative LFC values are above and under the threshold (0) respectively. *p<0.05, **p<0.01.

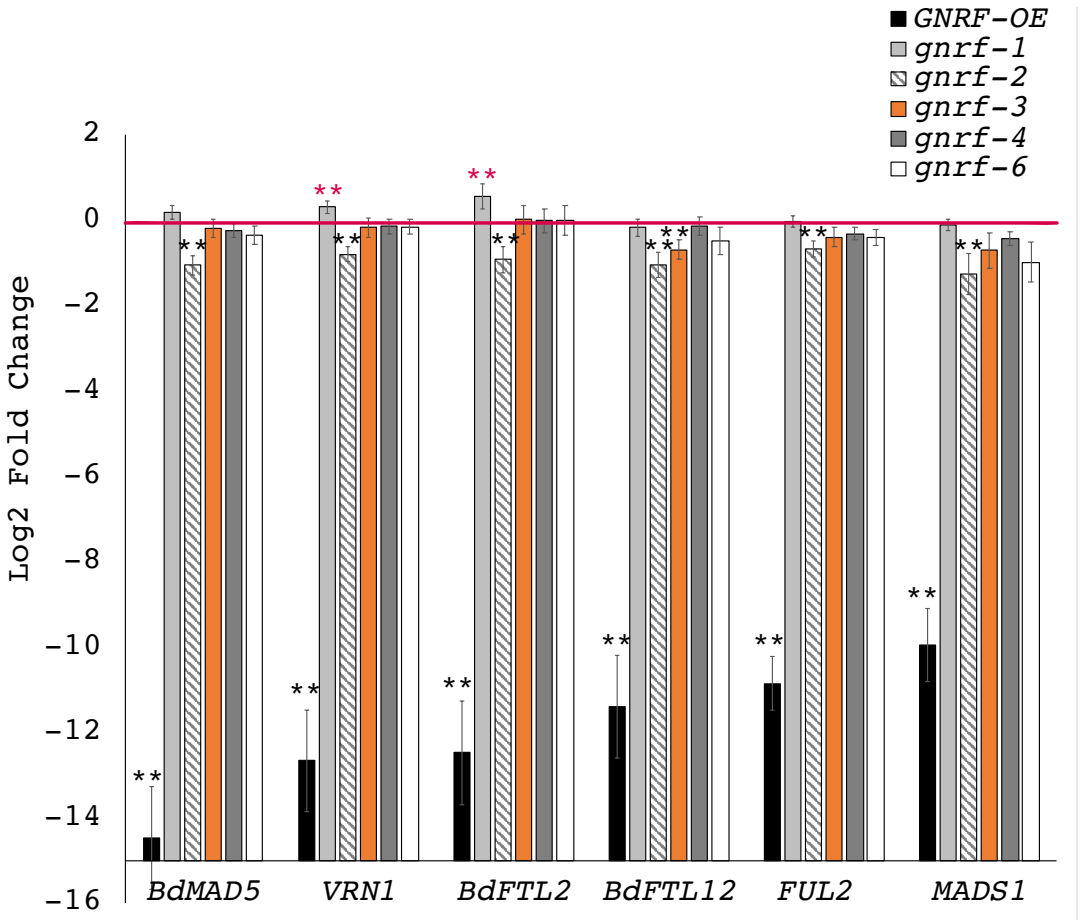


Figure 2. 25 **Flowering pathway genes that are differentially expressed in *GNRF* mutants relative to WT.** Log2 Fold change ± log fold change standard error (lfcSE). Positive and negative LFC values are above and under the threshold (0) respectively. *p<0.05, **p<0.01.

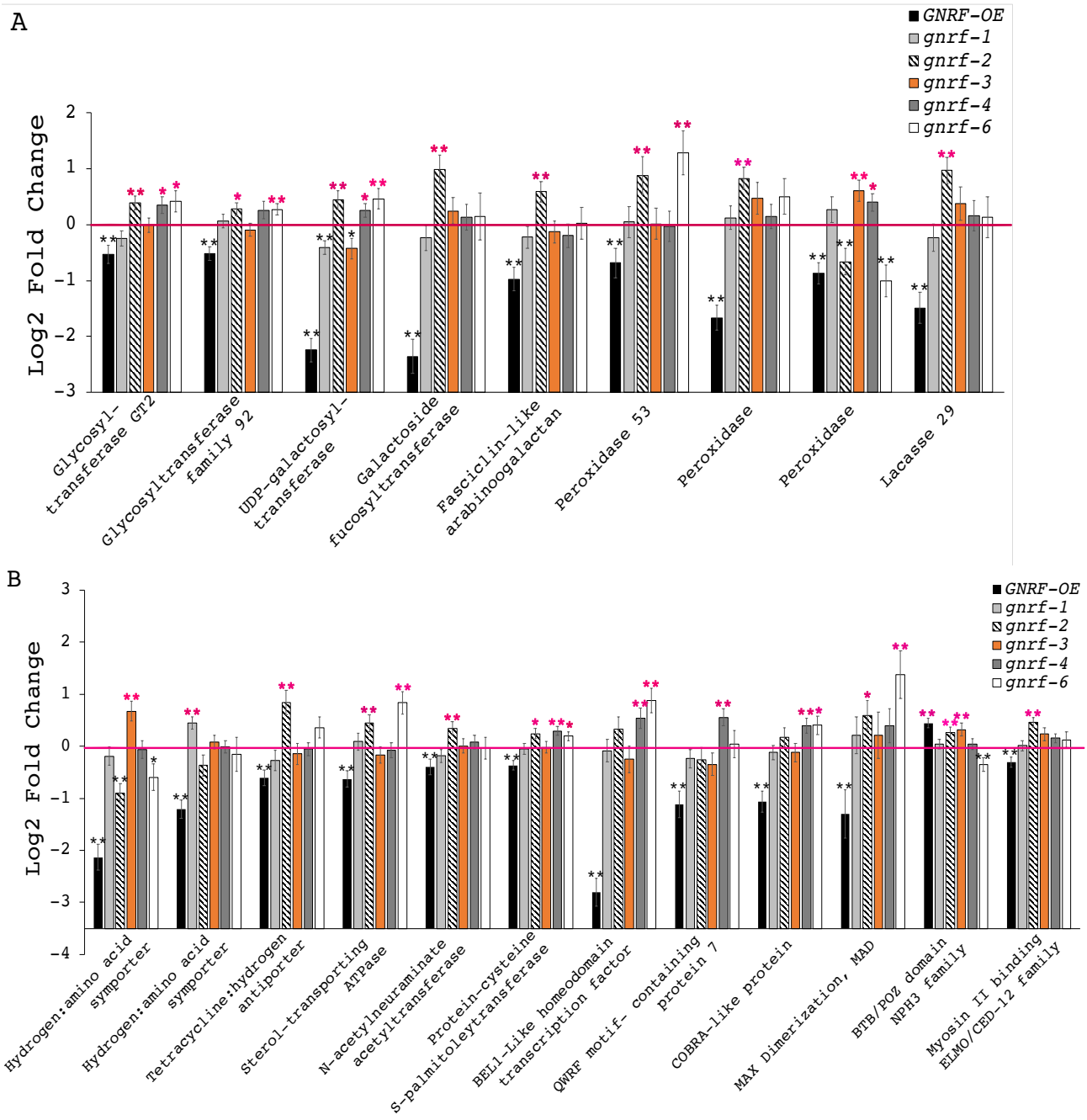


Figure 2. 26 **Genes of interest differentially expressed in *GNRF* mutants relative to WT.** The genes were selected based on: homology to genes associated to cell wall formation, listed in the GNRF cluster for co-expression analysis (Figure 2B), and shown differential expression in a microarray data from stem tissue of GNRF-OE plants. Log2 Fold change ± log fold change standard error (lfcSE). Positive and negative LFC values are above and under the threshold (0) respectively. *p<0.05, **p<0.01.

2.4 Discussion

Phylogenetic analysis revealed that *AtSND2* is the closest ortholog of *GNRF*. Effects on cell wall thickness in overexpression and dominant repressor lines of *AtSND2* suggests, that it activates the expression of secondary wall biosynthetic genes (Zhong et al., 2008). Based on protein similarity, it reasonable to expect GNRF to have a similar role in the regulation of the secondary cell wall. Indeed, orthologs of *AtSND2* in rice, poplar, and switchgrass have been identified as positive regulators of secondary cell wall formation (Hussey et al., 2011a; Rao and Dixon, 2018; Ye et al., 2018). The phylogeny suggests that grasses have two copies of SND2. *P. virgatum* has two copies of orthologous proteins in each clade, likely a result of its tetraploid genome (Ma et al., 2012). All species examined from the subfamily Pooideae, *B. distachyon*, wheat, and barley, have only one copy of the *SND2* ortholog. Thus, the Pooideae may have lost one copy of SND2. Interestingly, analysis of gene expression in rice describes the two rice orthologs of *SND2*, *Os01g48130*, and *Os05g48850*, as distinctly regulated and independent of secondary cell wall formation (Guo et al., 2014). Functional and comparative analysis of paralog *SND2* genes arranged in different clusters would clarify their role in a specific biological process, including cell wall formation.

Gene expression studies and comparative co-expression analysis have facilitated the identification of transcription factors functioning in complex networks; several of these transcription factors belong to the NAC and MYB protein families (Ferreira et al., 2016; Sibout et al., 2017; Smita et al., 2015). *GNRF* and *SWAM1* genes also clustered in this co-expression network with genes implicated in cell wall biosynthesis. This result would indicate that the function of *GNRF*, like *SWAM1*, may be related to this biological process in *B. distachyon*. In sorghum, higher expression of MYB and NAC genes has been detected in old and fully

expanded stem internodes where secondary wall biosynthesis occurs (Kebrom et al., 2017). This finding is consistent not only with the fact that *GNRF* and *SWAM1* are highly expressed in stems, compared to roots and leaves, but also with their possible relationship in the regulation of cell wall formation.

The prolonged vegetative growth in *GNRF-OE* appeared to induce node formation. This process occurred after the development of a short internode elongation and allowed for substantial above-ground biomass at senescence due to the extended number of days of growth. Stem photosynthetic and accumulate resources. The transition to the reproductive stage in *B. distachyon* Bd21-3 plants seems to require specific stem-elongating conditions (vernalization and longer photoperiod) to reach the transition to flower (Jensen and Wilkerson, 2017). *GNRF-OE* plants were vernalized and grown at long photoperiod; hence *GNRF* function seems to be related to internal cues associated with the flowering process. Despite the fact that *gnrf* mutants are heavily mutagenized, the overlapping difference in flowering, plant height and biomass compared to wild type, supports the notion that *GNRF* has a role in plant development and flowering.

Reduction in lignin content, stem area, and interfascicular wall thickness in *GNRF-OE* plants indicate that *GNRF* is a repressor of cell wall formation regulating lignin biosynthesis. The lighter stain pattern in interfascicular fibers of *gnrf* stem cross sections without a significant reduction in total lignin suggests that lignin is deposited at the similar amount level, but its deposition is occurring differentially in cell type. For instance, lignin staining appears to be lower in interfascicular fibers, so it is probably higher in other cell types, like vascular bundle cells. This changed in the distribution of lignin in cells carrying out secondary cell wall

formation would compensate the lignin content in the whole stem and would explain the fact that *gnrf* mutant plants are still straight.

While *AtSND2* and *OsSND2A* are positive regulators of cellulose and activate cell wall thickening (Hussey et al., 2011a; Zhong et al., 2008), GNRF is a repressor of cellulose biosynthesis. Little is known about the function of other NAC SND2 proteins, but the repressor function of *GNRF* and the finding of only one copy of SND2 proteins in the subfamily Pooideae that includes *B. distachyon* (Figure 2.1 pink), provides a notion of a unique regulation. This finding also suggests that other grasses that have two or more orthologous SND2 proteins may function either redundantly as activators or antagonistically as repressors.

Transcript abundance of *SWAM1*, a positive regulator of cellulose and lignin associated gene expression (Handakumbura et al., 2018), was found downregulated by *GNRF*. *SWAM1* was downregulated in stem samples of *GNRF-OE* plants grown at long and short days and analyzed by Q-RT-PCR and RNAseq analysis. Consistent with this result, *SWAM1* transcript expression was upregulated in stem samples of *gnrf-1* and *gnrf-2* in the Q-RT-PCR analysis and *gnrf-2*, *gnrf-3* and *gnrf-4* in the RNAseq analysis. The repression of *SWAM1* by *GNRF* supports the downregulation of cellulose and lignin genes observed.

Protein-DNA interaction analysis using the luciferase-compatible yeast-one-hybrid assay revealed that GNRF binds to the *CESA4* promoter, the *SWAM1* 5'UTR, the *SWAM4* promoter and the *SWAM4* 5'UTR in yeast. Interestingly, SWAM4 and SWAM1 proteins were found binding the *GNRF* 5'UTR and the *GNRF* promoter. These findings suggest that GNRF protein may regulate itself expression and downregulate *CESA4* either directly or indirectly by controlling *SWAM1* and/or *SWAM4* gene expression. Yeast-one-hybrid and DAP-seq results

(Figure 2.17) suggest that GNRF, SWAM1 and SWAM4 proteins are more likely to bind the 5'UTR and the region adjacent to the 5'UTR of the transcription factors tested. This result indicates a possible function of the untranslated region in the translation initiation and the regulation of gene expression that has been proposed in eukaryotic genes (Barrett et al., 2012) and *A. thaliana* genes (Kim et al., 2014) but that requires further investigation. NAC and MYB binding sites in cellulose biosynthetic genes were preferentially located along the promoter region. MYB binding sites were more frequent than NAC binding sites in the promoter of lignin biosynthetic genes. These binding patterns indicate preferential regions that could be used as binding sites and provide information that contributes to the elucidation of both, direct and indirect regulation. Overall, yeast-one-hybrid and DAP-seq approaches have provided an insight into the function of GNRF, SWAM1 and SWAM4 proteins in the cell wall regulation. However, protein-DNA approaches that provide information to identify the binding sites of these proteins *in vivo* are required to confirm the putative interactions found and to identify their function within the transcriptional network that regulates secondary cell wall biosynthesis in *B. distachyon*.

Transcriptomic analysis by RNAseq in *GNRF-OE* and *gnrf* mutant stems showed a higher correlation between *GNRF-OE*, *gnrf-2* and *gnrf-6* replicate samples compared to *gnrf-1*, *gnrf-2*, *gnrf-3*. Nonetheless, hierarchical clustering grouped replicates and showed differences between variables. RNAseq, microarray, and gene expression analysis by RT-Q-PCR have distinct parameters and metrics to quantify expression. However, the gene expression results obtained by the given statistical significance in each approach provided a similar trend. *GNRF* was found upregulated in *GNRF-OE* samples, as expected and downregulated in the *gnrf-6* mutant. Since the T-DNA insertion in *gnrf-6* interfered with *GNRF* expression, it is a suitable

mutant allele. Even though the non-synonymous mutations have a heavily mutagenized background and did not show a reduction in *GNRF* expression, they still are valuable due to their potential to alter protein function. The repression of *SWAM1* by *GNRF* in the transcriptomic analysis was confirmed and is consistent with the downregulation of cell wall associated genes observed. The fact that most of the genes, including *SWAM1*, that clustered in the co-expression analysis appeared to be regulated by *GNRF* provides a clear indication of a transcriptional network.

The strong flowering phenotype of *GNRF-OE* plants can be explained by the dramatic downregulation of flowering-associated genes. *GNRF* was also found related to the regulation of glycosyltransferases, peroxidases, laccases, and transport proteins. These results not only suggest that *GNRF* has a pleiotropic function but also give insight into the relationship of distinct proteins and components of the developmental processes of secondary cell wall biosynthesis and flowering in *B. distachyon*. Assays that test protein-DNA interactions *in planta* would clarify binding sites and functions.

Table 2 .1 **List of primers for Chapter 2**

Primer Name	Sequence
PH_N1F2_GNRF	ACATGGTGCAATAGCTTCAACGACG
PH_N1R2_GNRF	CCAGTCCTAATCGATCCGGGATC
IW_Gal4BDTOPOFWD	CACCATGAAGCTACTGTCTTCTATC
P01_GNRF_Pr_F1	CACCAATATATAGCACGGTCGATG
P02_GNRF_Pr_R1	TTTGGTTGTTGCTTCCCAAG
P03_GNRF_Pr_F2	CACCTGATTTCGTTGTTCTCTGAATTT
P04_GNRF_Pr_R2	CCATATGTGGCAAGATTAGAGA
P05_GNRF_Pr_F3	CACCGCGCGCAGAGATCATATCTCTCC
P06_GNRF_Pr_R3	GGCCGCCCGGGCTCGATCTC
P13_GNFR_attB5_F	GGGGACAACTTTGTATACAAAAGTTGCTATGACATGGTGCAATAGCT
P14_GNFR_attB2_R	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAGGGGCCAAAGCCTGT
P15_GNFRCDattL1_F	CACCATGACATGGTGCAATAGCTTCAAC
P16_GNFRCDattL1_R	TCAGGGGCCAAAGCCTGTCCCTCCC
P17_pDEST22_F	TATAACGCGTTTGGAATCACT
P18_pDEST22_R	AGCCGACAACCTTGATTGGAGAC
P19_pDEST15_F	GCGACCATCCTCCAAAATC
P20_pDEST15_R	TAAAGCCTGGGGTGCCTAAT
P76_Hyg_F	AGAATCTCGTGCTTTCAGCTTCGA
P77_Hyg_R	TCAAGACCAATGCGGAGCATATAC
P78_pMDC32_F	CGCACAATCCCACTATCCTT
P79_pMDC32_R	AAGACCGGCAACAGGATTC
P80_pLUC_F	TGTGCTCCTTCCTTCGTTCT
P81_pLUC_R	CTTCGTTGTTTTCGGTTGGT
P84_pLUC2_F	AAGCTTGAATTCGAGCTCGGTACCC
P85_pLUC2_R	CAGAGCACATGCCTCGAGGTCGA
P98_CaMV_35SPr_F	CTATCCTTCGCAAGACCCTTC
P99_attl_seq_F	AACAAGTTTGTACAAAAAAGC
P100_HygF	GCGAAGAATCTCGTGCTTTC
P101_HygR	GATGTTGGCGACCTCGTATT
P104_NaNGNRF2_F	TACACCAACTACGGCAAGCA
P105_NaNGNRF2_R	TAACATCGTTGTCCCAGTGC
P118_UbiGNRF_F	AGCTACGGGGGATTCCTTT
P119_UbiGNRF_R	GATGGACGGTGGAAGAAGTG
P120_35S_Hyg_F	GAAACCTCCTCGGATTCCAT
P121_35S_Hyg_R	ATAGGTCAGGCTCTCGCTGA
P132_M13F	GTAAAACGACGGCCAG
P133_M13R	CAGGAAACAGCTATGAC
P134_GNRFPrF4F	CACCACTTATATTCTTGTCGGTC
P135_GNRFPrF4R	CGTGATCGAGGGAGATTGTT
P136_GNRFPrF5F	CACCTCTACGGTGACCTCGGAT
P137_GNRFPrF5R	TGACCGATCACTTGCCAATA
P138_GNRFPrF6F	CACCTCTCCTAGCTTCCCTTTTCC
P139_GNRFPrF6R	ATCGATCACACCAAGGCGC
P140_GNRFPrF7F	CACCATATCATGTTGGGCTAGTTA
P141_GNRFPrF7R	AGAGAAATCGGACGGGATCT
P149_SWAM1PrF1F	CACCGGAAATAGGCAAAAGCTGGCTG
P150_SWAM1PrF1R	TGAACCGTGACAACCTGACGCTAC
P151_SWAM1PrF2F	CACCTTCGAGAAAAACAGCGTT
P152_SWAM1PrF2R	CATAGGCTGTGCGTTCCTT
P153_SWAM1PrF3F	CACCAATGTTCTGCATTGATGT
P154_SWAM1PrF3R	GGGATTTAAAATAGCTGCTGGGT
P155_SWAM1PrF4F	CACCACCTCCCAGAGACGGGATGT
P156_SWAM1PrF4R	AGCCCGCGGCCGGTTGAA
P157_GNRFCDSF	CACCATGACATGGTGCAATAGC

Primer Name	Sequence
P158_GNRFCDSR	TCAGGGGCCAAAGCCTGT
P159_SWAM4PrF1F	CACCGAAGGATGCCATTACAAT
P160_SWAM4PrF4R	TTCAGTGCAATGCAGAACCTG
P161_SWAM4Ex1R	AGTTTAGGGACTGAGCTCCAG
P162_SWAM4Ex3F	CAGTATGGGATTCGAGCCG
P163_CRESR	ACCACTTTGTACAAGAAAGCTGG
P164_SWAM4PrF1R	AGAGGGCATCTCCCTTTTGTTTAC
P165_SWAM4PrF2F	CACCCAATTATCTTTGCAATTATAGGA
P166_SWAM4PrF2R	ACTTCTACGTCACGAGTGTTCCA
P169_eGFPF	ATGGTGAGCAAGGGCGAG
P170_eGFPR	TCTTGTAGTTGCCGTCGTCCT
P171_SWAM4PrF3F	CACCACAGCACAGCTAACAAGATG
P172_SWAM4PrF3R	CGGCCATGCGTCAGGGTA
P173_SWAM4PrF4F	CACCGAAGCAACCAAGGGAAGA
P176_SWAM4CDS_F	CACCATGGGGAGGCATTCTT
P177_SWAM4CDS_R	CTACATATGCTCAAAAGACAAGGGCA
P178_SWAM1CDS_F	CACCAATGGGGCGGCAC
P179_SWAM1CDS_R	GCCTCAAAAGTACTCGAGGTTGAA
P180_pMDC32F_2	CTATCCTTCGCAAGACCCTTC
P191_attl_seq_F	AACAAGTTTGTACAAAAAAGC
P192_LUC_F	AGAACTGCCTGCGTGAGATT
P193_LUC_R	ATCCAGATCCACAACCTTCG
P194_LUC_Rev	ACTGCATACGACGATTCTGTG
JC_GNRF_TDNA_GSP_F	GCGATCTTCTTGGGAAGCAAC
JC_GNRF_TDNA_GSP_R	GGCCTGCAGGTATACACGTAC
JC_pJJ-LB-TDNARev3	AGCTGTTTCCTGTGTGAAATTG
P07_qPCR_Hyg_Fdw	ATTTCGGCTCCAACAATGTC
P08_qPCR_Hyg_Rev	GCGACCTCGTATTGGGAAT
P33_qPCRGNRFCDS_F	CAAGAAGCAGCAGCAACAGCAAC
P34_qPCRGNRFCDS_R	CGCAGCCTGCAACTGTTCATAC
P74_qPCR_GUS_F	CGACATGTGGAGTGAAGAGTATC
P75_qPCR_GUS_R	TCGGCGAAATTCCATACCTG
P124_SRUBIqPCRF	GGAGGCACCTCAGGTCATTT
P125_SRUBIqPCRR	ATAGCGGTCATTGTCTTGCG
P126_SRACTqPCRF	CCTGAAGTCCTTTTCCAGCC
P127_SRACTqPCRR	AGGGCAGTGATCTCCTTGCT
P187_GAPDHqPCR_F	TTGCTCTCCAGAGCGATGAC
P188_GAPDHqPCR_R	CTCCACGACATAATCGGCAC
P189_EF1alfaqPCR_F	CCATCGATATTGCCTTGTGG
P190_EF1alfaqPCR_F	GTCTGGCCATCCTTGGAGAT
PH_qPCRUBC18_F	TCACCCGCAATGACTGTAAG
PH_qPCRUBC18_R	ACCACCATCTGGTCTCCTTC
PH_qPCRGAPDH_F	TCACCCGCAATGACTGTAAG
PH_qPCRGAPDH_R	ACCACCATCTGGTCTCCTTC
PH_qPCRCAD1_F	AGGATAGAATGGGCAGCATCGC
PH_qPCRCAD1_R	ATCTTCAGGGCCTGTCTTCCTGAG
PH_qPCRCOMT4_F	TGGAGAGCTGGTACTACCTGAAG
PH_qPCRCOMT4_R	CGACATCCCGTATGCCTTGTTG
PH_qPCRCESA4_F	GCGTTTCGCATACACCAACACC
PH_qPCRCESA4_R	ACTCGCTAGGTTGTTCAGTGTGG
PH_qPCRCESA7_F	GCGATTCGCCTACATCAACACC
PH_qPCRCESA7_R	GGCTGGCAAATGTGCTAATCGG
PH_qPCRCESA8_F	CAAAGCACAAGTTCCGCCTGTG
PH_qPCRCESA8_R	TGGCTCGTATGCATCTGTCAAATC

Table 2.2 **List of transcription factors libraries used in Yeast-one-hybrid assays**

BradiTORFL

	1	2	3	4	5	6	7	8	9	10
A	Bradi3g52420	Bradi1g63780	Bradi3g44520	Bradi1g36750_1	Bradi4g04050_3	Bradi4g05300_1	Bradi2g05590_2	Bradi2g47887_1	Bradi5g25850_1	Bradi1g14750
B	Bradi2g08080	Bradi2g25400	Bradi2g11080	Bradi3g49440_1	Bradi4g01300_2	Bradi1g09090_1	Bradi4g02600_1	Bradi1g35990_1	Bradi2g54470	Bradi4g35630
C	Bradi5g10050	Bradi2g08460	Bradi1g48520	Bradi1g27120_1	Bradi4g01300_1	Bradi1g72150_5	Bradi1g73040_1	Bradi1g45510_1	Bradi2g57200	Bradi4g28280
D	Bradi4g33740	Bradi1g24390	Bradi1g47460	Bradi3g33570_1	Bradi5g21950_1	Bradi3g58220_1	Bradi3g50050_1	Bradi3g02730_3	Bradi2g49250	bHLH72
E	Bradi1g18200	Bradi1g01640	Bradi2g16120_1	Bradi3g07540_1	Bradi4g02580_1	Bradi3g13570_1	Bradi2g17982_1	Bradi3g21460_1	Bradi5g17810	Empty Vector
F	Bradi4g33750	Bradi3g34670	Bradi3g59290_1	Bradi1g72150_2	Bradi2g00740_1	Bradi3g17170_2	Bradi3g21460_2	Bradi3g45170_1	Bradi3g25670	
G	Bradi1g46210	Bradi4g42050	Bradi3g60557_1	Bradi1g72150_8	Bradi1g73040_3	Bradi3g46920_1	Bradi2g05590_3	Bradi3g49970_1	Bradi3g59380	
H	Bradi2g23890	Bradi1g16110	Bradi5g16127_1	Bradi5g25860_1	Bradi2g61000_1	Bradi3g01540_1	Bradi1g45510_2	Bradi5g18680_1	Bradi1g47690	

Sublibrary

	1	2	3	4
A	Bradi1g61400	Bradi5g20130	Bradi2g51990	Bradi1g17700
B	Bradi2g05700	Bradi1g13910	Bradi3g00300	Bradi3g38840
C	Bradi1g76970	Bradi1g10050	Bradi3g12920	Bradi3g37180
D	Bradi2g36730	Bradi1g13680	Bradi3g40090	Bradi4g06317
E	Bradi2g46197	Bradi1g72960	Bradi3g42430	Empty Vector
F	Bradi2g47590	Bradi2g17980	Bradi3g56290	
G	Bradi2g54940	Bradi2g40620	Bradi3g28970	
H	Bradi3g13120	Bradi2g48690	Bradi1g60670	

Sublibrary

	1	2	3	4
A	BdMYB18	BdMYB104	BdbHLH072	BdbZIP6
B	BdNAC28	BdHB7	Bradi3g00300	BdbZIP57
C	HB9	Bradi1g10050	Bradi3g12920	Bradi3g37180
D	BdMYB41/SWAM4	Bradi1g13680	BdNAC60	35 Myb-like
E	BdNAC38/GNRF	Bradi1g72960	BdMYB69	Empty Vector
F	BdMYB48/SWAM1	BdMYB31	BdbZIP61	
G	Bradi2g54940	BdMYB44	Bradi3g28970	
H	BdNAC49	Bradi2g48690	Bradi1g60670	

Table 2.3 **Genes in *GNRF* co-expression cluster.** Ninety-six genes in the co-expression network of *GNRF* (PlaNet).

Gene ID	Description
Bradi3g05750.1	4-Coumarate:CoA ligase (4CL)
Bradi1g31320.1	4-Coumarate:CoA ligase (4CL)
Bradi2g39420.1	ABC transporter, ATP-binding protein
Bradi4g19457.1	Adenosylhomocysteinase / SAHase
Bradi4g13580.1	Aerolysin toxin, agglutinin domain
Bradi2g21860.1	AMP binding, ADP binding, NADH-cytochrome b5 reductase
Bradi3g47950.1	ANKYRIN Repeat containing protein
Bradi3g59090.1	Aspartic-type endopeptidase activity
Bradi3g10270.1	Bidirectional Sugar Transporter Sweet6-Related
Bradi4g35477.1	Bile acid sodium symporter family protein
Bradi4g16327.1	BSD domain-containing protein
Bradi4g03397.1	BTB/POZ domain (BTB) // NPH3 family (NPH3)
Bradi4g25540.1	BTB9 - Bric-a-Brac, Tramtrack, Broad Complex BTB domain
Bradi1g45400.1	Cellulose Synthase-Interactive Protein 1
Bradi2g34240.1	CESA1 - cellulose synthase
Bradi3g28350.1	CESA7 - cellulose synthase, (BdCESA4)
Bradi1g54250.1	CESA8 - cellulose synthase, (BdCESA3)
Bradi4g30540.1	CESA9 - cellulose synthase, (BdCESA7)
Bradi4g34040.1	CHIT13 - Chitinase family protein precursor
Bradi3g36887.1	Cinnamoyl-CoA reductase
Bradi1g08310.2	COBRA-like protein
Bradi3g16530.1	COMT6, O-methyltransferase
Bradi2g10960.1	Copper ion binding, plastocyanin-like domain
Bradi1g25117.1	CSLF2 - cellulose synthase-like family F
Bradi2g17067.1	Cytochrome b-561
Bradi2g21300.1	Cytochrome P450
Bradi4g36240.1	Endoglucanase
Bradi1g09460.1	Endoglucanase
Bradi3g40820.1	Eukaryotic cytochrome b561 (Cytochrom_B561)
Bradi2g22170.1	Expressed protein
Bradi3g58150.1	Expressed protein
Bradi1g67870.1	Expressed protein
Bradi4g37490.1	F-box domain and kelch repeat containing protein
Bradi1g06290.1	Fasciclin domain containing protein
Bradi2g16560.1	Fasciclin domain containing protein
Bradi1g57040.1	Fasciclin domain containing protein
Bradi4g31130.1	Ferric reductase
Bradi2g12370.1	GDSL-like lipase/acylhydrolase
Bradi3g32180.1	GDSL/SGNH-like Acyl-Esterase
Bradi4g21240.1	Glucuronoxylan 4-O-methyltransferase (GXM)
Bradi4g40400.1	Glucuronoxylan 4-O-Methyltransferase 2-related
Bradi5g15527.1	Glycosyl Hydrolase
Bradi3g54370.1	Glycosyl hydrolase (GH), subfamily GH79
Bradi3g14080.1	Glycosyltransferase, CAZy family GT2
Bradi2g33090.1	Glycosyltransferase, CAZy family GT31
Bradi1g65750.1	Glycosyltransferase, CAZy family GT43
Bradi1g72350.1	Glycosyltransferase, CAZy family GT8
Bradi3g53170.1	Glycosyltransferase family 92 domain

Gene ID	Description
Bradi1g68250.1	GTPase activity
Bradi3g45160.1	Harpin-induced protein 1 domain containing protein
Bradi1g72762.1	Hydrolase, alpha/beta fold family protein
Bradi5g14720.1	Hydroxycinnamoyl-CoA:shikimate/quinate HCT
Bradi2g60310.1	Hydroxyproline-rich Glycoprotein
Bradi1g47767.1	Inorganic H+ pyrophosphatase
Bradi3g37530.1	Ion channel activity, ferric reductase
Bradi2g54680.1	Lacasse 10
Bradi2g54690.1	Laccase 11
Bradi1g03940.1	Leaf senescence related protein
Bradi1g00710.1	Lzipper-MIP1
Bradi2g08790.1	Manganese ion binding, Cupin domain
Bradi4g01200.2	Methyltransferase, 5-methyltetrahydropteroyltriglutamate-h.
Bradi2g47590.1	MYB family transcription factor, BdMYB48, (SWAM1)
Bradi3g50370.1	Myosin II binding, ELMO/CED-12 family protein
Bradi2g15300.2	N-acetylneuraminate acetyltransferase, CAS1 domain
Bradi2g46197.1	NAC Domain Containing Protein 73, (GNRF)
Bradi1g76362.1	Non-specific serine/threonine protein kinase
Bradi4g08917.1	NPH3 family
Bradi2g37000.1	Peroxidase / Lactoperoxidase
Bradi3g49250.2	Phenylalanine ammonia-lyase, BdPAL1
Bradi3g49260.1	Phenylalanine ammonia-lyase, BdPAL2
Bradi1g11040.1	Phosphatidylinositol-4-phosphate 5-kinase
Bradi1g67460.1	Phospholipase A2 activity
Bradi1g54680.1	Protein-cysteine S-palmitoleyltransferase activity
Bradi1g34670.1	Putative glycosyltransferase belonging to CAZy family GT61
Bradi1g75410.1	Putative glycosyltransferase CAZy family GT14
Bradi2g34470.1	Pyrimidine nucleotide sugar transmembrane transporter
Bradi3g33070.1	QWRF motif- containing protein 7
Bradi5g04540.1	RING, subfamily zinc finger (C3HC4-type RING finger)
Bradi2g12150.2	S-adenosylmethionine synthetase
Bradi2g59400.1	Similar to glucuronyltransferase. CAZy family GT47
Bradi2g12710.1	Similar to RRM, putative AtKinesin-13A/Kinesin-13A
Bradi1g35477.1	Strubbelig-receptor family 7
Bradi3g05670.1	Strubbeling-Receptor family 3 precursor
Bradi1g72430.1	Strubbeling-Receptor family 6 precursor
Bradi2g18447.1	Sulfotransferase domain containing protein
Bradi2g55340.1	Transmembrane amino acid transporter protein
Bradi2g10970.1	Tubulin/FtsZ domain containing protein
Bradi1g10150.1	Tubulin/FtsZ domain containing protein
Bradi2g52790.1	Tubulin/FtsZ domain containing protein
Bradi1g31580.1	Ubiquitin-conjugating enzyme
Bradi1g19660.1	Uncharacterized GPI-anchored protein At5g19240 precursor
Bradi2g59410.1	Xylosyltransferase, CAZy family GT47-2
Bradi1g19920.1	Xylulokinase activity, carbohydrate kinase
Bradi2g23300.1	Unknown
Bradi4g28260.1	Unknown
Bradi2g00880.1	Unknown

Table 2.4 **Replicate samples of WT, *GNRF-OE* and *gnrf* mutants for RNA-seq analysis.**

Genotype	Replicate 1	Replicate 2	Replicate 3	Replicate 4	Total samples
GNRF-OE	SR2	SR3	SR5		3
gnrf-1	SR17	SR19	SR24	SR25	4
gnrf-2	SR27	SR28	SR32	SR34	4
gnrf-3	SR35	SR38	SR40		3
gnrf-4	SR45	SR46	SR47	SR49	4
gnrf-6	SR61	SR62	SR63	SR64	4
WT	SR65	SR66	SR67	SR68	4

Table 2.5 **List of a subset of genes differentially expressed in *GNRF-OE* and *gnrf* plants identified by RNA-seq (p<0.01)**

Subcategories

- Genes grouped in the same co-expression cluster (PlaNet)
- Genes differentially expreesed in RNA-seq, microarray analysis, and present in the co-expression analysis (PlaNet)
- Genes of interest (Cell wall formation and flowering pathway)
- Downregulation
- Upregulation

Gene ID	Gene Description	RNAseq Fold *GNRF-OE*	Microarray Fold *GNRF-OE*	RNA-seq	
				gnrf	Fold
Bradi1g21980	MADS transcription factor, BdMAD5	-14.47			
Bradi1g08340	BdAP1, Homologous to Arabidopsis APETALA1 (MADS33/VRN1/FUL1)	-12.67	-53.8	g-1	0.31
Bradi1g48830	BdFTL2, Homologous to Rice Flowering Time Locus T-Like2	-12.48		g-1	0.57
Bradi2g07070	BdFTL1, Homologous to Rice Flowering Time Locus T -Like1	-11.43			
Bradi1g38150	BdFTL12, Homologous to Rice Flowering Time Locus T - Like4	-11.39			
Bradi1g16240	F-Box	-11.00			
Bradi1g59250	MADS transcription factor, BdMAD16 (MADS10/FUL2)	-10.85	-51.3		
Bradi1g08326	MADS BOX PROTEIN, MADS1	-9.96	-33.6		
Bradi3g08301	RNA HELICASE // SUBFAMILY NOT NAMED	-9.06			
Bradi1g13040	BdGLO1C globulin	-7.91			
Bradi1g27910	Peroxidase	-7.80	-11.3		
Bradi1g33016	Unknown	-7.53			
Bradi2g25490	EamA-like transporter family (EamA)	-7.41			
Bradi5g15647	MULE transposase domain (MULE)	-7.32			
Bradi5g26380	Unknown	-7.32			
Bradi5g19120	K18635 - protein SPIRAL1 and related proteins (SPR1)	-7.26			
Bradi4g01506	unknown	-7.13			
Bradi3g05570	Potassium ion transporter	-7.09	-6.2		
Bradi4g34510	PINFORMED-Like auxin efflux carrier	-4.71	-2.5		
Bradi2g23370	Laccase 8	-4.65	-12.6		
Bradi3g51800	MADS TF (MAD42)	-4.65	-2.8		
Bradi3g58560	Copper ion binding	-4.59	-11.6		
Bradi1g69890	MADS11	-4.49	-5.7		
Bradi1g10520	BdPHYA, Phytochrome A	-4.42			
Bradi3g48950	Ammonium transporter	-4.39	-5.0		
Bradi1g66720	Laccase 5	-4.37	-12.3		
Bradi1g57607	KNOX6 Homeobox	-4.35	-4.5		
Bradi3g37850	Potassium ion transporter	-4.03	-4.7		
Bradi1g10047	KNOX2 Homeobox	-3.75	-3.2		
Bradi1g73710	C2C2-Dof family	-3.67	-2.6		
Bradi5g04120	Glycoside hydrolase, !-Expansin	-3.64	-3.9		
Bradi2g17530	Lipid transfer protein	-3.26	-7.8		
Bradi2g17550	Lipid transfer protein	-3.12	-8.5		
Bradi1g12290	Glycosyltransferase, GT47 family	-3.10	-6.8		
Bradi2g30490	Lipid transfer protein	-3.08	-3.8		
Bradi3g30590	Ferulic acid 5-hydroxylase 1 (FAH1)	-3.06	-5.1		
Bradi1g34210	Cation transmembrane transporter	-2.97	-2.3		
Bradi3g26690	BEL1-LIKE homeodomain transcription factor	-2.80	-7.0	g-4, g-6	0.54,0.88
Bradi1g03500	Proton-dependent oligopeptide transporter	-2.79	-5.3		
Bradi5g10640	NAC family, XND1-like	-2.76	-3.6		
Bradi1g64560	Glycosyltransferase, GT34 family, xylosyltransferase	-2.68	-7.6		
Bradi2g54970	Lipid transfer protein	-2.65	-4.9		
Bradi2g34650	Fasciclin-like arabinogalactan protein	-2.63	-8.1		
Bradi4g28000	Sugar transporter	-2.57	-2.7		
Bradi5g17930	Lipid transfer protein	-2.51	-4.2		
Bradi2g00220	Fasciclin-like arabinoogalactan protein	-2.46	-4.6		
Bradi2g17540	Lipid transfer protein	-2.46	-4.4		
Bradi4g13697	Galactoside 2-alpha-L-fucosyltransferase	-2.36	-6.2	g-2	0.98
Bradi2g09690	Peroxidase	-2.31	-6.9		
Bradi2g08310	UDP-galactosyltransferase activity	-2.24	-7.1	g-2, g-6	0.44, 0.46

Bradi1g45190	Hydrogen:amino acid symporter activity	-2.13	-6.3	g-3	0.67
Bradi1g17830	Potassium transporter	-2.08	-2.6		
Bradi1g33827	Glycosylhydrolase, GH6, xyloglucan endotransglycosylase 6	-2.05	-4.9		
Bradi1g78100	Arsenite transport	-1.90	-6.5		
Bradi1g72350	Glycosyltransferase, CAZy family GT8	-1.88			
Bradi3g50067	No apical meristem (NAM) protein, SWN3	-1.85			
Bradi2g54680	Lacasse 10	-1.81			
Bradi4g06317	MYB transcrition factor, BdMYB1	-1.75			
Bradi4g21790	Proton-dependent oligopeptide transporter	-1.73	-5.0		
Bradi3g16530	COMT6, O-methyltransferase	-1.71			
Bradi1g21800	Sugar transporter	-1.68	-4.6		
Bradi2g20840	Peroxidase	-1.66	-5.9	g-2	0.83
Bradi4g30540	CESA9 - cellulose synthase, (BdCESA7)	-1.65	-2.8	g-4	0.54
Bradi3g39800	High affinity sodium:dicarboxylate symporter activity	-1.61	-7.8		
Bradi3g28350	CESA7 - cellulose synthase, (BdCESA4)	-1.60	-3.0		
Bradi1g25117	CSLF2 - cellulose synthase-like family F	-1.58	-3.8		
Bradi1g25937	EamA-like transporter	-1.58	-3.1		
Bradi3g51250	Mechanosensitive ion channel	-1.58	-2.4		
Bradi5g24170	Sulfate transporter	-1.58	-2.4		
Bradi3g16515	MYB59/LHY	-1.55	-2.9		
Bradi2g36910	PMT Transferase family	-1.55		g-4, g-6	0.42, 0.55
Bradi2g53580	Glycoside hydrolase, ! -Expansin	-1.54	-3.9		
Bradi2g49912	CELLULOSE SYNTHASE A CATALYTIC SUBUNIT 8, CESA8	-1.49		g-3	0.75
Bradi4g44810	Lacasse 29	-1.49	-4.3	g-2	0.97
Bradi3g28920	UDP-glucuronic acid transporter	-1.48	-3.6		
Bradi2g48690	MADS transcription factor, MADS28	-1.48	-4.3		
Bradi1g21990	Glycosyltransferase, GT75 family	-1.46	-4.8		
Bradi3g21480	Homeobox family	-1.46	-3.8		
Bradi3g04080	Glycosylhydrolase, GH9 family glycosyl hydrolase 9B8	-1.44	-4.0		
Bradi2g55340	Transmembrane amino acid transporter protein	-1.43			
Bradi3g49250	Phenylalanine ammonia-lyase, BdPAL1	-1.43			
Bradi1g69770	Aluminum activated citrate transporter	-1.42	-3.4		
Bradi1g73170	Sucrose transporter	-1.37	-2.7		
Bradi3g06480	CAD1	-1.35			
Bradi1g12710	Glycosyl hydrolase, GH10 family	-1.31	-4.8		
Bradi2g59410	Xylosyltransferase, CAZy family GT47-2	-1.30	-3.6		
Bradi1g71990	MAX DIMERIZATION, MAD	-1.30	-3.2	g-6	1.38
Bradi2g08790	Manganese ion binding, Cupin domain	-1.30			
Bradi5g14720	Hydroxycinnamoyl-CoA:shikimate/quinate hydroxycinnamoyltransferase HCT	-1.28			
Bradi4g33370	WRKY60	-1.25	-2.2		
Bradi3g16130	Sterol-transporting ATPase activity/ ABC transporter	-1.21	-2.7		
Bradi2g24910	Hydrogen:amino acid symporter activity	-1.21	-2.5	g-1	0.45
Bradi1g34140	ATPase-like zinc transporter	-1.21	-3.3		
Bradi3g36887	Cinnamoyl-CoA reductase	-1.16			
Bradi4g21240	Glucuronoxylan 4-O-methyltransferase (GXM)	-1.14	-3.7		
Bradi3g34567	WRKY13, DISEASE RESISTANCE PROTEIN-LIKE-RELATED	-1.13	-2.5		
Bradi3g33070	QWRF motif- containing protein 7	-1.11		g-4	0.56
Bradi2g10970	Tubulin/FtsZ domain containing protein	-1.11			
Bradi1g59880	COBRA-like protein	-1.06	-3.6	g-4	0.40
Bradi1g59830	L-amino acid transmembrane transporter activity	-1.05	-2.3		
Bradi4g25540	BTB9 - Bric-a-Brac, Tramtrack, Broad Complex BTB domain	-1.05			
Bradi5g04540	RING, subfamily zinc finger (C3HC4-type RING finger) family protein	-1.03			
Bradi2g12150	S-adenosylmethionine synthetase	-1.00			
Bradi4g33490	Fasciclin-like arabinoogalactan protein	-0.97	-3.8	g-2	0.59
Bradi1g03940	Leaf senescence related protein	-0.94			
Bradi1g24880	Laccase	-0.92	-4.3		
Bradi3g05750	4-Coumarate:CoA ligase (4CL)	-0.89	-3.1		
Bradi2g23300	Expressed protein	-0.88			
Bradi1g38297	Peroxidase	-0.87	-3.2	g-3	0.60
Bradi2g21300	Cytochrome P450	-0.81			
Bradi2g02320	Glycosylhydrolase, GH10 family	-0.81	-3.7		
Bradi1g13910	Homeobox transcription factor, BdHB5, (HB7)	-0.79			
Bradi5g22920	bHLH transcription factor, BdbHLH148	-0.79	-2.2		
Bradi1g67460	phospholipase A2	-0.78			
Bradi1g16540	BdGT43-1, Xylosyltransferase, CAZy family GT43	-0.77			
Bradi3g13727	NAC DOMAIN-CONTAINING PROTEIN 12, SWN8	-0.72			
Bradi3g39420	Caffeoyl CoA 3-O-methyltransferase, CCoACOMT1	-0.71	-2.8	g-3	0.39
Bradi4g40400	GLUCURONOXYLAN 4-O-METHYLTRANSFERASE 2-RELATED	-0.69			

Bradi1g27920	PEROXIDASE 53-RELATED	-0.68	-5.3	g-2, g-6	0.88,1.29
Bradi1g00710	Lzipper-MIP1	-0.66			
Bradi3g32180	GDSL/SGNH-like Acyl-Esterase	-0.64			
Bradi1g50057	No apical meristem (NAM) protein, SWN7	-0.64		g-4	0.47
Bradi4g08130	Sterol-transporting ATPase activity	-0.63	-2.1	g-2, g-6	0.46, 0.84
Bradi3g51280	Tetracycline:hydrogen antiporter activity	-0.61	-3.7	g-2	0.84
Bradi3g47950	ANKYRIN Repeat containing protein	-0.58			
Bradi2g47590	MYB family transcription factor, BdMYB48, (SWAM1)	-0.55	-3.0	g-2, g-3,	0.8, 0.32, 0.43
Bradi3g14080	Glycosyltransferase, CAZy family GT2	-0.53		g-2	0.38
Bradi3g53170	F2103.9 protein-related, Glycosyltransferase family 92 domain	-0.51		g-6	0.27
Bradi3g32390	Tetracycline transporter	-0.51	-2.5		
Bradi1g67870	Expressed protein	-0.50		g-2	0.49
Bradi1g31320	4-Coumarate:CoA ligase (4CL)	-0.47	-2.4		
Bradi5g17990	ATP dependent copper transporter	-0.47	-3.2		
Bradi2g52790	Tubulin/FtsZ domain containing protein	-0.47			
Bradi4g36240	Endoglucanase	-0.45			
Bradi1g10150	Tubulin/FtsZ domain containing protein	-0.43			
Bradi4g01200	Amino acid biosynthesis, homocysteine methyltransferase	-0.43			
Bradi2g15300	N-acetylneuraminate acetyltransferase, CAS1 domain	-0.39		g-2	0.35
Bradi2g59400	Similar to glucuronyltransferase. CAZy family GT47	-0.39			
Bradi1g54680	Protein-cysteine S-palmitoleyltransferase activity	-0.37		g-4,g-6	0.29
Bradi3g50370	Myosin II binding, ELMO/CED-12 family protein	-0.30		g-2	0.46
Bradi2g10960	Copper ion binding, plastocyanin-like domain	-0.30			
Bradi3g58150	Expressed protein	-0.27			
Bradi2g23710	MYB transcrition factor, BdMYB36, SWAM9	0.37			
Bradi2g39420	ABC transporter, ATP-binding protein	0.37			
Bradi1g72762	Hydrolase, alpha/beta fold family protein	0.39			
Bradi4g03397	BTB/POZ domain (BTB) // NPH3 family (NPH3)	0.45		g-6	-0.35
Bradi1g08310	COBRA-like protein	0.47			
Bradi1g54250	CESA8 - cellulose synthase, (BdCESA3)	0.47			
Bradi3g33170	Phytochrome interacting factor like, BdbHLH90, WRIB	0.51			
Bradi2g34240	CESA1 - cellulose synthase	0.51			
Bradi1g20250	MYB transcrition factor, BdMYB4, SWAM6	0.55			
Bradi4g28260	Glycosyltransferase, GT77 family, Extensin	0.59	-2.5		
Bradi1g17700	bZIP transcription factor, BdbZIP6, SWIZ	0.61			
Bradi3g55890	S-adenosylmethionine-dependent methyltransferase, COMT7	0.66			
Bradi3g45160	Harpin-induced protein 1 domain containing protein	0.67			
Bradi2g36730	MYB transcrition factor, BdMYB41, SWAM4	0.75			
Bradi1g43680	PEROXIDASE	0.75	-4.0		
Bradi1g29060	Cellulose synthas, BdCESA5	0.76			
Bradi2g12370	GDSL-like lipase/acylhydrolase	0.81			
Bradi1g48370	Caffeoyl-CoA 3-O-methyltransferase, CCoACOMT3	1.18			
Bradi5g21550	Cinnamyl alcohol dehydrogenase (CAD), CAD7	1.18			
Bradi1g32870	Peroxidase	1.27	-3.1		
Bradi2g11080	MYB transcrition factor, BdMYB27	1.31			
Bradi2g46197	NAC Domain Containing Protein 73, GNRF	2.34		g-6	-9.02

CHAPTER 3

SECONDARY WALL ASSOCIATED MYB4 (*SWAM4*) IS A TRANSCRIPTIONAL ACTIVATOR OF CELL WALL BIOSYNTHESIS IN *BRACHYPODIUM DISTACHYON*

3.1 Introduction

In higher eukaryotes, the family of MYB transcription factors (from the avian myeloblastosis virus), characterized for sharing the MYB DNA-binding domain, are involved in cell differentiation and proliferation. In the eudicot plant *A. thaliana*, MYB proteins containing two highly conserved DNA-binding domains called R2R3 domain are related to the c-Myb proteins and have been functionally characterized by genetic screenings (Stracke et al., 2001). Several MYB proteins determined as transcriptional regulators of secondary cell wall biosynthesis were first identified by gene expression pattern. Genes overexpressed, mutated and fused with a dominant repressor motif were used to characterize MYB and NAC transcription factors as master regulators and target genes involved in the regulation of cell wall biosynthetic genes. *SND1* (SECONDARY WALL-ASSOCIATED NAC DOMAIN PROTEIN1) was identified as a regulator of *AtMYB46* and a master regulator itself of 11 more transcription factors including the NAC and MYB proteins *AtSND2*, *AtSND3*, *AtMYB103*, *AtMYB85*, *AtMYB52*, *AtMYB54*, *AtMYB69*, *AtMYB42, AtMYB43*, *AtMYB20*, and the homeobox protein *AtKNAT7* (Zhong et al., 2008; Zhong et al., 2007a). Co-regulation between SND and MYB transcription factors was also observed (Yao et al., 2012). Two closely related proteins *AtMYB58* and *AtMYB63* are expressed in vessels and fibers where secondary cell wall formation occurred

and identified as positive regulators of the lignin biosynthetic genes (Zhou et al., 2009). A study using an estradiol-inducible system in *B. distachyon* revealed that *BdSWN5*, a NAC transcription factor, resulted in ectopic secondary cell wall formation and up-regulation of *BdCESA4,* as well as *BdXCP1*and *BdMYB1,* an ortholog to *A. thaliana MYB46* (Valdivia et al., 2013). The *B. distachyon* gene *SWAM1* (*SECONDARY WALL ASSOCIATED MYB1*) was identified as a MYB activator of lignin and cellulose biosynthetic genes that promotes cell wall thickening (Handakumbura et al., 2018). In *Pinus taeda*, the *PtMYB4* was found to regulate lignification when lignin deposition was increased after overexpressing *PtMYB4* in tobacco plants (Patzlaff et al., 2003a).

Transcriptional regulation of cell wall biosynthesis in grasses has been less studied, and due to distinct secondary cell wall compositions between eudicots and monocots, there is likely a difference in the transcriptional network as well (Handakumbura and Hazen, 2012). Comparative genomic analysis revealed that R2R3 MYB secondary cell wall regulators in *A. thaliana* are highly conserved in grasses (Zhao and Bartley, 2014). Functional orthologs of well-characterized MYB transcription factors in *A. thaliana* have been characterized in monocots and conifers. The rice *OsMYB46* and maize *ZmMYB46* orthologs of *AtMYB46* were overexpressed in *A. thaliana,* and they activated secondary wall biosynthesis (Zhong et al., 2011). *OsMYB58*/*63* and *OsMYB55*/*61* were identified as regulators of secondary cell wall formation in rice (Hirano et al., 2013a; Hirano et al., 2013b). Overexpression of *PvMYB4* in transgenic switchgrass and tobacco resulted in the downregulation of lignin biosynthetic genes, a reduction in lignin content, and an increase in saccharification efficiency. Comparison of transcriptional regulation of secondary cell wall biosynthesis between eudicots and grasses reinforces the idea that certain structural features of the regulation are conserved but that differences in regulatory functions also exist.

(Shen et al., 2012).

Well-conserved *cis*-regulatory AC-type elements or the SMRE (secondary wall MYB responsive element) binding element are targets of specific cell wall regulating MYB proteins (Noda et al., 2015; Zhong and Ye, 2012, 2014; Zhong et al., 2015). AC elements were identified as targets of the pine protein *PtMYB1* by transcriptional activation assays using promoters of lignin biosynthetic genes in vitro (Patzlaff et al., 2003b). The SMRE, ACC(A/T)A(A/C)(T/C), interacts with several MYBs *in vitro* and *in planta* as well (Shen et al., 2012; Zhong et al., 2011; Zhong and Ye, 2012).

AtMYB61 encodes a MYB protein that regulates stomatal aperture and xylem cell structure. This pleiotropic function was identified in overexpression *35S:MYB61* and loss-of-function mutant *atmyb61* mutant plants of *A. thaliana* (Liang et al., 2005; Newman et al., 2004; Romano et al., 2012). *AtMYB61* target binding sites have been recognized (Prouse and Campbell, 2013; Romano et al., 2012) and orthologous genes were identified and functionally associated with diverse developmental processes including secondary cell wall formation (Huang et al., 2015; Li et al., 2018; Matias-Hernandez et al., 2017; Ye et al., 2018). This chapter describes the functional characterization of Bradi2g36730 *SWAM4* (*SECONDARY WALL ASSOCIATED MYB4*), a *B. distachyon* ortholog of *AtMYB61* and *OsMYB61b,* that encodes a protein closely related to SWAM1 protein. Overexpression, mutant, and dominant repressor plants were studied with genetic and biochemical approaches to gain insight into the function of SWAM4.

3.2 Materials and Methods

3.2.1 Phylogenetic analysis

Corresponding genome annotations from NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi), TAIR BLAST v2.2.8 (https://www.arabidopsis.org/Blast/), and Phytozome v12.1 (https://phytozome.jgi.doe.gov) servers were used to download *SWAM4* (Bradi2g36730) and MYB homologous protein sequences. An extensive phylogenetic tree was constructed with 67 protein sequences including six *SWAM* genes from *B. distachyon;* from this arrangement, 24 protein sequences grouped on a clade that includes SWAM4 were selected to construct a smaller phylogenetic tree. Protein sequences from the following plants were used: *Arabidopsis thaliana*, *Capsella rubella*, *Brachypodium distachyon*, *Sorghum bicolor, Oryza sativa*, *Setaria viridis*, *Ananas comosus*, *Eucalyptus grandis*, *Pabies glauca, Pinus taeda, Pseudotsuga menziessi, Populus thrichocarpa, Medicago truncatula.* Multiple sequence alignment with the iterative refinement method L-INS-I were aligned using MAFFT interphase (Katoh et al., 2017). The unrooted phylogenetic trees were constructed using the neighbor-joining method, a bootstrap resampling value of 1000, and visualized on Phylo.io (http://phylo.io) (Robinson et al., 2016)

3.2.2 Plant material and growth conditions

B. distachyon accession Bd21-3 was used as genetic background (Vogel and Hill, 2008). Seed was imbibed in water at 4°C for ten days and planted in pots with a mix 3:1 ratio of potting mix (Sun Gro Sunshine #8 / Fafard 2 Mix, Burton, OH) and turface (Pro's choice Sports Field Products, Chicago, IL). Treatment with Gnatrol (Valent Bioscience Corporation, Libertyville,

IL) was made 24 hours prior to planting. Plants were grown in long day conditions 20-hour light/4-hour dark at 26°C and 18°C respectively in a Conviron growth chamber (Controlled Environments Ltd, Manitoba, Canada) at a fluence rate of 220 μmol m^{-2} sec^{-1} and 68% relative humidity.

3.2.3 *SWAM4* overexpression (*SWAM4-OE*) and *SWAM4* dominant repressor (*SWAM4-DR*) plants

To create *SWAM4* overexpression plants, the full-length coding region of *SWAM4* (Bradi2g36730) was cloned into the modified version of the destination vector pOL001. This vector contains the constitutive *ZmUbi* promoter and the hygromycin resistant gene. The construct was transformed into *B. distachyon* by *A. tumefaciens*-mediated transformation (Vogel et al., 2006; Vogel and Hill, 2008). The same vector was used to generated dominant repressor plants by cloning a 39-nucleotide dominant repressor motif, CRES (Hiratsu et al., 2003; Kagale and Rozwadowski, 2011) adjacent to the *SWAM4* full-length coding sequence. Seed previously generated from *SWAM4-OE* and *SWAM4-DR* of three and six independent events respectively were planted. For genotyping, leaf tissue was used to extract DNA, and PCR amplification of the hygromycin resistant gene was performed as described in chapter 2. Additional PCRs to confirm the genotyping to amplify the junction between *ZmUbi-SWAM4* and *SWAM4-CRES* were carried out. Both junction fragments were amplified under the following conditions: 95°C for 30 s followed by 40 cycles of 95°C for 20 s, 52°C for 45 s, and 68°C for 30 s, with a final extension step at 68°C for 5 min. Primers are listed in Table 3.1

3.2.4 *swam4-1*, *swam4-2* mutant allele identification

Two mutant lines with individual non-synonymous mutations in *SWAM4* coding sequence were selected from a sodium azide (NaN3) mutagenized TILLING (Targeting Induced Local Lesion IN Genome) collection (Dalmais et al., 2013). Seeds from both lines were planted, and DNA was extracted from leaf tissue as described (Handakumbura et al., 2013). *swam4-1* and *swam4-2* mutant alleles were identified by PCR amplification and sequence confirmation. Briefly, a 460 bp fragment was amplified using Taq DNA polymerase (New England Biolabs) under the following conditions: 95°C for 30s followed by 40 cycles of 95°C for 20 s, 50°C for 45 s, and 68°C for 30 s, with a final extension step at 68°C for 5 min. PCR amplicons were sent to Macrogen USA for purification and sequencing (Macrogen, Boston, MA).

3.2.5 Plant phenotyping

Flowering time data was collected from 3-14 independent events and was determined when the inflorescence emerged from the flag leaf at BBCH stage 5-1 (Hong et al., 2011). To estimate above ground-biomass, data from stem and seed weight were separately collected. Measurements were made from 3 plants per genotype at senescence at BBCH stage 9. Plant height data was from 3-6 plants per genotype at both BBCH stages 5-1 and 5-9. Error bars correspond to means ± SEM (Standard Error of the Mean). Two-tailed Student's *t*-tests were performed, and significance was set to $p<0.05$ and $p<0.01$.

3.2.6 Cross sections, histochemical analysis of stem lignification and cell wall measurements

The first internode of the tallest stem of the plants at complete senescence was sectioned using a vibrotome Leica VT 1200S (Leica Biosystems). Stem cross sections of 55 μm were stained with phloroglucinol-ethanol as described (Matos et al., 2013). The images were captured using a PixeLINK 3 MP camera (PixeLINK), 4x, 10x and 20x microscope objective lenses of a Nikon Eclipse E200MV microscope (Nikon). Measurements of stem area, stem perimeter, interfascicular fiber and xylem wall thickness of cells were made in ImageJ using the images captured at 20x. Three independent cross sections were analyzed with a total of forty-five thickness measurements per plant as previously described (Handakumbura et al., 2018).

3.2.7 Acetyl bromide soluble lignin measurement

Fully senesced stem tissue from three plants per genotype at BBCH stage 9 (Hong et al., 2011) was pulverized and processed for lignin content as previously described (Foster et al., 2010; Handakumbura et al., 2018). A grass coefficient factor of 17.75 was used to calculate the percentage of lignin in the samples using the defined formula (Foster et al., 2010).

3.2.8 RNA extraction and Q-RT-PCR analysis

For gene expression analysis by Q-RT-PCR, RNA was extracted from stem tissue at the beginning of heading when the inflorescence began to emerge from the flag leaf, at stage 5-1 BBCH-scale (Hong et al., 2011). Stem tissue was collected from three biological replicates of

each genotype: WT, three independent events of *SWAM4-OE*, and three of *SWAM4-DR* mutant lines. Samples were collected into a 1.5 ml microcentrifuge tube containing two metal beads at ZT22 during the 4-hour dark cycle. Tissue was flash-frozen in liquid nitrogen and ground in a ball mill grinder. Total RNA was extracted using the RNeasy Plant mini kit (Qiagen, Gaithersburg, MD) and treated with RNase-Free DNase set (Qiagen). RNA was resuspended in 30 µl DEPC-treated water.

cDNA was synthesized from 200 ng of total RNA by using the SuperScript III First-Strand synthesis SuperMix for Q-RT-PCR (Life Technologies). cDNA was diluted four-fold with RNase-free water (Qiagen) and used to make a pool of like samples. To confirm primer efficiency, serial dilutions of the pool were used to make standard curves. QuantiPrime primer design webtool was used for qPCR primer design (Arvidsson et al., 2008), and synthesized by Fisher (FisherScientific). The Quantitative PCR reactions were carried out in triplicated using 1 µl of diluted cDNA and the QuantiFast SYBR Green PCR Kit (Qiagen). Real-time amplification was performed in an Eppendorf Realplex2 Mastercycler, and cycling conditions were: 95°C for 2 min, 40 cycles of 95°C for 15s, 60°C for 15s, and 68°C for 20s, followed by a melting curve analysis. Transcript levels were determined by normalizing threshold cycle (Ct) values from target genes to the reference genes *BdACTIN7* and *BdGAPDH* (Hong et al., 2008). Primers are listed in Table 3.1

3.2.9 Yeast-one-hybrid assay

Three overlapping fragments of the *SWAM4* promoter (from -1470 bp to -391 bp upstream of the transcription start site) and one fragment of *SWAM4* 5'UTR region (-390 bp -1

bp) were cloned into pENTR/D-TOPO (Invitrogen, Life technologies), and subsequently recombined independently into a Gateway compatible pLUC destination vector (Bonaldi et al., 2017; Pruneda-Paz et al., 2014). All constructs containing the designated fragment 1 (-1470-1056bp), fragment 2 (-1211-744bp), fragment 3 (-856-391) and fragment 4 (-390-1bp) were linearized by restriction digestion with StuI (New England Biolabs) and transformed in yeast as previously described (Taylor-Teeples, 2015). After the non-self-activation test, yeast colonies were transformed with pDEST22:*GNRF*, pDEST22:*SWAM1* and pDEST22:*SWAM4* and pDEST22-empty vector constructs from the sub-library collection. Fold change relative to empty vector was calculated as described in Chapter 2. Error bars correspond to means ± SEM (Standard Error of the Mean). Two-tailed Student's *t*-tests were performed, and significance was set to $p<0.05$ and $p<0.01$.

3.2.10 RNA extraction and RNA-seq analysis

For RNA-seq transcriptome analysis, stem tissue from four plants each of WT, *SWAM4-OE*, *SWAM4-DR*, *swam4-1*, and *swam4-2* lines were individually taken at the end of heading when inflorescence fully emerged, stage 5-9 BBCH-scale (Hong et al., 2011). Samples were collected at ZT22 during the 4-hour dark cycle. Stem tissue was flash-frozen in liquid nitrogen and ground as described above. Total RNA was extracted using the RNeasy Plant mini kit (Qiagen) and treated with RNase-Free DNase set (Qiagen). RNA samples were resuspended in 50 µl DEPC-treated water, and RNA concentration was estimated by Qubit RNA BR assay kit (ThermoFisher, life technologies, Waltham, MA) in a Qubit 2.0 fluorometer (Life technologies) following manufacturer's recommendation.

3.2.11 RNA Library construction and sequencing analysis

Quality assessment of RNA was performed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Lexington, MA). Samples with an RNA integrity number (RIN) above 7.0 were selected for library construction and sequencing. Seventeen mRNA libraries were created by poly A selection and sequenced on the Illumina NovaSeq 6000 system -flow cell type S2 2 x 100bp. Library preparation and sequencing was performed at Yale Center for Genome Analysis (YCGA Yale-West campus, Orange, CT) using the Illumina platform. RNA-seq analysis was performed as described in chapter 2, methods section 2.2.14

3.3 Results

3.3.1 *SWAM4* is an ortholog of *OsMYB61b* in rice and is closely related to *AtMYB50* and *AtMYB61*.

SWAM4 (*Bradi2g36730*) is an R2R3-type MYB transcription factor included in the phylogenetic analysis of SWAM1 and SWAM-related proteins in *B. distachyon* (Handakumbura et al., 2018). A broad phylogenetic tree illustrating amino acid sequence similarity was constructed with 67 protein sequences by using the MAFFT sever and Phylo.io visualization application. Similar to the previous SWAM phylogenetic tree presented (Handakumbura et al., 2018), three clades were observed; however, a different arrangement was obtained with the clade that contains SWAM4 amino acid sequence (Figure 3.1A). Twenty-five proteins were used to

construct a second phylogenetic tree to determine protein similarity neighboring SWAM4 (Figure 3.1B). *SWAM4* is most similar to AtMYB50 and AtMYB61, well-characterized with pleiotropic functions (Liang et al., 2005; Matias-Hernandez et al., 2017; Newman et al., 2004; Prouse and Campbell, 2013; Romano et al., 2012) in *C. rubella*, *M. truncatula*, and two uncharacterized genes from *P. trichocarpa* (Figure 31.B). From the monocot plants, *SWAM4* orthologous genes were found in *B. distachyon* (*BdMYB27*), *S. bicolor, O. sativa*, and *S. viridis.* *OsMYB61a* (*Os01g18240*) was found closely related to *SWAM4*, nonetheless the orthologous gene of *SWAM4* in rice was identified as *OsMYB61b* (*Os05g04820*). Interestingly, both genes have been described as positive regulators of cell wall biosynthesis in rice (Huang et al., 2015; Ye et al., 2018).

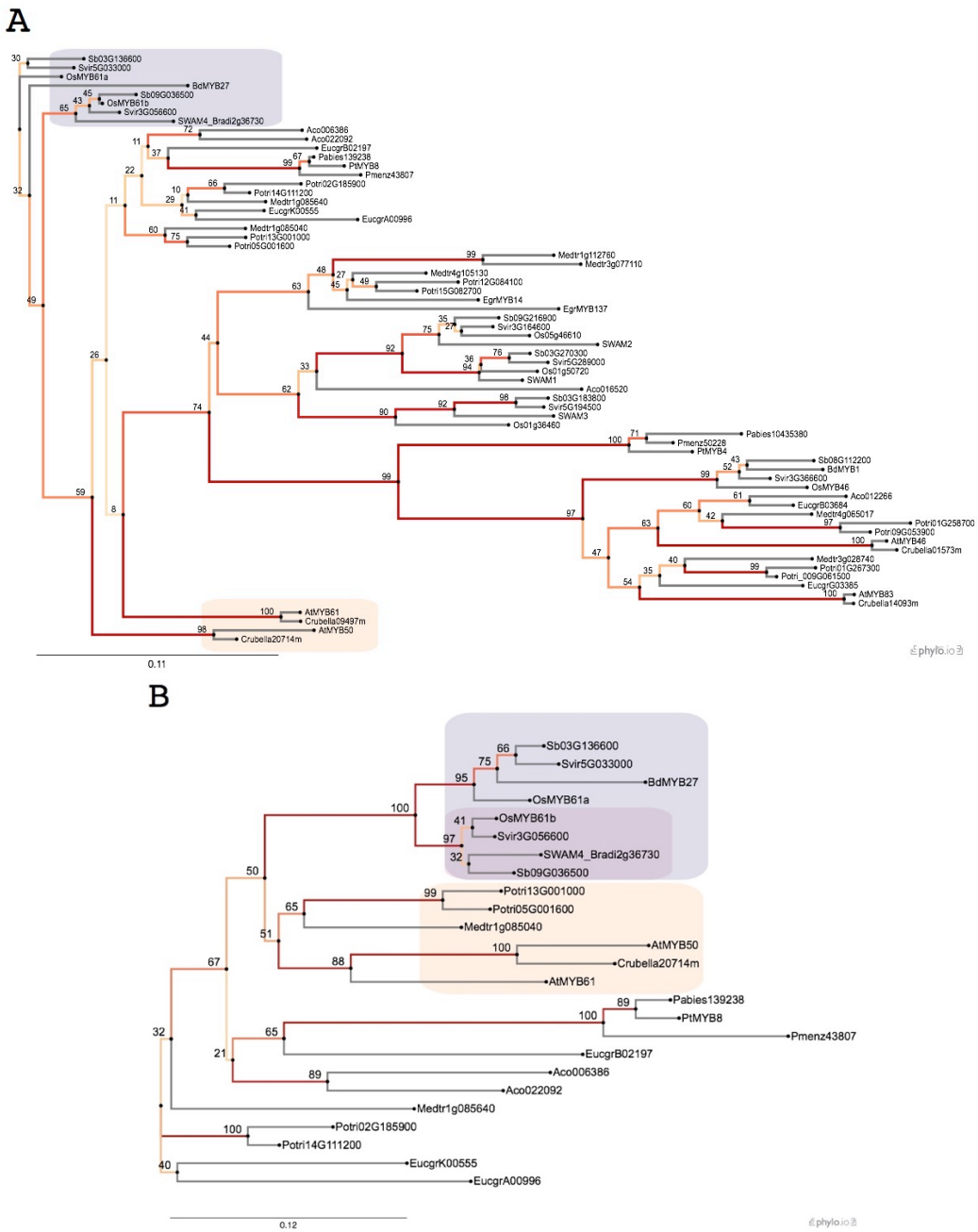


Figure 3. 1 **Phylogeny of *SWAM4* (Bradi2g36730).** SWAM4 phylogeny illustrating amino acid sequence similarity. (A) Related proteins from the R2R3 MYB subgroup. (B) *SWAM4* and *BdMYB27* (purple), and a small cluster lacking SWAM proteins (ivory-color) Sequence similarity between the eudicot plants *Arabidopsis thaliana*, *Capsella rubella*, and the monocot plants *Brachypodium distachyon*, *Sorghum bicolor*, *Oryza sativa*, *Setaria viridis*, *Ananas comosus*, *Eucalyptus grandis*, *Pabies glauca*, *Pinus taeda*, *Pseudotsuga menziessi*, *Populus trichocarpa*, *Medicago truncatula*. Amino acid sequences were aligned with MAFFT server for multiple sequence alignment. Iterative refinement method L-INS-I**,** neighbor-joining method and bootstrap resampling value of 1000 were applied. Color scheme of the branches corresponds to the bootstrap support numbers that indicate similarity. The Phylo.io application was used for visualization.

3.3.2 *SWAM4* transcript is relative abundant in stem tissue and it is clustered in a co-expression arrangement that contains putative cell wall-associated gene

SWAM4 was identified as part of a collection of genes co-regulated with cellulose and lignin biosynthetic genes in *B. distachyon* (Handakumbura et al., 2018; Handakumbura et al., 2013). Transcript abundance in the stem was approximately 27-fold and two-fold greater than leaf and root respectively (Figure 3.2A). Consistent with a putative role for cell wall formation, regulators of cell wall thickening are highly expressed in stems, where fundamental metabolic functions and storage of sugars occur (Jensen and Wilkerson, 2017). The co-expression analysis showed that *SWAM4* is part of a 75 gene co-expression cluster (Table 3.2).

The cluster includes the cellulose synthase gene *BdCESA9,* a Fisciclin-like arabinogalactan protein, the homeobox protein knotted-1 protein (*KNOB7*), and the *BdMYB1* transcription factor (Handakumbura et al., 2013; Valdivia et al., 2013); these genes are associated with cell wall biosynthesis.

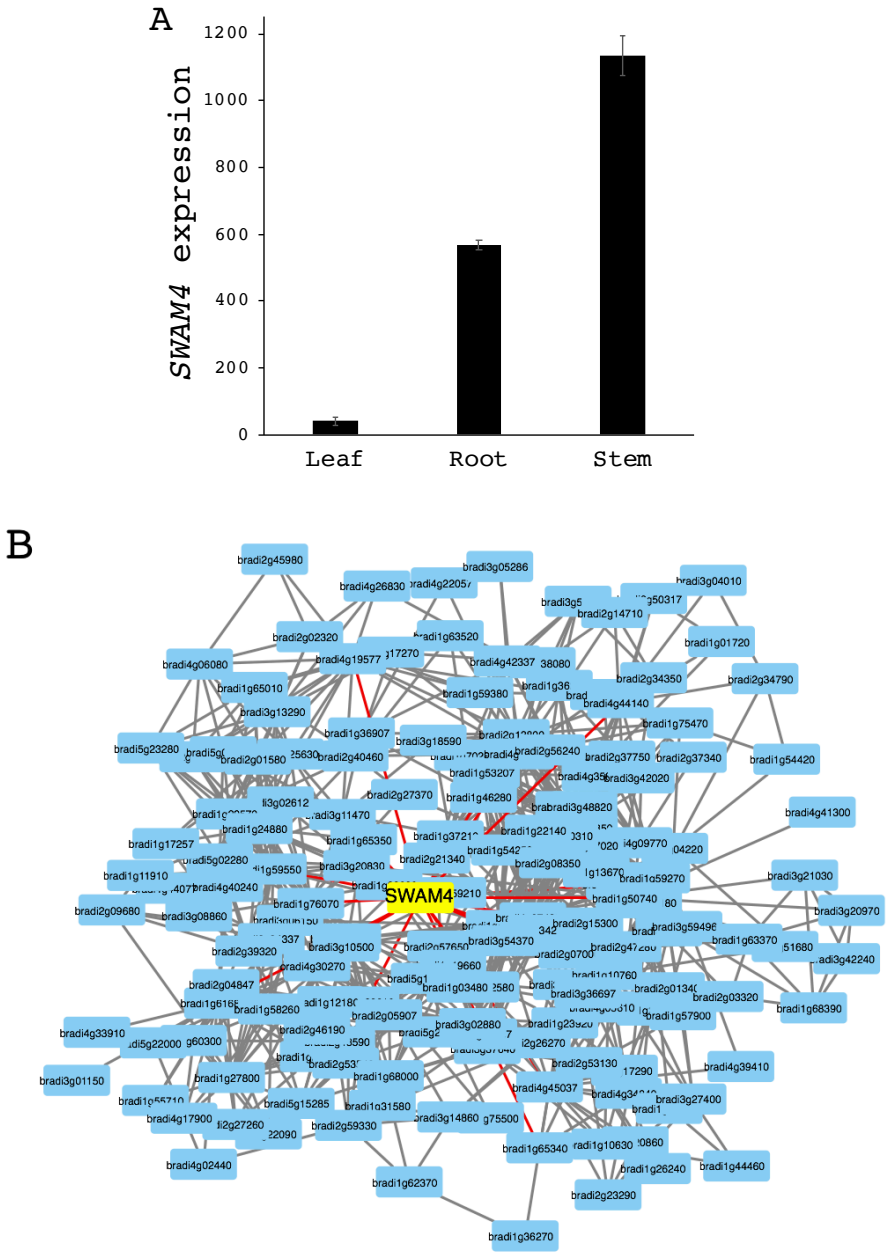


Figure 3. 2 **Relative expression and co-expression analysis of SWAM4.** (A) Relative transcript abundance of *SWAM4* in leaf, root and stem tissue in a whole-genome tiling microarray. SEM (Standard Error of the Mean) of three biological replicates. (B) PlaNet co-expression neighborhood of SWAM4.

3.3.3 Overexpression and dominant repressor *SWAM4 t*ransgenics result in developmental changes

To investigate the function of *SWAM4*, *SWAM4* overexpression plants were generated by cloning the full-length coding region of *SWAM4* downstream of the *ZmUbi* promoter in the vector pOL001(Vogel et al., 2006). Agrobacterium-mediated transformation protocol was used to transform the calli of *B. distachyon* (Vogel and Hill, 2008*)*, and three independent events were used. *SWAM4* overexpression construct (*SWAM4-OE*) plants did not show an apparent phenotypic change compared to WT controls during growth (Figure 3.3A); however, these transgenic plants flowered significantly earlier (Figure 3.5A) and were shorter compared to WT (Figure 3.5B). Even though the two transformation events *SWAM4-OE2* and *SWAM-OE3* appeared similar to WT and *SWAM4-OE3* developed more stems, all independent events showed higher total biomass compared to WT with a significative increment in stem weight (Figure 3.5C).

Two homozygous lines from a TILLING (Targeting Induced Local Lesion IN Genome) collection mutagenized with sodium azide (NaN_3) were genotyped, and non-synonymous mutations were identified in *SWAM4* (Dalmais et al., 2013). These lines were used to investigate *SWAM4* function. The *swam4-1* allele was identified carrying a non-synonymous mutation that replaces a glycine (G) with glutamic acid (E) in amino-acid position 287 of the protein. The *swam4-2* allele presented a non-synonymous mutation that replaces threonine (T) with isoleucine (I) in the position 366 of SWAM4 protein. Both lines were confirmed to be homozygous mutants. *swam4-1* and *swam4-2* plants were initially shorter than WT controls (Figure 3.4A);

however, the mutant plants continued growing and reached a relatively normal height compared to WT plants at flowering and senescence stage. In contrast to *SWAM4-OE* plants, *swam4-1* and *swam4-2* flowered significantly later than for WT plants (Figure 3.5A). Total biomass was also affected in *swam4* mutants; both stem and seed weight were significantly higher than WT plants (Figure 3.5C). To investigate if the phenotypes observed in the *swam4* mutants might be related to the lack of functionality of other proteins than *SWAM4*, a review of additional genes mutated in the lines was conducted. The *swam4-1* mutant also has a splice site donor mutation in *VRN1/FUL* (Bradi1g08340), which is homologous to Arabidopsis *APETALA1. VRN1* is induced by cold treatment, and its expression is high after vernalization. In wheat, *VRN1* expression promotes flowering by downregulating *VRN2*, a flowering repressor, but is not indispensable for flowering. While a *vrn-1* null mutant in wheat experienced delayed flowering (Chen and Dubcovsky, 2012), a line overexpressing *BdVRN1* flowered rapidly without vernalization (Ream et al., 2014). The delayed flowering phenotype in *swam4-1* plants might be caused by the mutation of *BdVRN1*; however, the fact that *SWAM4-OE* and *swam4-2* show early and late flowering phenotypes, respectively, still suggests that the non-synonymous mutation in *SWAM4* might cause the phenotype. An inspection of other genes mutated in the *swam4-2* line indicated that the line contains a mutation in an intron of the homeobox gene *KNOB7*.

3.3.4 Dominant repressor *SWAM4* plants are severely dwarfed

To further investigate the function of *SWAM4*, dominant repressor (*SWAM4-DR*) lines were created to suppress the expression of *SWAM4* target genes. *SWAM4-DR* lines were generated by overexpressing the full-length coding region under the *ZmUbi* promoter fused to a 39-base pair

dominant repressor sequence (Hiratsu et al., 2003). Five independent events were used, and all *SWAM4-DR* plants were dramatically shorter than control plants (Figure 3.3C-D). Leaves of *SWAM4-DR* plants were pointed upwards (Figure 3.3C) but turned curly when plants were fully senesced (Figure 3.3C). *SWAM4-DR* plants were significantly shorter than control plants at both flowering and senescent stages (Figure 3.5B). As expected, total biomass of *SWAM4-DR* plant was significantly reduced, though it was due to the weight of the stems, not changes in seed weight (Figure 3.5C). Flowering time was not altered in *SWAM4-DR* plants compared to WT controls (Figure 3.5A).

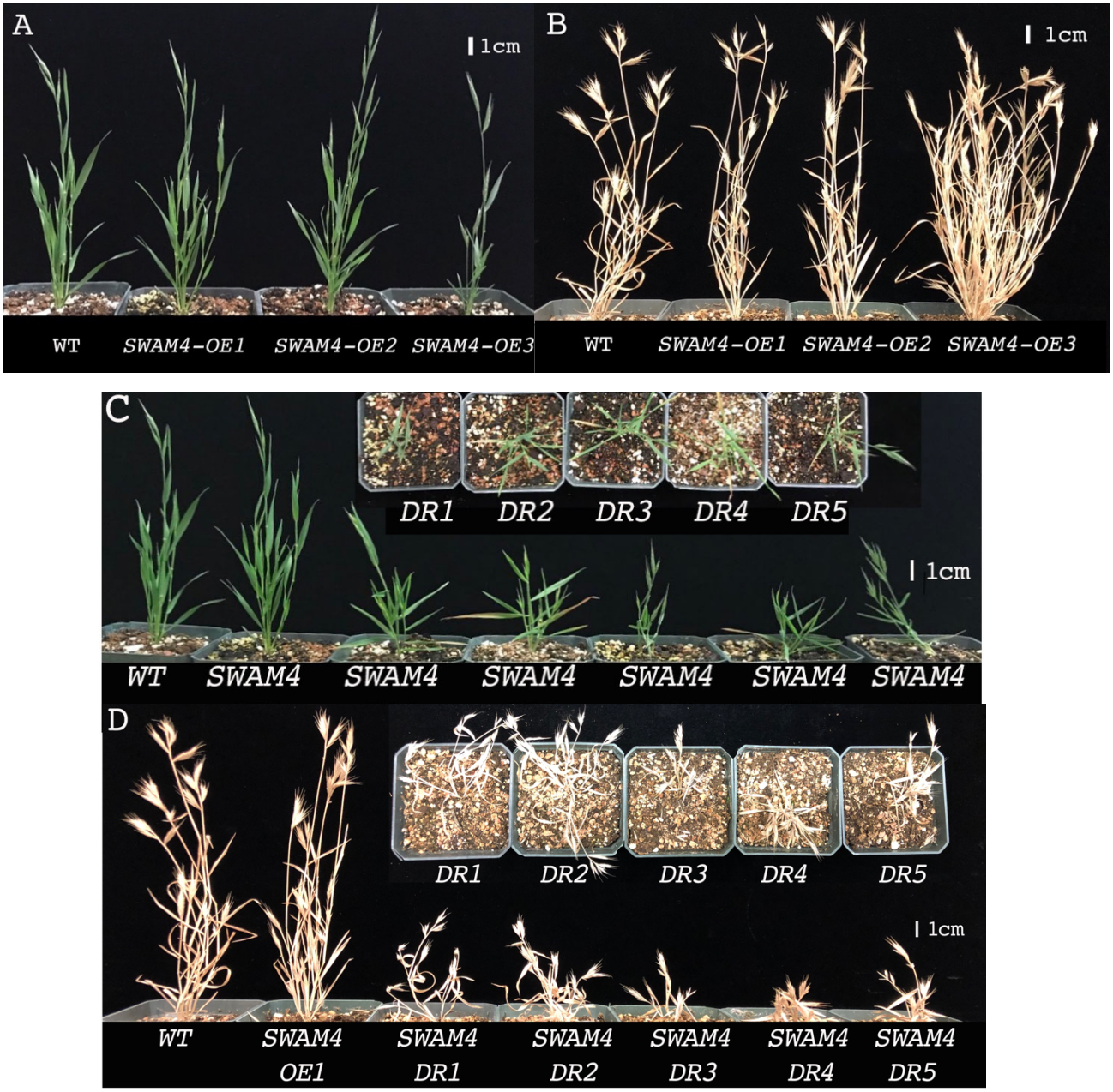


Figure 3. 3 **Plant phenotypes of WT and mutant plants *SWAM4-OE* and *SWAM4-DR***. Plants from three and five independent transformation events for *SWAM4-OE* and *SWAM4-DR* respectively. (A, C) Plants are from juvenile to early reproductive stage. The image was captured 29 DAG. (B, D) Plants are completely senesced. The image was captured 120 DAG.

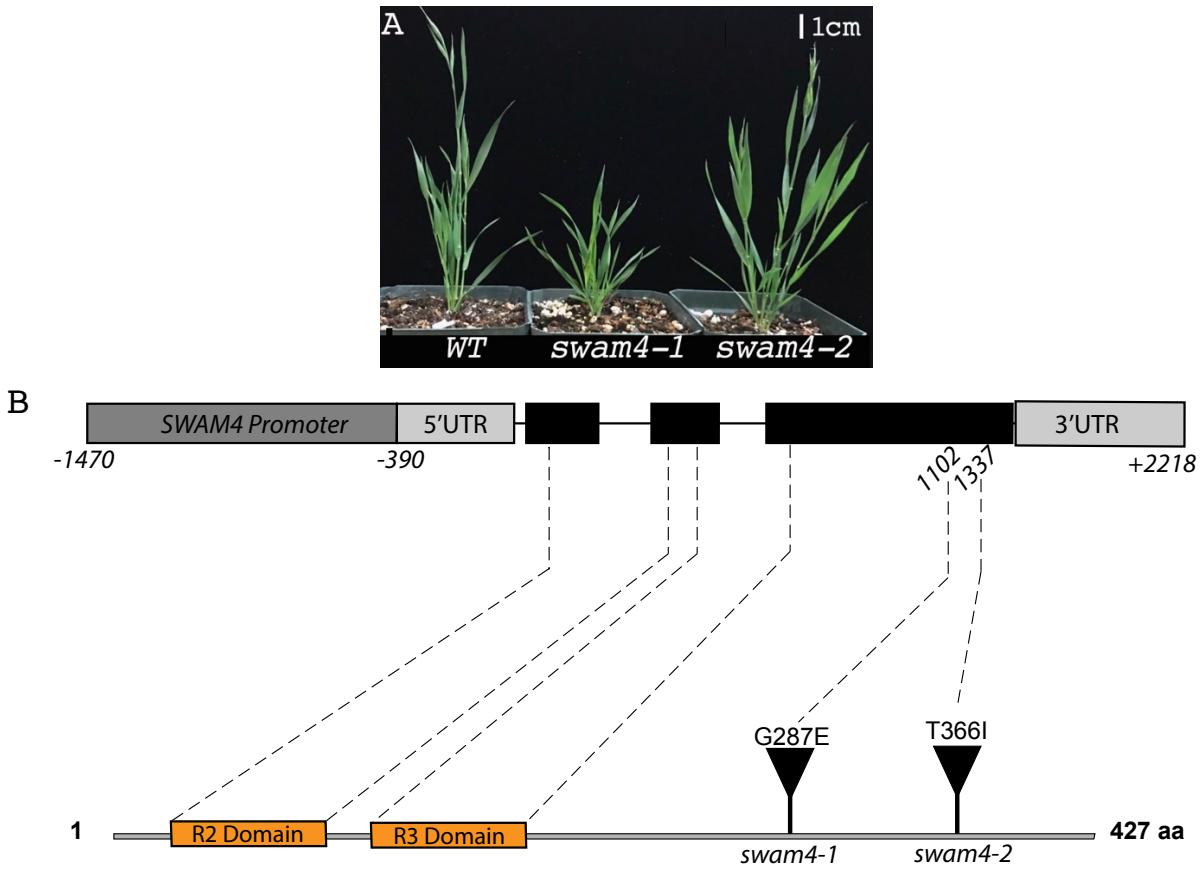


Figure 3. 4 ***swam4* mutant alleles**. (A) Plant phenotypes of WT and two mutant plants swam4-1 and swam4-2. Plants are from juvenile to early reproductive stage. The image was captured 29 DAG. (B) Top. Diagram of SWAM4 genomic structure indicating the location of the non-synonymous mutations of the two mutant lines in the exon 3 (black box) of *SWAM4*. Bottom. Diagram of SWAM4 protein with the location of the swam4-1 and swam4-2 mutant alleles. Dash lines show the position of the R2R3 protein domains and mutant alleles within the *SWAM4* coding region. Black triangles represent the position of each non-synonymous mutation. G, glycine; E, glutamic acid; T, threonine; I, isoleucine; aa, amino acids.

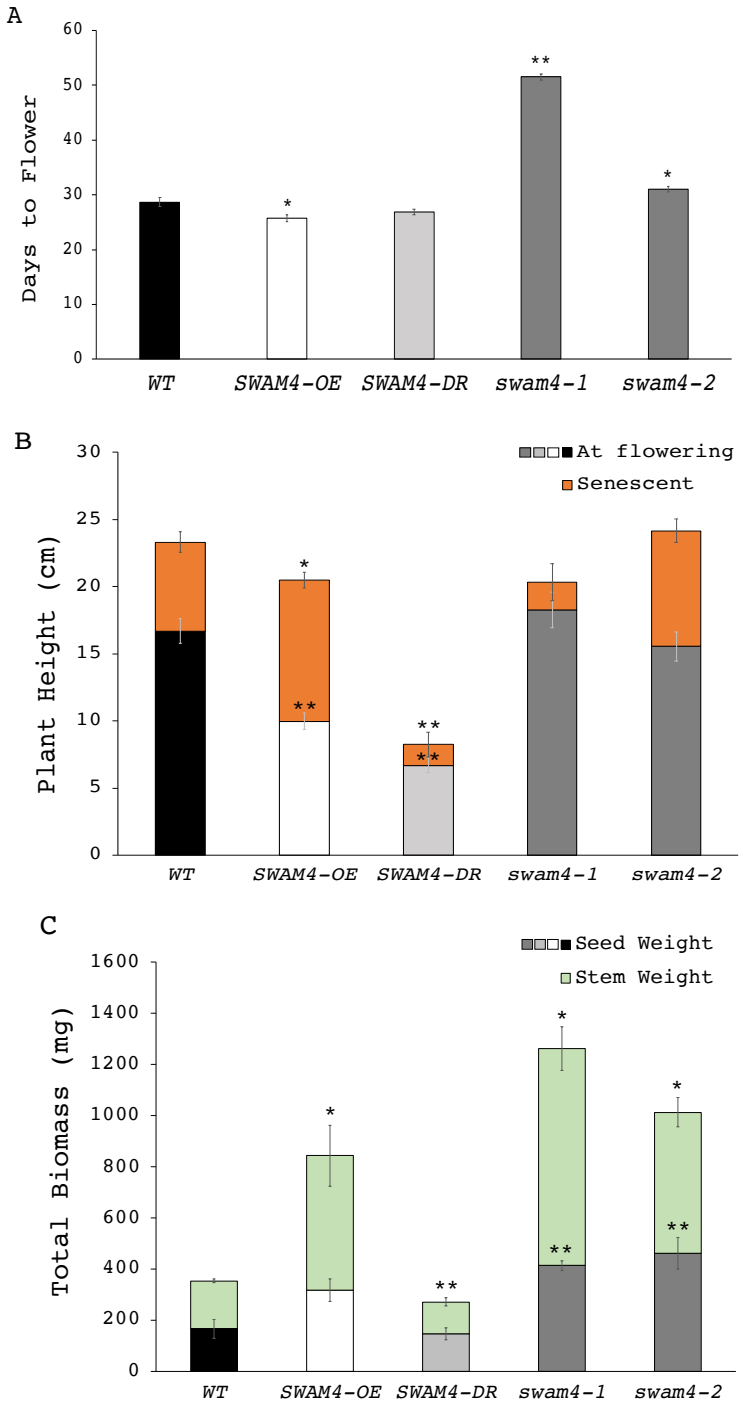


Figure 3. 5 **Plant phenotypes.** (A) Days until the start of heading when the inflorescence emerges from the flag leaf, are shown. n=6-40 per genotype. (B) Plant heights at flowering (white, black, and grey) and at senescent stage (orange) are shown, n=3-23. (C) Total above-ground biomass was measured at complete senescence, n=3-15 plants per genotype. Means ± SEM are shown. Two-tailed Student's t-tests were performed *p<0.05, **p<0.01.

3.3.5 *SWAM4* is a regulator of lignin biosynthesis

To determine if *SWAM4* function is related to cell wall formation, fully senesced stems of three independent events of *SWAM4-OE* and *SWAM4-DR* and the two *swam4* mutants were sectioned and stained with phloroglucinol-HCl (Figure 3.6A). Lighter staining of interfascicular fibers and epidermal cells suggests a reduction in lignin deposition in the *SWAM4-DR* and *swam4* mutants (Figure 3.6M-Za). Lighter staining was also observed in *SWAM4-DR5* vascular bundles (Figure 3.6S-U). *SWAM4-OE* cross sections were not different from WT (Figure 3.6D-L). Interestingly, *SWAM4-DR2* and to a lesser extent *swam4-1* stem cross sections exhibited an atypical stem shape (Figure 3.6P-R, V-X). Stem tissue at complete senescence of three and five independent events of *SWAM4-OE* and *SWAM4-DR*, respectively, and both *swam4* mutants was pulverized and processed for lignin quantification as acetyl bromide soluble lignin (ABSL). Consistent with the qualitative observations, *SWAM4-DR* plants had significantly reduced lignin content compared to WT control. No significant changes were observed in *SWAM4-OE* and *swam4* mutants (Figure 3.6B). To further investigate the function of *SWAM4* in stem shape and thickness, stem area and perimeter were measured in three independent sections of *SWAM4-OE* and *SWAM4-DR* (Figure 3.7A, B), but the reduced values in the transgenics was not significant. Interfascicular fiber wall thickness and xylem wall thickness were measured in *SWAM4-OE* and *SWAM4-DR*. Interestingly, interfascicular fiber wall thickness was dramatically reduced in *SWAM4-DR* stems, and no significative changes were observed in *SWAM4-OE* (Figure 3.7C). There were no differences in xylem wall thickness (Figure 3.7D)

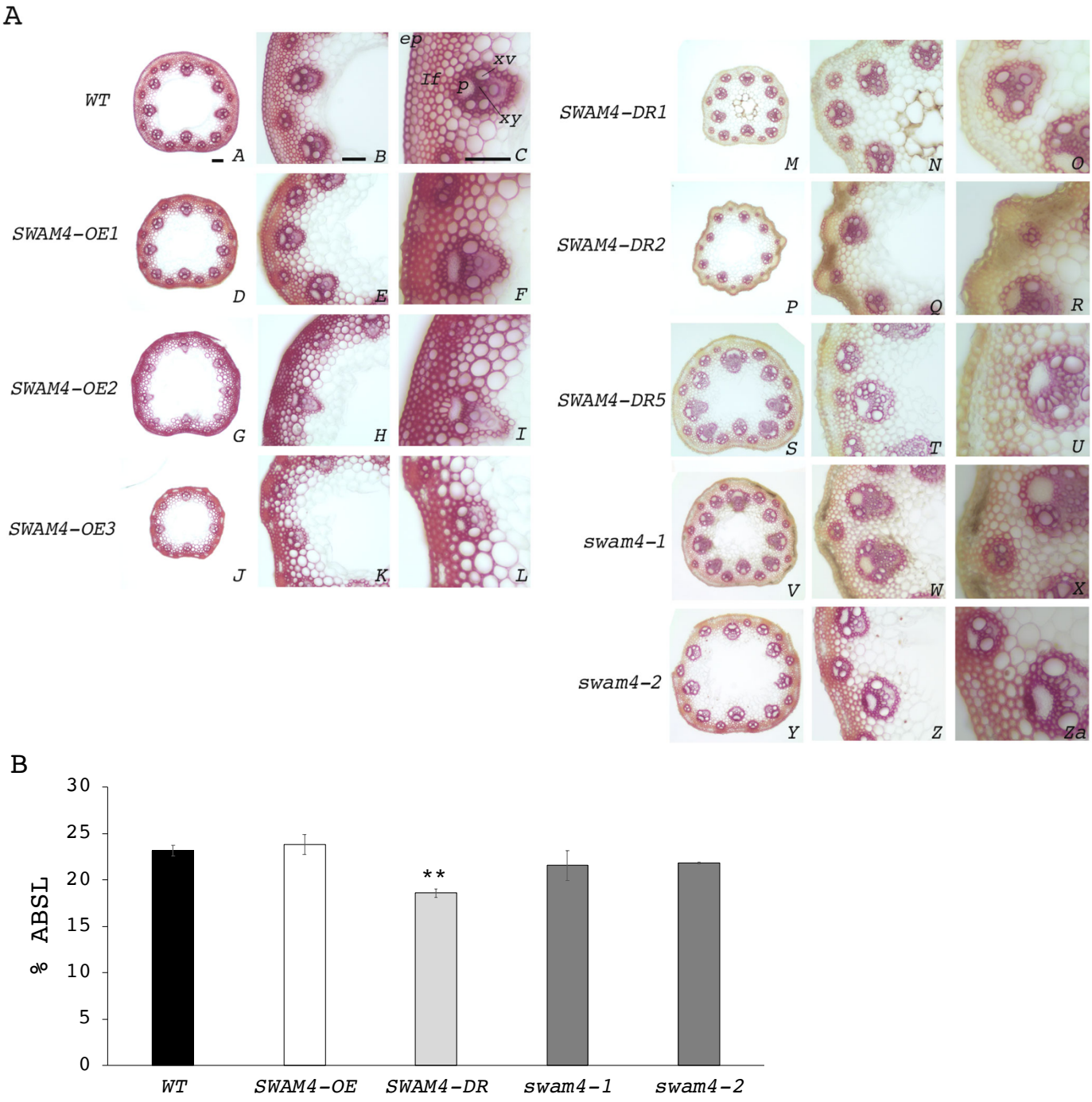


Figure 3. 6 **Histo-chemical and quantitative analysis for lignin composition in stems.** (A) Cross sections of the first internode of fully mature plants of WT(*A-C*), *SWAM4-OE* (*D-L*), *SWAM4-DR* (*M-U*), *swam4-1* (*V-X*) and *swam4-2* (*Y-Za*) stained with phloroglucinol-HCl. The images were captured using 4x, 10X and 20x objectives (left to right). ep, epidermis; if, interfascicular fibers; p, phloem; xv, xylem vessels; xy, xylem. Scale bar = 0.1 mm. (B) Acetyl bromide soluble lignin content of completely senesced stem tissue. Three and five independent events of *SWAM4-OE* and *SWAM4-DR* respectively, and *swam4* mutant plants were tested. n=3-15 plants per genotype were analyzed and samples were tested in triplicate. Means ± SEM are shown. Two- tailed Student's *t*-tests were performed *p<0.05

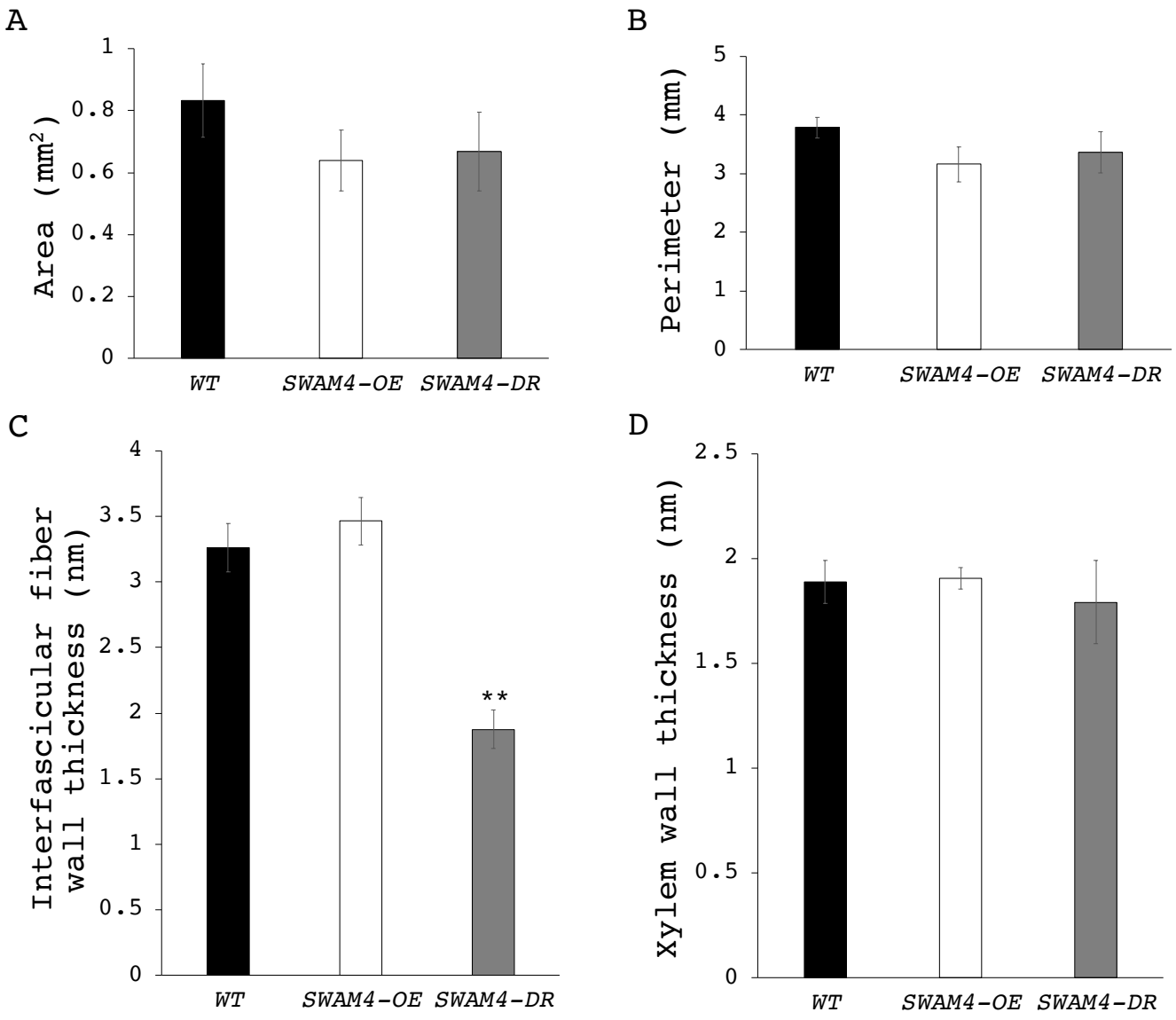


Figure 3. 7 **Stem and cell wall measurements of cross sections from the first internode of fully senesced plants**. Area (A) and perimeter (B) of WT, *SWAM4-OE* and *SWAM4DR*. n=3. Wall thickness of interfascicular fibers (C) and xylem cells (D) of three independent events and five sections event. n=15 per section. Means ± SEM. Two-tailed Student's t-tests were performed **p<0.01

Quantitative real-time PCR was used to measure transcript abundance of *SWAM4* in *SWAM4-OE*, *SWAM4-DR*, and WT from stems of plants grown under diurnal temperature and light cycle. Transcript abundance of *SWAM* in *SWAM4-OE* was not different from WT. On the other hand, the transgene was thirteen-fold greater in *SWAM4-DR* than control plants (Figure

3.8A). To explore the function of *SWAM4* in secondary cell wall biosynthesis, Q-RT-PCR was performed on stem samples to obtain gene expression data of *CESA4*, *CESA7*, *CESA8*, *CAD1*, and *COMT6* in *SWAM4-OE*, *SWAM4-DR*, and WT. The lignin biosynthetic genes *CAD1* and *COMT6* were significative upregulated in *SWAM4-OE* plants and downregulated in *SWAM4-DR* plants. Similarly, the expression of cellulose genes seemed to be reduced in *SWAM4-OE* plants and increased in *SWAM4-DR*; however, these results by Q-RT-PCR were found non-significant. (Figure 38B).

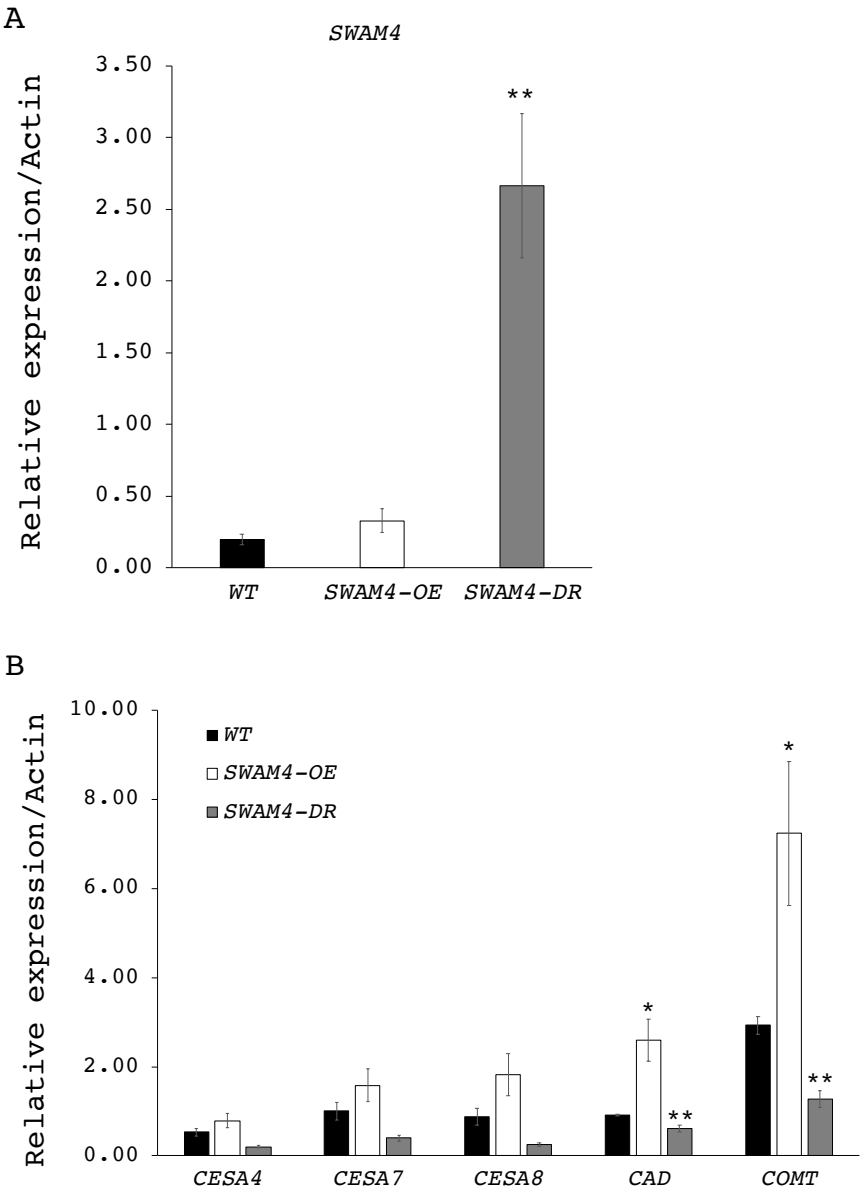


Figure 3. 8 **Transcript abundance of *SWAM4* and cell wall biosynthetic genes in WT, *SWAM4-OE* and *SWAM4-DR* mutant plants.** Q-RT-PCR was performed in WT, *SWAM4-OE*, *SWAM4-DR* plants grown under diurnal temperature and light cycle (LDHC; 20-hour light/4-hour dark at 26°C and 18°C respectively). Stem tissue was collected at ZT22 during the 4-hour dark cycle. Gene expression of *SWAM4* (A) and cell wall biosynthetic genes, *CESA4*, *CESA7*, *CESA8*, *CAD* and *COMT* were measured (B). Data were normalized to Actin housekeeping gene and six to three individuals were analyzed in triplicate. Means ± SEM were shown. Two-tailed Student's t-tests were performed *$p<0.05$, **$p<0.01$.

3.3.6 *SWAM4* is a regulator of cell wall biosynthesis

Transcriptome analysis using RNA-seq was performed using RNA from stems of WT, *SWAM4-OE*, *SWAM4-DR*, *swam4-1* and *swam4-2* plants. Samples were collected in the middle of the dark cycle (Table 2.4). Then, cDNA libraries were sequenced with the Illumina NovaSeq 6000 paired-end sequencing x100bp. The assessment of quality was described in Chapter 2, Section 2.311 (Figure 2.18). A stringent p-value<0.005 was used to generate heatmaps and hierarchical analysis, principal component analysis (PCA) and scatter plots; less stringent p-values p<0.05 and p<0.01 were used to determine significantly differential expression in further analysis. PCA of the transcriptomic dynamics with all genotypes and individual comparison with WT samples revealed that *SWAM4-DR* and *swam4-1* clusters were well-correlated and differentiated from WT. *SWAM4-OE* and *swam4-2* were less similar compared to WT (Figures 3.9 - 3.10). Similar to PCA results, upregulated (yellow) and downregulated (red) deferentially expressed genes were better clustered and distinguished in *SWAM4-OE* and *swam4-2* within replicates (Figure 3.11). All genotypes presented differentially expressed genes with a stringent significant threshold p<0.005, plotted in red. *swam4-1* and *SWAM4-DR* plants showed a higher number of genes differentially expressed (Figure 3.12). A subset list of genes of interest that were differentially expressed within all genotypes with a p-value <0.01 is presented in Table 3.4.

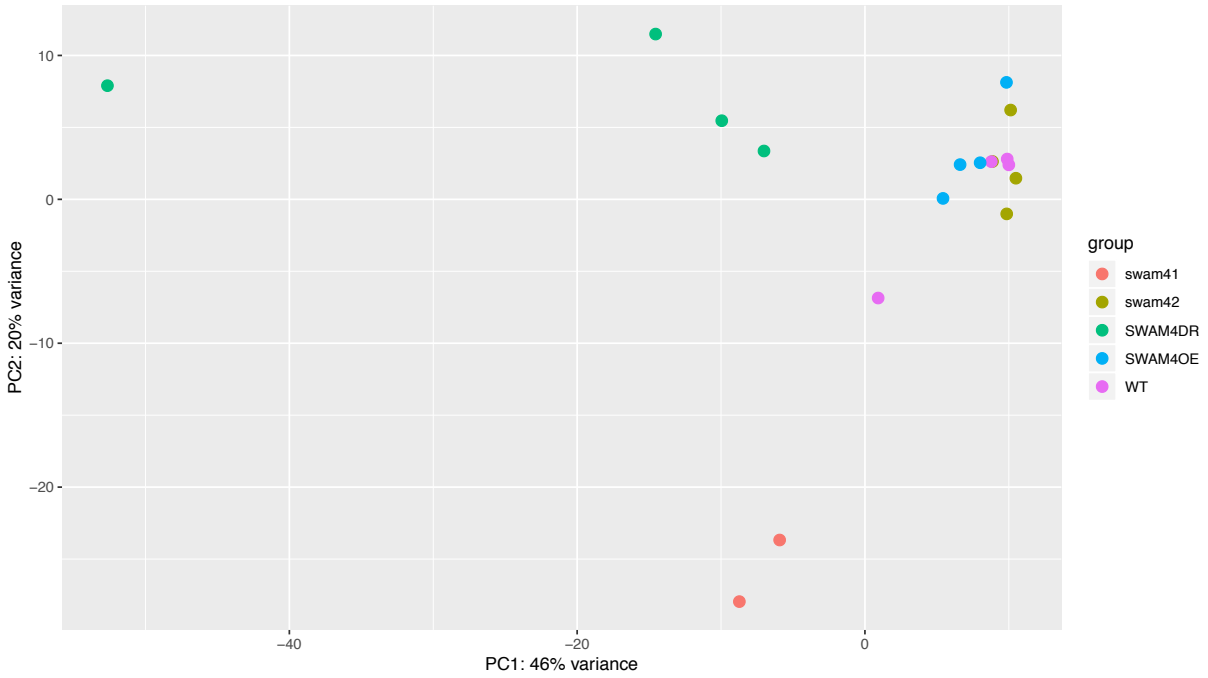


Figure 3. 9 **Genetic distribution of WT, *SWAM4-OE*, *SWAM4-DR*, *swam4-1*, and *swam4-2* individual plants.** Principal Component Analysis (PCA) illustrating the distribution of individual RNA-seq samples along the two PCs.

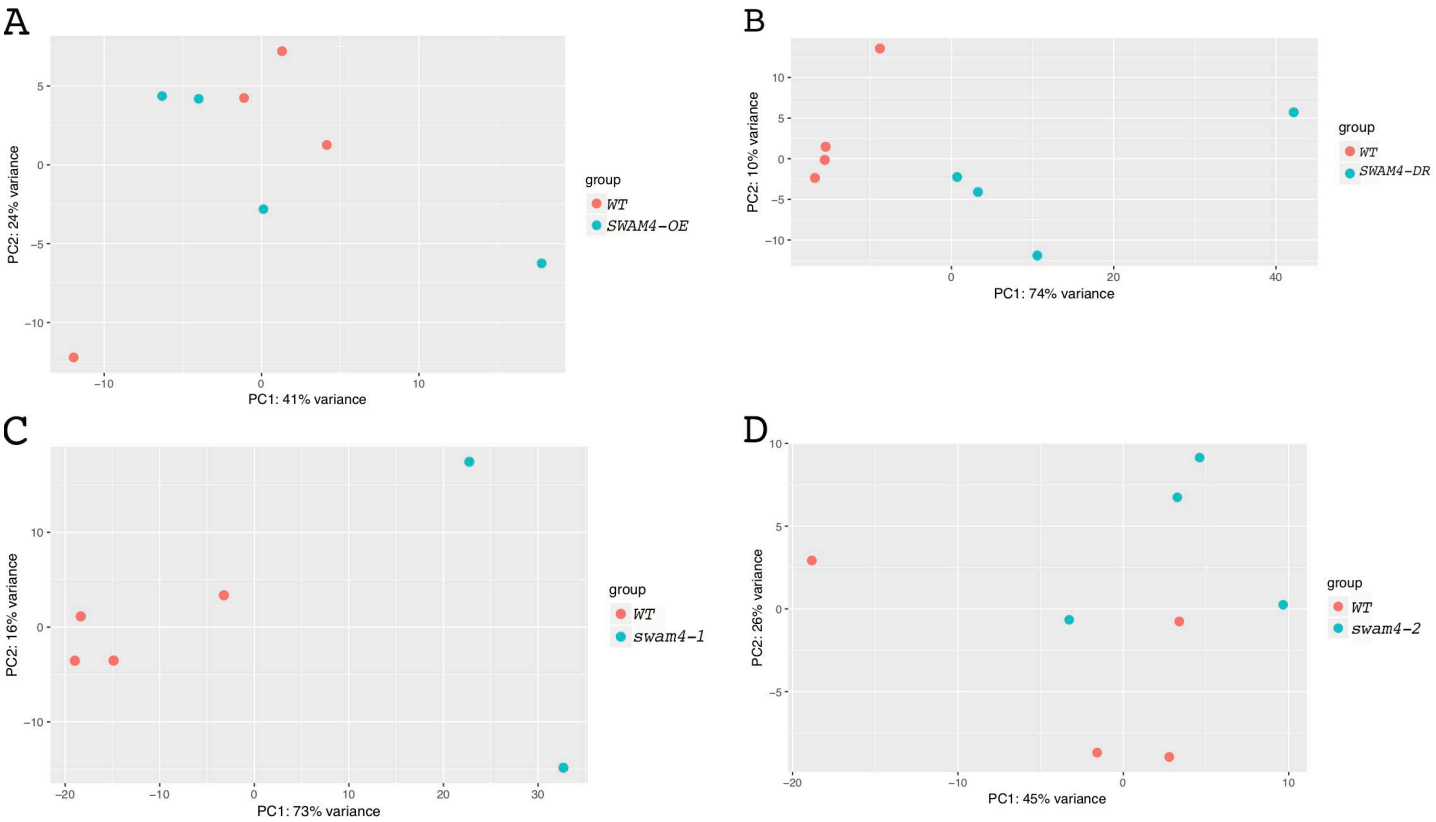


Figure 3. 10 **Total variation and distribution of WT and *SWAM4* mutants. PCA.** All genotypes were compared to WT samples individually. Percentages of variance are indicated.

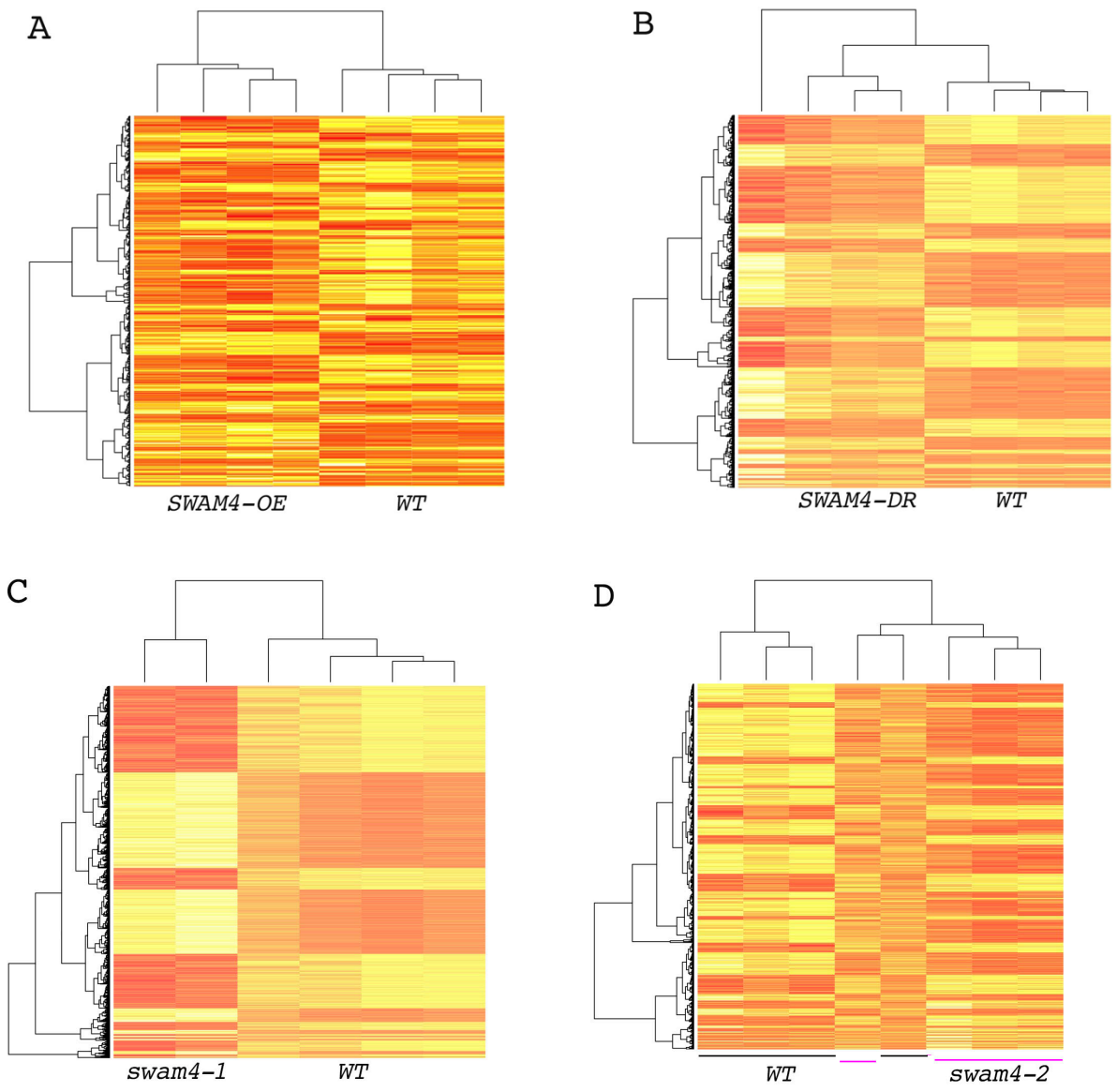


Figure 3. 11 **Differential gene expression of WT and S*WAM4* mutants. Heatmap and hierarchical analyses.** Upregulation is indicated in yellow and downregulation in red.

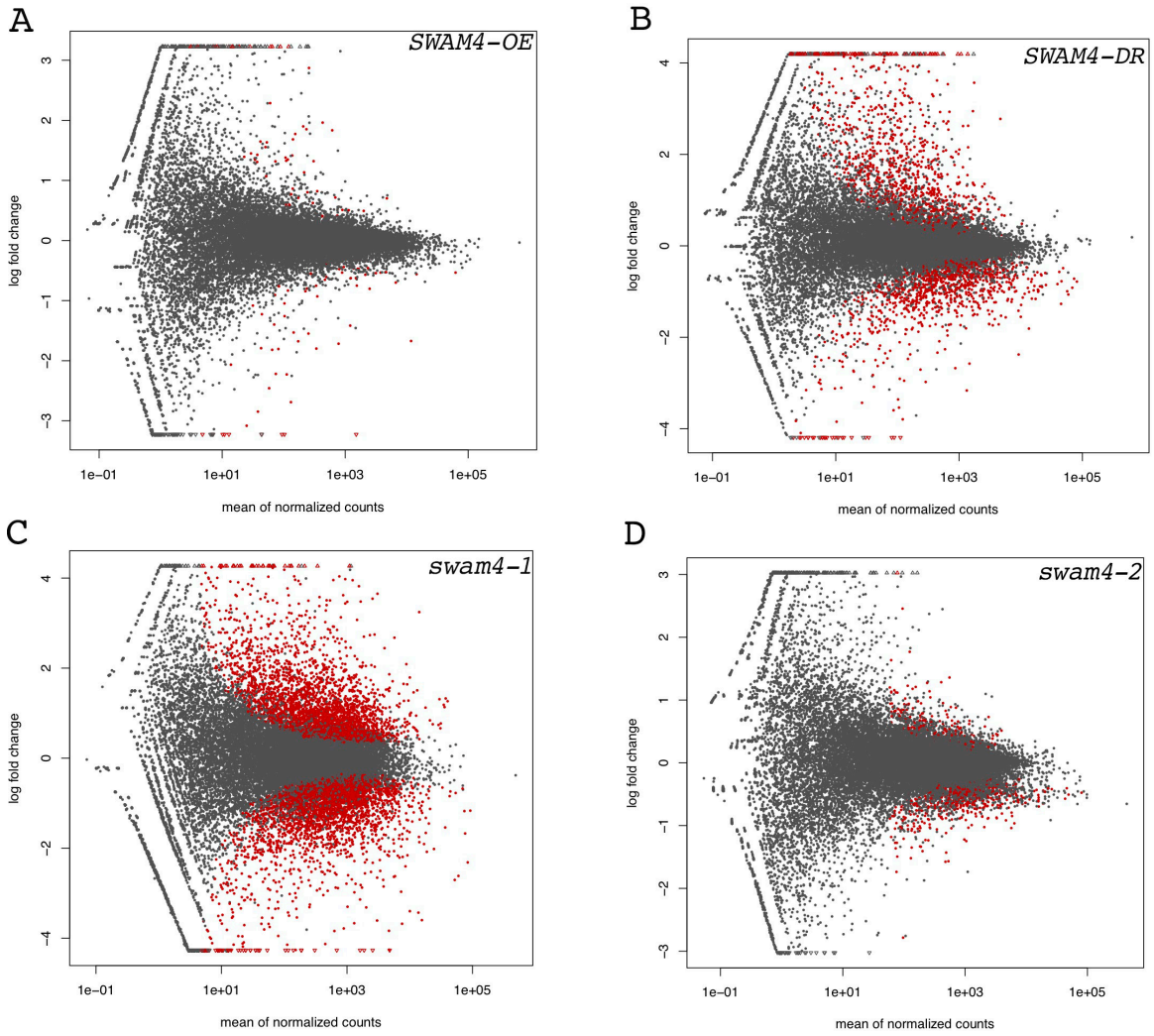


Figure 3. 12 **Genes differentially expressed (Log2 FC) of SWAM4 mutants relative to WT. Scatter plots.** Upregulation (above 0) and downregulation (below 0) are illustrated. Significance threshold p<0.005 (red dots).

Transcription factors, cell-wall-associated genes, glycosyltransferases, glycosyl hydrolases, and laccases were identified as genes putatively regulated by *SWAM4*. Similar to gene expression results obtained by Q-RT-PCR (Figure 3.8A), *SWAM4* expression was found upregulated in *SWAM4-DR* plants, and no significant changes in SWAM4 expression were observed in *SWAM4-OE* plants (Figure 3.13). Since dominant suppression of *SWAM4* targets should occur in *SWAM4-DR* plants, all genes of interest that showed significant downregulation

in *SWAM4-DR* plants were subject to investigation in other experiments. *GNRF*, *SWAM7*, *SWAM8*, *KNOB7*, and *NAC97*, all genes associated with cell wall formation in *B*. *distachyon*, were found to be downregulated in *SWAM4-DR* plants. Differential expression of some of these genes was also observed in *swam4* mutants; however, due to the fact that *swam4* mutant alleles were not complemented, their results were only comparative (Figure 3.13). If the *swam4* mutants are crossed with WT, their genotypes would be cleaner from other mutations and they would be more informative to compare with WT. *SWAM4-OE* plants showed a significant downregulation of cellulose and lignin genes; nevertheless, since *SWAM4-OE* plants are not overexpressing the *SWAM4* gene expression as expected, the results obtained were not used to deduce *SWAM4* function. Interestingly and consistent with quantitative and qualitative results of lignin composition and gene expression results obtained by Q-RT-PCR (Figures 3.6, 3.8B), cellulose and lignin biosynthetic genes were significantly downregulated in *SWAM4-DR* plants. Together, these findings suggest that *SWAM4* is a regulator of cell wall biosynthetic genes including *CESA4*, *CESA7*, *CESA8*, *CAD1*, and *COMT6* (Figure 3.14). Glycosyltransferases, Glycosyl-hydrolases, and laccases were particular genes that were found downregulated in both *SWAM4-DR* plants and *SWAM4-OE* plants (Figure 3.15). Additional genes with significant regulation among all genotypes are listed in Figure 3.16. In general, most of the genes identified as potentially regulated by *SWAM4* were repressed in both, *SWAM4-DR* and *SWAM4-OE* plants; a lower repression, neighboring the threshold (0) but still significant was observed in *SWAM-OE* plants.

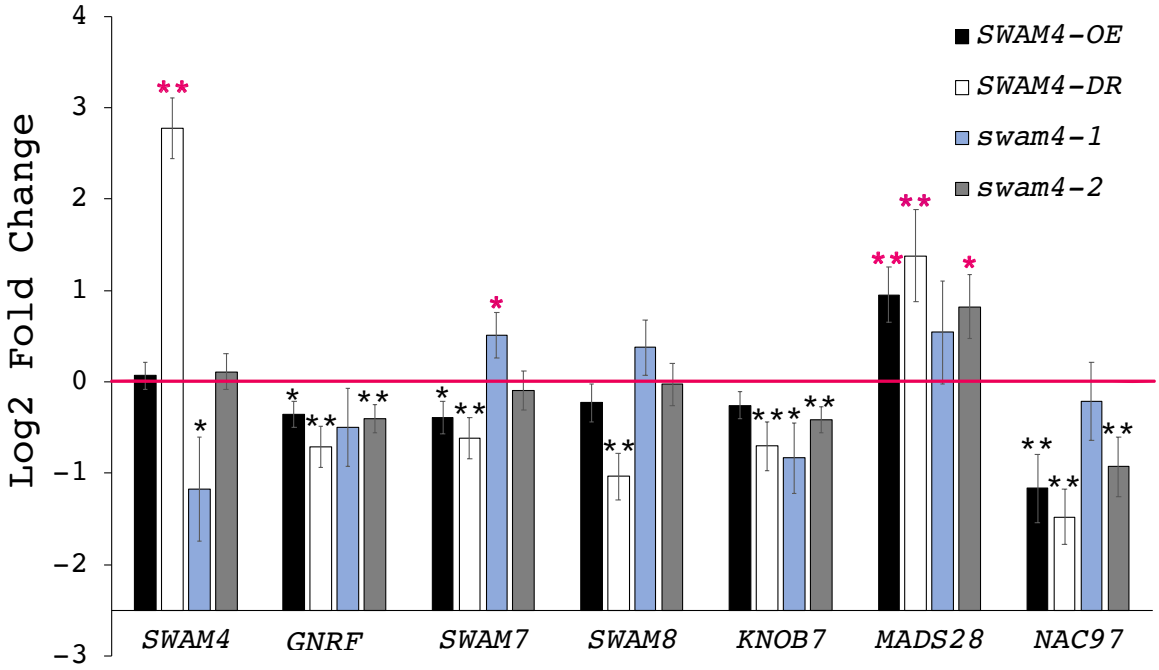


Figure 3. 13 **Transcription factors that are differentially expressed in *SWAM4* mutants relative to WT.** Log2 Fold change (LFC) of SWAM4, GNRF, SWAM7, SWAM8 and KNOB7. Log2 Fold change ± log fold change standard error (lfcSE). Positive and negative LFC values are above and under the threshold (0) respectively. *p<0.05, **p<0.01.

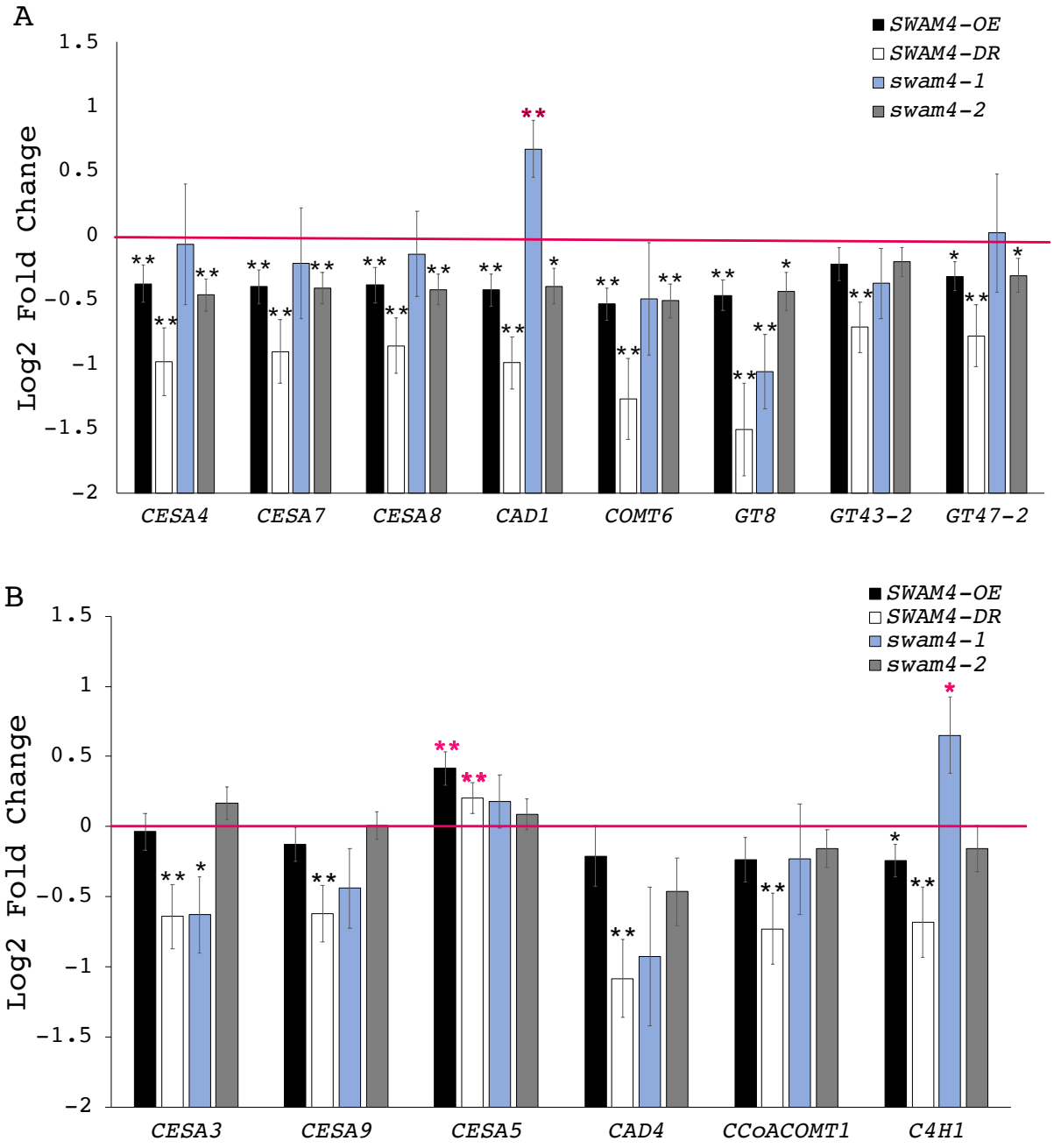


Figure 3. 14 **Cell wall biosynthetic genes that are differentially expressed in *SWAM4* mutants relative to WT.** (A) Log2 Fold change (FLC) of cellulose (*CESA4*, *CESA7*, *CESA8*), lignin (*CAD1*, *COMT6*), and hemicellulose biosynthetic genes (GT8, GT43-2, GT47-2). (B) Log2 Fold change of additional genes associated with cell wall formation. Log2 Fold change ± log fold change standard error (lfcSE). Positive and negative LFC values are above and under the threshold (0) respectively. *p<0.05, **p<0.01.

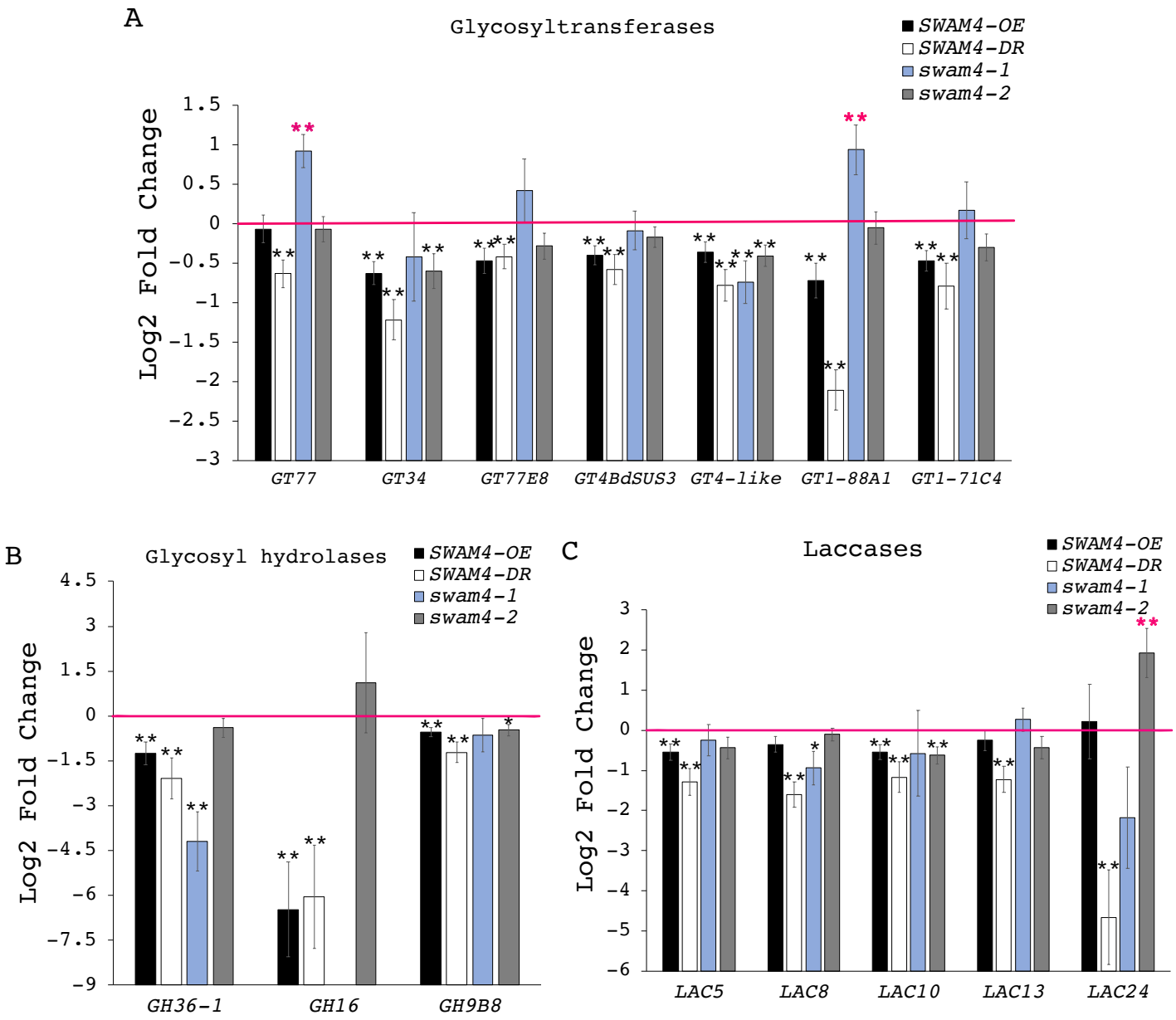


Figure 3. 15 **Genes of interest differentially expressed in *SWAM4* mutants relative to WT.** (A) Glycosyltransferases. (B) Glycosyl hydrolases. (C) Laccases. Log2 Fold change ± log fold change standard error (lfcSE). Positive and negative LFC values are above and under the threshold (0) respectively. *p<0.05, **p<0.01.

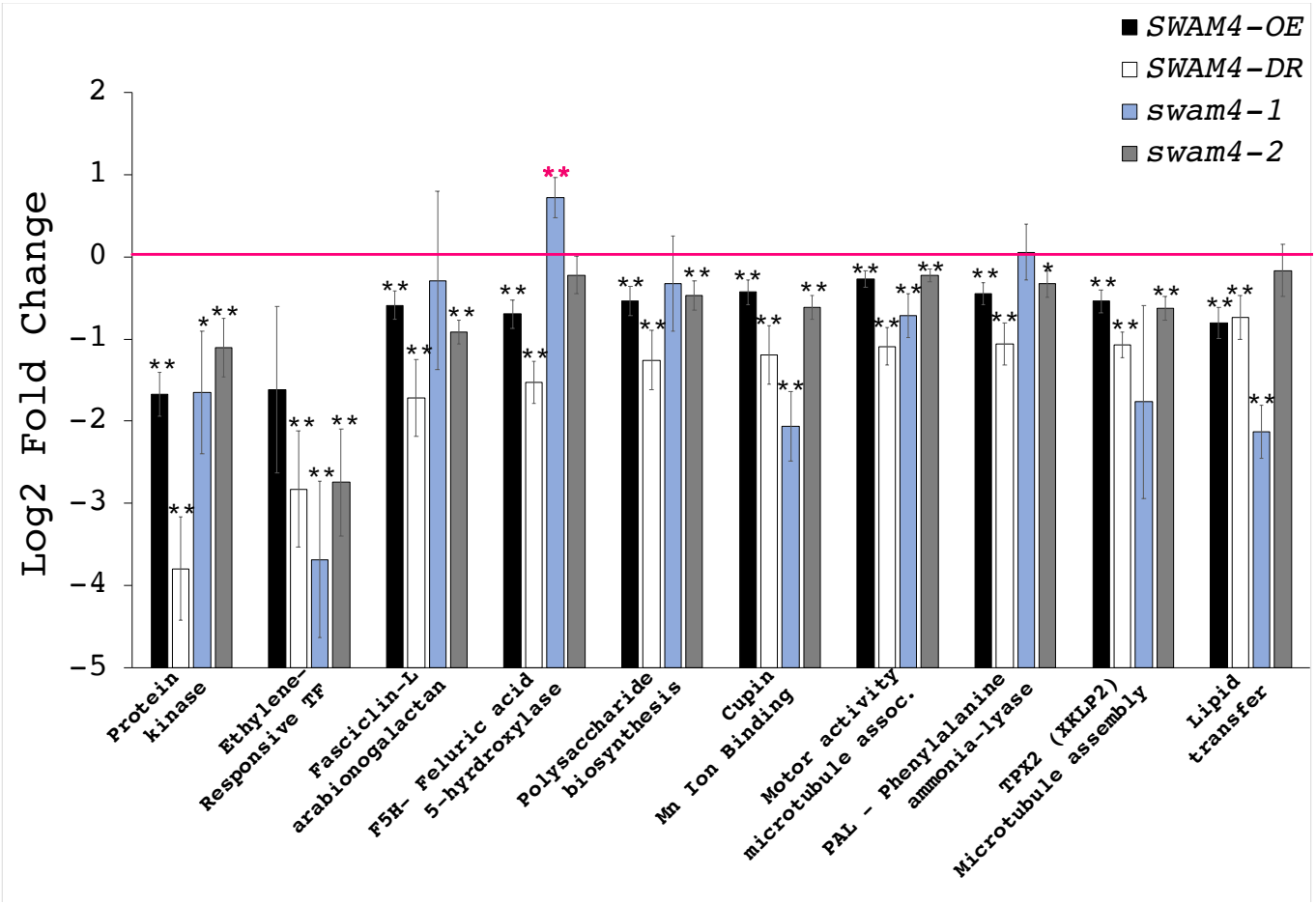


Figure 3. 16 **Additional genes of interest differentially expressed in *SWAM4* mutants relative to WT.** The log2 Fold change ± log fold change standard error (lfcSE). Positive and negative LFC values are above and under the threshold (0) respectively. *p<0.05, **p<0.01.

3.4 Discussion

The R2-R3 MYB protein family of transcription factors has been well-studied in the eudicot *A. thaliana*, and their function has been associated with several plant growth and developmental mechanisms (Dubos et al., 2010; Stracke et al., 2001). Some of the MYB transcription factors have been identified as regulators of secondary cell wall biosynthesis in eudicots as well as monocots (McCarthy et al., 2009; Nakano et al., 2015; Zhong et al., 2008; Zhong and Ye, 2014; Zhong et al., 2015). *SWAM4* was identified among 26 MYB transcription factors in *B. distachyon*, as part of a collection of genes co-regulated with cellulose and lignin

biosynthesis genes. Moreover, *SWAM4* was also included in the phylogenetic analysis of MYB proteins related to *SWAM1*, a positive regulator of secondary cell wall formation (Handakumbura et al., 2018). While *SWAM1* did not have an orthologous gene in *A. thaliana*, *SWAM4* was shown to be closely related to *AtMYB50* and *AtMYB61* genes. These genes encode two proteins that share 66% identity and a conserved N-terminal amino acid sequences but a distinctive C-termini (Prouse and Campbell, 2012; Zhao and Bartley, 2014). *AtMYB50* has not been studied in detail compared to *AtMYB61*; however, functional modules and co-expression network analysis have grouped *AtMYB50* with genes expressed mainly in the xylem (Cai et al., 2014). Functional genomics and transcriptomic analysis revealed that *AtMYB61* has pleiotropic effects, including controlling stomatal aperture, terpene metabolism, and ectopic lignification (Hirano et al., 2013b; Liang et al., 2005; Matias-Hernandez et al., 2017; Newman et al., 2004; Romano et al., 2012). Mutations in *AtMYB61* promoted changes in xylem cell structure and decreased xylem formation (Romano et al., 2012). Functional genomics and transcriptional activation assays have shown that *PtoMYB74*, an *AtMYB61* ortholog in poplar, positively regulates cell wall formation (Li et al., 2018). Interestingly, two NAC proteins, NAC29 (SND1-like) and NAC31 (VND-like) were found regulating *OsMYB61a* directly, via gibberellin signaling, and subsequently, OsMYB61a protein was found to function as an activator of cellulose biosynthetic genes in rice (Huang et al., 2015). Furthermore, transcriptional activation assays in yeast and rice protoplasts showed that *OsMYB61*a and *OsMYB61b, SWAM4* closest ortholog in rice*,* were regulated by *OsSND2A* (*Os05g48850*) (Chapter 2, Figure 2.1), a positive regulator of cell wall biosynthesis in rice (Ye et al., 2018). Overall, *SWAM4* function might be related to secondary cell wall formation due to its amino acid similarity with MYB61 proteins that are implicated in cell wall regulation.

In grasses, stems are central organs that function as specialized compartments where biomass is accumulated. Additionally, stems contain specific genes that are co-expressed to modulate internode elongation and cell wall formation (Jensen and Wilkerson, 2017; Kebrom et al., 2017). High transcript abundance of genes related to secondary cell wall biosynthesis has been located in stems, and changes in the expression of these genes have caused either reduction or ectopic deposition of cell wall components (Huang et al., 2018; Kebrom et al., 2017; Zhong et al., 2008). *SWAM4* transcript abundance was higher in stems compared to roots and leaves (Figure 3.2A). Comparative co-expression network analyses showed that genes expressed in mature stems were also involved in cell wall formation (Sibout et al., 2017). Interestingly, the co-expression analysis shown that *SWAM4* clusters with *BdCESA9*, a primary-cell-wall-related cellulose synthase gene that is highly expressed in stems and roots compared to leaves (Handakumbura et al., 2013); a gene encoding a fasciclin-like arabinogalactan protein (*Bradi3g39740*), a cell wall component that was found to be associated with somatic embryogenesis in *B. distachyon* (Betekhtin et al., 2018); the gene *KNOB7*, an ortholog of the *A. thaliana* homeobox transcription factor gene *KNAT7* that is upregulated during cell wall formation, and has been found to regulate lignin and xylan gene expression (He et al., 2018; Li et al., 2012; Zhong et al., 2008; Zhong et al., 2007a; Zhong and Ye, 2012); and the gene *BdMYB1* transcription factor related in the upregulation of cell wall formation, was clustered also in the SWAM protein phylogenetic tree (Figure 3.1A) as an ortholog of *AtMYB46*, a target of *AtSWN*, a secondary wall NAC related to the cell wall regulation of *A. thaliana* (Valdivia et al., 2013). Collectively, the fact that *SWAM4* is highly expressed in stems and is clustered in the co-expression network with many genes associated with cell wall regulation in *B. distachyon* suggests that *SWAM4* might function in cell wall formation as well.

To a certain extent, *SWAM4-OE* and *SWAM4-DR* mutants exhibit the whole plant phenotype of *SWAM1-OE* and *SWAM1-DR* mutant plants (Handakumbura et al., 2018) with some subtle differences. Since the two SWAM proteins are close homologs in *B. distachyon,* this observation is expected and presumes that they possibly perform similar functions and share genetic redundancy. The reciprocal flowering phenotype of *SWAM4-OE* and both *swam4* mutants showed similarities between *SWAM4* and *AtMYB61* mutant phenotypes; *35S::AtMYB61* plants flowered early whereas *atmyb61* T-DNA mutant plants were late-flowering (Romano et al., 2012). Thus, perturbing *SWAM4* and *AtMYB61* in *B. distachyon* and *A. thaliana*, respectively, resulted in similar phenotypes. While *SWAM4-OE* and *AtMYB61* overexpression plants developed rapidly, compared to WT controls, *swam4* and *atmyb61* plants developed slowly. Although no flowering phenotype was observed in *SWAM4-DR* plants, their severe dwarf phenotype and the fact that the plant height of *SWAM4-OE* plants was also reduced suggest that *SWAM4* function is associated with plant growth and development.

B. distachyon is a suitable model to study unique biological traits in temperate grasses such as maize, rice, wheat, barley and switchgrass, to obtain a better understanding of plant biomass accumulation processes and cell wall development (Girin et al., 2014; Jensen and Wilkerson, 2017; Vogel et al., 2006). Seed biomass was significantly increased only in *swam4* mutants, but interestingly stem biomass and total biomass were significantly different in all mutants compared to WT control plants; while stem biomass was increased in *SWAM4-OE* and *swam4* mutant plants, it was reduced in *SWAM4-DR* plants. These results were similar to

SWAM1 mutant phenotypes (Handakumbura et al., 2018) and suggest that the function of *SWAM4* is related to stem development and probably associated with seed formation as well.

Dominant repression of the expression of *SWAM4* targets in *SWAM4-DR* plants caused a significant qualitative and quantitative reduction of total lignin content and a reduction of interfascicular fiber wall thickness. Even though *swam4-1* and *swam4-2* plants did not have a significant reduction in percent lignin content as measured by the ABSL assay, the qualitative observation in the cross-section showed lighter staining and presumably a slight decrease in lignin content. Although *SWAM4-OE* plants did not show an increase in lignin content, the phenotypes observed in *SWAM4-DR*, and *swam4* mutants suggest that *SWAM4* has a role in the regulation of lignin in fiber cells. No changes were observed in xylem cell wall thickness. Comparatively, functional characterization of *AtMYB61* in *A. thaliana* indicated that *AtMYB61* is a positive regulator of xylem vessel formation and seemes to promote differentiation of cambial cells into vessels or fiber cells. While the *atmyb61* mutant had fewer xylem cells with irregular shape and thinner walls compared to the control, the overexpression mutant *35S::MYB61* had a higher number of xylem cells and ectopic lignification (Romano et al., 2012). Similarly, the poplar homolog of *SWAM4* in *Populus tomentosa PtoMYB74* was found to be highly expressed in stems and xylem tissues. Overexpression of *PtoMYB74* in *P. tomentosa* and *A. thaliana* resulted in an increased cell wall thickness in the vessels and the accumulation of lignin and cellulose (Li et al., 2018). Eudicot and monocot plants exhibit a distinct cell composition and organization in the stem and vascular system. The eudicots have vascular bundles arranged in a ring and a cambium layer that differentiates into xylem and phloem. In contrast, monocots lack a cambium layer and have vascular bundles well-defined in a bundle sheath that includes xylem

and phloem cells (Handakumbura and Hazen, 2012). The *Pinus taeda* MYB protein PtMYB8, homologous to SWAM4 (Figure 3.1B), similarly regulates lignin biosynthesis in secondary xylem (Bomal et al., 2008). Conifers also exhibit a different cell organization. Pine stems, for example, contain rays of xylem growth composed of cells with thin cell walls. Wood (secondary xylem) is made of tracheids, and wood cells derive from vascular cambium activity (Plomion et al., 2001). These data are consistent with the functional association of *SWAM4* and the regulation of lignin. Presumably, some of the differences in the vascular system of eudicots, monocots, and conifers are useful to determine the biological reasons that explain the preferential regulation of *SWAM4* in interfascicular fibers and the irregular shape observed in the stems of *SWAM4-DR* plants.

Collectively, gene expression analysis showed evidence that *SWAM4* functions in the regulation of lignin and cellulose biosynthetic genes (*CESA4, CESA7*, *CESA8*, *CAD1*, and *COMT6*). The observation that *SWAM4* was not overexpressed, as expected, in *SWAM4-OE* plants but there was a significant change in the expression of genes that potentially are regulated by *SWAM4* (based on *SWAM4-DR* gene expression patterns) was similar to the results described for the *SWAM1-OE* mutant plants (Handakumbura et al., 2018). The failure of the overexpression system is still unclear, and future experiments might clarify a possible reason, such as the potentially toxic effects of overexpressing these transcription factors. *GNRF* was significantly repressed in *SWAM4-DR* plants and *swam4-2* plants, and reciprocally, *SWAM4* was repressed when *GNRF* was overexpressed (Chapter 2, Figure 2.22A). Consistently, 70 out of the 164 genes listed in Chapter 2 (Table 2.5) as putatively regulated by *GNRF* were also found differentially expressed in *SWAM4* RNA-seq data (Table 3.4, blue). Among these genes,

BdCESA3 (Bradi1g54250), a glycosyltransferase, GT77 family member (Bradi4g28260), a harpin-induced protein (Bradi3g45160), a GDSL-like lipase/acylhydrolase (Bradi2g12370), and *BdCESA5* (Bradi1g29060) were upregulated in *GNRF-OE* and downregulated (except for *BdCESA5*) in *SWAM4-DR* plants. Conversely, *BdMADS28* (Bradi2g48690), an EamA-like transporter (Bradi2g25490), and an arsenite secondary active transmembrane transporter (Bradi1g78100) were upregulated in *SWAM4-DR* and downregulated in *GNRF-OE* plants. These findings might be used to identify a relationship between GNRF and SWAM4 and perhaps to elucidate a possible transcriptional directionality. Even though yeast-one-hybrid assays showed reciprocal protein-DNA binding by GNRF and SWAM4 (Chapter 2, Figures 2.14 and 2.15B), additional experiments *in planta* are required to identified direct and indirect binding among targets and to clarify their functional relationship in the regulation of cell wall formation.

Seventeen of the genes that belong to the SWAM4 cluster in the co-expression analysis were found differentially expressed in the *SWAM4* RNA-seq analysis. Among them, the homeobox gene *KNOB7* (*Bradi1g76970*), an ortholog of *KNAT7* in *A. thaliana*, was downregulated in *SWAM4-DR* and *swam4-2* plants. *KNAT7*, as described above, was identified as a downstream component of secondary cell wall regulation (Zhong et al., 2008; Zhong et al., 2007a; Zhong and Ye, 2012) and regulates cell wall formation in *A. thaliana* (He et al., 2018; Li et al., 2012; Liu et al., 2014). *KNAT7* was selected as a target of *AtMYB61,* a *SWAM4* ortholog in *A. thaliana,* by transcriptomic analysis of *atmyb61* and 35S::MYB61 plants (Romano et al., 2012). Additionally, a relationship between *KNAT7* and *AtMYB61* was supported by the similar xylem-reduction phenotypes in mutant plants of both genes and by binding of protein to DNA via an AC binding element, as detected by electrophoretic mobility shift assays (Romano et al.,

2012). To date, *KNOB7* has remained uncharacterized in *B. distachyon*; however, the fact that KNOB7 clusters with SWAM4 in the co-expression analysis and is downregulated in the *SWAM4-DR* and *swam4-2* plants suggest that it might be, like *KNAT7*, involved in cell wall formation in *B. distachyon*. Among the other transcription factors regulated by SWAM4, *BdNAC97* (Bradi5g15587) is uncharacterized; however, mutations of its orthologues gene *Dry* in sorghum plants resulted in collapsed cells and changes in the composition of the cell wall in stems (Zhang et al., 2018b). Overall, the functional association of *SWAM4* with other genes related to cell wall formation would provide information about a cascade of transcription factors and molecular interactions within a transcriptional network that regulates cell wall biosynthesis in grasses.

The function of *SWAM4* in the regulation of secondary cell wall biosynthesis was additionally supported by the identification of glycosyltransferases, glycosyl hydrolases, laccases and other proteins associated with cell wall formation in the *SWAM4* RNA-seq analysis presented here. Further analyses are needed to identify direct and indirect interactions, their directionality, and the protein binding sites implicated. Collectively, the findings of the putative targets of *SWAM4* in *B. distachyon* can be used to identify orthologous genes in eudicots, conifers, and grasses that contribute to the understanding of the transcriptional regulation and molecular processes underlying secondary cell wall formation.

Table 3. 1 **List of primers for Chapter 3**

Primer Name	Sequence
Gal4BDTOPOFWD	CACCATGAAGCTACTGTCTTCTATC
P17_pDEST22_F	TATAACGCGTTTGGAATCACT
P18_pDEST22_R	AGCCGACAACCTTGATTGGAGAC
P19_pDEST15_F	GCGACCATCCTCCAAAATC
P20_pDEST15_R	TAAAGCCTGGGGTGCCTAAT
P76_Hyg_F	AGAATCTCGTGCTTTCAGCTTCGA
P77_Hyg_R	TCAAGACCAATGCGGAGCATATAC
P78_pMDC32_F	CGCACAATCCCACTATCCTT
P79_pMDC32_R	AAGACCGGCAACAGGATTC
P80_pLUC_F	TGTGCTCCTTCCTTCGTTCT
P81_pLUC_R	CTTCGTTGTTTTCGGTTGGT
P84_pLUC2_F	AAGCTTGAATTCGAGCTCGGTACCC
P85_pLUC2_R	CAGAGCACATGCCTCGAGGTCGA
P98_CaMV_35SPr_F	CTATCCTTCGCAAGACCCTTC
P99_att1_seq_F	AACAAGTTTGTACAAAAAAGC
P100_HygF	GCGAAGAATCTCGTGCTTTC
P101_HygR	GATGTTGGCGACCTCGTATT
P116_NaNSWAM4_F	TGCAGGCTGAGATGGATAAA
P117_NaNSWAM4_R	CATACACTGGCAGCAGCATT
P118_UbiSWAM4_F	AGCTACGGGGGATTCCTTT
P120_35S_Hyg_F	GAAACCTCCTCGGATTCCAT
P121_35S_Hyg_R	ATAGGTCAGGCTCTCGCTGA
P132_M13F	GTAAAACGACGGCCAG
P133_M13R	CAGGAAACAGCTATGAC
P159_SWAM4PrF1F	CACCGAAGGATGCCATTACAAT
P160_SWAM4PrF4R	TTCAGTGCAATGCAGAACCTG
P161_SWAM4Ex1R	AGTTTAGGGACTGAGCTCCAG
P162_SWAM4Ex3F	CAGTATGGGATTCGAGCCG
P163_CRESR	ACCACTTTGTACAAGAAAGCTGG
P164_SWAM4PrF1R	AGAGGGCATCTCCCTTTTGTTTAC
P165_SWAM4PrF2F	CACCCAATTATCTTTGCAATTATAGGA
P166_SWAM4PrF2R	ACTTCTACGTCACGAGTGTTCCA
P171_SWAM4PrF3F	CACCACAGCACAGCTAACAAGATG
P172_SWAM4PrF3R	CGGCCATGCGTCAGGGTA
P173_SWAM4PrF4F	CACCGAAGCAACCAAGGGAAGA
P176_SWAM4CDS_F	CACCATGGGGAGGCATTCTT
P177_SWAM4CDS_R	CTACATATGCTCAAAAGACAAGGGCA
P178_SWAM1CDS_F	CACCAATGGGGCGGCAC

Primer Name	Sequence
P179_SWAM1CDS_R	GCCTCAAAAGTACTCGAGGTTGAA
P180_pMDC32F_2	CTATCCTTCGCAAGACCCTTC
P191_att1_seq_F	AACAAGTTTGTACAAAAAAGC
P192_LUC_F	AGAACTGCCTGCGTGAGATT
P193_LUC_R	ATCCAGATCCACAACCTTCG
P194_LUC_Rev	ACTGCATACGACGATTCTGTG
P195_NANSWAM4_F	AGCCCTCTCAACATCAATGG
P196_NANSWAM4_R	CAAGGGCATTCTTTGCAAGT
P07_qPCR_Hyg_Fdw	ATTTCGGCTCCAACAATGTC
P08_qPCR_Hyg_Rev	GCGACCTCGTATTGGGAAT
P33_qPCRGNRFCDS_F	CAAGAAGCAGCAGCAACAGCAAC
P34_qPCRGNRFCDS_R	CGCAGCCTGCAACTGTTCATAC
P74_qPCR_GUS_F	CGACATGTGGAGTGAAGAGTATC
P75_qPCR_GUS_R	TCGGCGAAATTCCATACCTG
P124_SRUBIqPCRF	GGAGGCACCTCAGGTCATTT
P125_SRUBIqPCRR	ATAGCGGTCATTGTCTTGCG
P126_SRACTqPCRF	CCTGAAGTCCTTTTCCAGCC
P127_SRACTqPCRR	AGGGCAGTGATCTCCTTGCT
P187_GAPDHqPCR_F	TTGCTCTCCAGAGCGATGAC
P188_GAPDHqPCR_R	CTCCACGACATAATCGGCAC
P189_EF1alfaqPCR_F	CCATCGATATTGCCTTGTGG
P190_EF1alfaqPCR_F	GTCTGGCCATCCTTGGAGAT
qPCRUBC18_F	TCACCCGCAATGACTGTAAG
qPCRUBC18_R	ACCACCATCTGGTCTCCTTC
qPCRGAPDH_F	TCACCCGCAATGACTGTAAG
qPCRGAPDH_R	ACCACCATCTGGTCTCCTTC
qPCRCAD1_F	AGGATAGAATGGGCAGCATCGC
qPCRCAD1_R	ATCTTCAGGGCCTGTCTTCCTGAG
qPCRCOMT4_F	TGGAGAGCTGGTACTACCTGAAG
qPCRCOMT4_R	CGACATCCCGTATGCCTTGTTG
qPCRCESA4_F	GCGTTTCGCATACACCAACACC
qPCRCESA4_R	ACTCGCTAGGTTGTTCAGTGTGG
qPCRCESA7_F	GCGATTCGCCTACATCAACACC
qPCRCESA7_R	GGCTGGCAAATGTGCTAATCGG
qPCRCESA8_F	CAAAGCACAAGTTCCGCCTGTG
qPCRCESA8_R	TGGCTCGTATGCATCTGTCAAATC
P122_UBC18qPCRF	TCACCCGCAATGACTGTAAG
P123_UBC18qPCRR	ACCACCATCTGGTCTCCTTC

Table 3. 2 **Genes in *SWAM4* co-expression cluster.** Seventy-five genes in the co-expression network of SWAM4 (PlaNet). Top genes are related to secondary cell wall formation.

Gene ID	Description
Bradi1g02510	CESA5 - cellulose synthase, (BdCESA9)
Bradi3g39740	Fasciclin-like arabinogalactan protein 7
Bradi1g76970	Homeobox protein knotted-1, putative, expressed (KNOB7)
Bradi4g06317	MYB transcrition factor, BdMYB1
Bradi2g36730	MYB transcrition factor, BdMYB41, SWAM4
Bradi1g20860	3-oxoacyl-synthase
Bradi3g37020	6-phosphofructokinase activity
Bradi2g13800	Actin filament binding, fimbrin-like protein 2
Bradi2g49340	Actin monomer binding, profilin domain containing protein
Bradi1g36220	Alpha/beta hydrolase superfamily
Bradi3g09890	ANKYRIN repeat protein, calcium channel activity
Bradi1g26240	Antiporter activity, MATE efflux family protein
Bradi4g09770	ATP binding, ATP-grasp domain containing protein, expressed
Bradi4g42230	ATPase activity, coupled, rhoGAP domain containing protein
Bradi1g01720	B-cell receptor-associated protein 31
Bradi2g45090	Butyrate response factor 1 (ZFP36L), Zinc finger C-x8-C-x5-C-x3-H type
Bradi1g08130	COBRA-like protein
Bradi1g75470	Cytoplasmic FMR1 interacting protein (CYFIP), PIP Protein-Related
Bradi4g22250	Dirigent, putative, expressed
Bradi3g13990	DNA-dependent ATPase activity
Bradi4g44490	Dual specificity protein tyrosine/serine/threonine phosphatase
Bradi2g01340	EamA-like transporter family (EamA)
Bradi3g27400	F-box domain containing protein
Bradi2g10302	Flavin-containing dimethylaniline monooxygenase
Bradi1g43560	GDP-fucose protein O-fucosyltransferase (O-FucT)
Bradi2g04220	GDP-fucose protein O-fucosyltransferase (O-FucT)
Bradi2g07000	GDSL/SGNH-like Acyl-Esterase family found in Pmr5 and Cas1p
Bradi3g31870	Glutathione S-transferase
Bradi4g05050	Glycosyl hydrolase (GH), subfamily GH28
Bradi3g00910	Glycosyl hydrolase (GH), subfamily GH32
Bradi3g36697	Glycosyl transferase family group 2, Mannan synthase 7-related
Bradi2g56240	HAUS AUGMIN-Like Complex Subunit 7, nuclear matrix protein 1
Bradi3g55277	HEAT repeat family protein, CLASP
Bradi2g04230	Hydrolase, alpha/beta fold family protein, putative, expressed
Bradi2g37340	Inorganic diphosphatase activity
Bradi2g53130	Inositol 5-phosphatase endonuclease/exonuclease/phosphatase
Bradi4g05610	Kelch repeat-containing protein
Bradi1g65340	Leucine-rich repeat protein kinase, subfamily LRR-VI
Bradi1g37210	Leucine-rich repeat protein kinase, subfamily LRR-VI
Bradi3g50570	Leucine-rich repeat protein kinase, subfamily LRR-III
Bradi1g56020	Major Facilitator Superfamily (MFS_1), BT1 family protein
Bradi4g39410	Member of 'GDXG' family of lipolytic enzymes, carboxylesterase
Bradi3g48820	Microtubule binding, HEAT repeat family protein
Bradi2g23290	Microtubule motor activity, kinesin motor domain containing protein
Bradi3g52530	Mitogen-Activated Protein Kinase
Bradi3g51680	NAD Dependent Epimerase/Dehydratase, RmlD substrate binding domain
Bradi3g14260	NAD dependent epimerase/dehydratase, UDP-glucose 4-epimerase activity
Bradi4g35800	NC domain-containing protein, Lecithin retinol acyltransferase
Bradi3g42020	NCK-associated protein 1 (NCKAP1, NAP125)
Bradi4g45037	Omega-3 fatty acid desaturase, chloroplast precursor
Bradi4g02580	OsIAA31 - Auxin-responsive Aux/IAA gene family member
Bradi2g23460	Oxidoreductase, aldo/keto reductase family protein
Bradi2g47280	Phosphatidylinositol binding, clathrin assembly protein
Bradi3g12950	Plasma-membrane choline transporter
Bradi3g01137	Protein kinase
Bradi4g44140	Protein kinase APK1B, chloroplast precursor
Bradi2g45850	Protein kinase family protein, subfamily RLCK-V
Bradi1g60520	Protein serine/threonine phosphatase activity
Bradi1g70290	Putative galacturonosyltransferase, CAZy family GT8
Bradi3g14860	Putative glycosyltransferase CAZy family GT14
Bradi4g07430	Ribosomal Protein S6 Kinase, AGC PVPK like kinases
Bradi3g59496	RING-H2 finger protein ATL1Q
Bradi1g10760	RNA recognition motif containing protein
Bradi5g17900	Vesicle-Associated Membrane Protein, MSP domain
Bradi5g08810	Zinc ion binding, phosphoric diester hydrolase, HIRAN domain
Bradi2g50317	Expressed protein
Bradi4g03730	Expressed protein
Bradi1g59270	Expressed protein
Bradi1g22140	Expressed protein
Bradi5g24560	Expressed protein
Bradi3g41830	Expressed protein
Bradi4g06680	Expressed protein
Bradi2g26270	Expressed protein
Bradi4g34840	Expressed protein

Table 3. 3 **Replicate samples of WT, *SWAM4-OE*, *SWAM4-DR*, *swam4-1*, and *swam4-2* mutants for RNA-seq analysis.**

Genotype	Replicate 1	Replicate 2	Replicate 3	Replicate 4	Total samples
SWAM4-OE	SR73	SR74	SR83	SR84	4
SWAM4-DR	SR136	SR141	SR142	SR143	4
swam4-1	SR180	SR184			2
swam4-2	SR187	SR188	SR190	SR192	4
WT	SR65	SR66	SR67	SR68	4

Table 3. 4 **List of a subset group of genes differentially expressed in *SWAM4-OE*, *SWAM4-DR*, *swam4-1* and *swam4-2* plants as identified by RNA-seq (p<0.01)**

Subcategories

Genes grouped in the same co-expression cluster (PlaNet)
Genes of interest related to cell wall formation
Downregulation
Upregulation
Genes differentially expressed in *GNRF-OE* and *gnrf* plants identified by RNA-seq. Chapert 2 (Table 2.5)

Gene ID	Gene Description	RNA-seq (Fold)			
		SWAM4-OE	*SWAM4-DR*	*swam4-1*	*swam4-2*
Bradi5g26380	Unknown	-7.94	-8.52	1.56	
Bradi4g01506	Unknown	-7.32	-6.12		
Bradi3g08301	RNA helicase	-6.81	-9.06		
Bradi4g11886	Unknown	-6.66	-3.08		
Bradi1g09690	Glycosyl hydrolase (GH), subfamily GH16	-6.47	-6.05		
Bradi1g20715	ACIN1, ACINUS, apoptotic chromatin condensation inducer in the nucleus	-6.28	-6.58		
Bradi5g26375	Protease do_like9	-5.91	-4.77		
Bradi3g22870	hydrogen-exporting ATPase activity,	-5.66	-4.76		
Bradi2g23797	A/G-SPECIFIC ADENINE GLYCOSYLASE/ENDONUCLEASE III	-3.45	-3.16		
Bradi3g48970	Histone lysine N-methyltransferase activity (H3-K9 specific), (H3-K4 specific)	-2.85	-4.24		
Bradi2g02830	Unknown	-2.69	-3.54	1.87	
Bradi3g18640	Unknown	-2.06	-1.31		
Bradi3g11270	Alpha-glucan, water dikinase	-1.99	-5.42		
Bradi1g58930	Protein kinase	-1.68	-3.79		-1.11
Bradi4g00467	AAA-Type ATPase Family Protein-Related	-1.41		1.73	-0.77
Bradi2g33690	Unknown	-1.26	-1.44		
Bradi1g44930	Glycosyl hydrolase (GH), BdGH36_1	-1.25	-2.09	-4.19	
Bradi4g27383	Unknown	-1.18	-1.50		-2.66
Bradi5g15587	No apical meristem (NAM) protein (NAM) BdNAC97	-1.17	-1.46		-0.93
Bradi1g66490	Rab GDP-dissociation inhibitor	-0.91	-1.46		
Bradi2g42840	KOG0156 - Cytochrome P450 CYP2 subfamily	-0.84	-2.47		
Bradi3g30193	UYnknown	-0.83	-1.14		
Bradi2g17530	Lipid transfer protein	-0.80	-0.74	-2.13	
Bradi1g65780	Hydrophobic Protein RCI2	-0.80	-1.41		
Bradi4g30060	Light-harvesting complex I chlorophyll a/b binding protein 4 (LHCA4)	-0.79		-2.49	-0.80
Bradi3g22786	Cycloartenol synthase / 2,3-epoxysqualene--cycloartenol cyclase	-0.75	-1.04		-0.91
Bradi3g50947	Protein T01H10.8	-0.72	-1.38		0.60
Bradi1g26790	UDP-Glucosyl transferase 88A1	-0.72	-2.11		
Bradi3g30590	Ferulic acid 5-hydroxylase 1 (FAH1)	-0.70	-1.53	0.72	
Bradi2g13110	Glutathione s-transferase, GST	-0.69		-1.87	-0.59
Bradi4g24131	Unknown	-0.66	-1.01		
Bradi2g58616	Protein O-GlcNAc transferase (OGT)	-0.66	-0.51		
Bradi5g10640	NAC transcrition factor, XND1-like BdNAC89	-0.65		-1.32	
Bradi1g64560	Glycosyltransferase, GT34 family, xylosyltransferase	-0.62	-1.21		-0.60
Bradi2g34650	Fasciclin-like arabinogalactan protein	-0.59	-1.72		-0.92
Bradi1g69870	ATPase activity	-0.57	-1.10		
Bradi1g25117	CSLF2 - cellulose synthase-like family F	-0.56	-1.55		
Bradi5g22800	Aldo-keto reductase	-0.55	-1.21	-0.86	
Bradi1g24760	Light-harvesting complex II chlorophyll a/b binding protein 3	-0.55		-2.85	-0.56
Bradi5g09610	Probable lipid transfer	-0.54	-0.94	-2.70	
Bradi4g27760	TPX2 (Targeting Protein For XKLP2)	-0.54	-3.43		-0.62
Bradi1g66720	Laccase 5	-0.54	-1.29		
Bradi3g53494	Unknown	-0.54	-1.74		
Bradi2g54680	Lacasse 10	-0.54	-1.17		-0.63
Bradi3g04460	Polysaccharide biosynthesis	-0.54	-1.26		-0.47
Bradi3g16530	COMT6, O-methyltransferase	-0.53	-1.27		-0.51
Bradi3g04080	Glycosylhydrolase, GH9 family glycosyl hydrolase 9B8	-0.53	-1.22		
Bradi1g37220	Unknown	-0.53	-1.14		-0.53
Bradi2g54970	Probable lipid transfer (LTP_2)Links	-0.53			
Bradi1g32990	Phosphosulfolactate synthase	-0.50	-0.85	-0.91	
Bradi5g23460	ATP binding	-0.50	-0.95	1.11	
Bradi1g43990	Constans-like, BdCO16-like	-0.50		-1.58	-0.81
Bradi1g61577	K13083 - flavonoid 3',5'-hydroxylase	-0.49	-1.22		
Bradi3g58560	Copper ion binding	-0.48	-2.58		

Bradi2g45090	Butyrate response factor 1 (ZFP36L), Zinc finger C-x8-C-x5-C-x3-H type	-0.48	-0.65		-0.57
Bradi2g45220	Zinc Finger Fyve domain containing protein	-0.47	-0.58		
Bradi2g33980	Unknown	-0.47	-0.41	1.01	
Bradi3g33070	QWRF motif- containing protein 7	-0.47	-0.87		-0.48
Bradi1g67230	Zinc ion binding	-0.47	-0.67		
Bradi2g08310	UDP-galactosyltransferase, UDP-glucosyl transferase 71C4	-0.47	-0.79		
Bradi1g52610	Glycosyltransferase, BdGT77E8	-0.47	-0.41		
Bradi4g16560	Steroid 17-alpha-monooxygenase activity	-0.46	-1.13		
Bradi1g72350	Glycosyltransferase, CAZy family GT8	-0.46	-1.50	-1.06	
Bradi3g56787	Pectate lyase superfamily protein	-0.46	-1.15		
Bradi5g18880	Predicted transporter/transmembrane protein	-0.45	-1.33	1.15	
Bradi3g49250	Phenylalanine ammonia-lyase, BdPAL1	-0.45	-1.06		
Bradi2g12150	S-adenosylmethionine synthetase	-0.45	-1.16		-0.44
Bradi2g36910	PMT Transferase family	-0.45	-0.70		
Bradi2g08790	Manganese ion binding, Cupin domain	-0.43	-1.20	-2.07	-0.62
Bradi3g06480	Cinnamyl alcohol dehydrogenase, CAD1	-0.42	-0.99	0.67	
Bradi3g05240	Aminomethyltransferase activity	-0.42	-1.02		
Bradi1g60750	Protein ESKIMO 1	-0.42	-1.22		
Bradi1g01820	3-deoxy-7-phosphoheptulonate synthase	-0.42	-0.84		
Bradi3g58830	Phospholipase A2 inhibitor	-0.42	-0.98	-2.32	
Bradi3g06050	Chorismate Mutase 2	-0.42		-0.68	-0.38
Bradi1g36770	Zinc-Fingers And Homeoboxes Related	-0.41			
Bradi1g46670	Putative sucrose synthase. Glycosyltransferase in family GT4, BdSUS3	-0.40	-0.58		
Bradi4g30540	CESA9 - cellulose synthase, (BdCESA7)	-0.40	-0.90		-0.41
Bradi1g43090	DNAJ Homolog Subfamily C Member	-0.39	-0.71		
Bradi2g57567	Protein IQ-domain 14-related	-0.39	-0.99		
Bradi3g53480	Unknown	-0.39	-0.47		
Bradi2g49912	Cellulose Synthase A Catalytic Subunit 8, (BdCESA8)	-0.38	-0.86		-0.42
Bradi4g21240	Glucuronoxylan 4-O-methyltransferase (GXM)	-0.38	-0.91		-0.64
Bradi3g48850	ATP-dependent RNA helicase activity	-0.38	-0.32		
Bradi3g28350	CESA7 - cellulose synthase, (BdCESA4)	-0.37	-0.98		-0.46
Bradi3g07047	Hypothetical protein (K09955)	-0.37	-0.61		
Bradi3g14277	Zinc Transport Protein ZNTB	-0.36	-0.60		
Bradi2g50010	3-hydroxyisobutyryl-CoA hydrolase	-0.36	-0.51		
Bradi2g11664	cAMP-regulated phosphoprotein	-0.36	-0.83		-0.54
Bradi1g60320	Putative sucrose synthase. Glycosyltransferase in family GT4	-0.36	-0.78		-0.41
Bradi1g31820	RING FINGER Protease, VACUOLAR-SORTING RECEPTOR 5-RELATED	-0.35	-0.61		-0.43
Bradi4g01200	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	-0.35	-0.92		
Bradi1g08301	Mitochondrial Substrate Carrier Family Protein	-0.35	-0.63		-0.57
Bradi1g64830	Galacturonosyltransferase-Like 6-Related	-0.35	-1.21		
Bradi1g75730	Alkane 1-monooxygenase / Omega-hydroxylase	-0.34	-1.56		-0.41
Bradi3g14080	Glycosyltransferase, CAZy family GT2	-0.33			
Bradi1g74590	Phosphoserine transaminase activity	-0.32	-0.52		
Bradi1g24830	Unknown	-0.30	-1.13		
Bradi1g68710	Motor activity, ATPase activity	-0.27	-1.09		-0.23
Bradi1g44560	Unknown	-0.25	-0.60		-0.32
Bradi1g10150	Tubulin/FtsZ domain containing protein	-0.25	-0.72		-0.28
Bradi3g39880	Unknown	-0.21	-0.62		
Bradi2g45520	Glucose transmembrane transporter activity		-5.64		-2.35
Bradi2g20830	Peroxidase / Lactoperoxidase		-4.74		-1.02
Bradi4g11850	Laccase 24		-4.66		1.91
Bradi4g41616	Ethylene-Responsive transcription factor 15-related		-2.82	-3.68	-2.74
Bradi2g49340	Actin monomer binding, profilin domain containing protein		-1.85		
Bradi3g31870	Glutathione S-transferase		-1.75		
Bradi3g33140	Rare lipoprotein A		-1.68	-1.61	
Bradi2g23370	Laccase 8		-1.60		
Bradi1g03500	Proton-dependent oligopeptide transporter		-1.52		
Bradi3g39800	High affinity sodium:dicarboxylate symporter activity		-1.50		
Bradi5g17900	Vesicle-Associated Membrane Protein, MSP domain		-1.33	-1.12	
Bradi1g00710	Lzipper-MIP1		-1.33		
Bradi2g17067	Predicted membrane protein, DoH and Cytochrome b-561/ferric reductase		-1.23		-0.29
Bradi2g55050	Laccase 13		-1.23		
Bradi2g18447	Sulfotransferase domain containing protein		-1.21		
Bradi2g00220	Fasciclin-like arabinoogalactan protein		-1.16		-0.49
Bradi2g16560	Fasciclin domain containing protein		-1.16		
Bradi4g33490	Fasciclin-like arabinoogalactan protein		-1.14		
Bradi4g29770	Cinnamyl alcohol dehydrogenase, CAD4		-1.08		
Bradi4g22250	Dirigent protein related		-1.06		

Bradi2g01340	EamA-like transporter family (EamA)		-1.05		
Bradi3g42430	MYB transcrition factor, BdMYB69, SWAM8		-1.03		
Bradi2g10970	Tubulin/FtsZ domain containing protein		-1.00		
Bradi5g04540	RING, subfamily zinc finger (C3HC4-type RING finger) family protein		-0.99		-0.34
Bradi4g40400	Glucuronoxylan 4-0 Methlytransferase 2 related		-0.98		
Bradi3g41830	Expressed protein		-0.97	-0.97	
Bradi1g09460	Endoglucanase		-0.96		
Bradi1g59880	COBRA-like protein		-0.92		-0.42
Bradi1g67460	Phospholipase A2		-0.92	-1.84	
Bradi3g55277	HEAT repeat family protein, CLASP		-0.88	-3.30	
Bradi2g04230	Hydrolase, alpha/beta fold family protein		-0.88		
Bradi2g23300	Expressed protein		-0.87		
Bradi3g05750	4-Coumarate:CoA ligase (4CL)		-0.85		
Bradi1g57040	Fasciclin domain containing protein		-0.85		
Bradi3g26690	BEL1-LIKE homeodomain transcription factor		-0.81	0.86	
Bradi5g14720	Hydroxycinnamoyl-CoA:shikimate/quinate hydroxycinnamoyltransferase HCT		-0.80		
Bradi3g45160	Harpin-induced protein 1 domain containing protein		-0.80		
Bradi4g34040	CHIT13 - Chitinase family protein precursor		-0.79		
Bradi3g49260	Phenylalanine ammonia-lyase, BdPAL2		-0.78	0.84	
Bradi1g31320	4-Coumarate:CoA ligase (4CL)		-0.78		
Bradi4g25540	BTB9 - Bric-a-Brac, Tramtrack, Broad Complex BTB domain		-0.78		-0.64
Bradi2g59410	Xylosyltransferase, CAZy family GT47-2		-0.78		
Bradi2g55340	Transmembrane amino acid transporter protein		-0.76		
Bradi3g28920	UDP-glucuronic acid transporter		-0.75		
Bradi5g15527	Glycosyl Hydrolase		-0.74		
Bradi3g37530	Ion channel activity, ferric reductase		-0.73		
Bradi3g39420	Caffeoyl CoA 3-O-methyltransferase, CCoACOMT1		-0.73		
Bradi2g37970	Xylosyltransferase, CAZy family GT43-2		-0.71		
Bradi2g46197	NAC Domain Containing Protein 73, GNRF		-0.71		-0.40
Bradi1g76970	Homeobox protein knotted-1, putative, expressed (KNOB7)		-0.70		-0.41
Bradi4g36240	Endoglucanase		-0.70		
Bradi1g27920	Peroxidase 53-related		-0.69		
Bradi3g54370	Glycosyl hydrolase (GH), subfamily GH79		-0.69	-1.65	
Bradi1g06290	Fasciclin domain containing protein		-0.68		
Bradi2g53470	Trans-cinnamate 4-hydroxylase, C4H1		-0.68		
Bradi2g10302	Flavin-containing dimethylaniline monooxygenase		-0.67		
Bradi2g12370	GDSL-like lipase/acylhydrolase		-0.65		
Bradi2g23460	Oxidoreductase, aldo/keto reductase family protein		-0.64	-1.00	
Bradi2g04220	GDP-fucose protein O-fucosyltransferase (O-FucT)		-0.64	-0.85	
Bradi2g37000	Peroxidase / Lactoperoxidase		-0.64		
Bradi1g54250	CESA8 - cellulose synthase, (BdCESA3)		-0.64		
Bradi1g72430	Strubbeling-Receptor family 6 precursor		-0.64		
Bradi4g31130	Ferric reductase		-0.63	-1.00	
Bradi4g28260	Glycosyltransferase, GT77 family, Extensin		-0.63	0.92	
Bradi2g30490	Lipid transfer protein		-0.63	-1.47	
Bradi2g10960	Copper ion binding, plastocyanin-like domain		-0.62		
Bradi1g02510	CESA5 - cellulose synthase, (BdCESA9)		-0.62		
Bradi5g20130	MYB transcrition factor, BdMYB104, SWAM7		-0.62		
Bradi2g21300	Cytochrome P450		-0.60		
Bradi2g52790	Tubulin/FtsZ domain containing protein		-0.59		
Bradi4g19457	Adenosylhomocysteinase / SAHase		-0.58		
Bradi1g59830	L-amino acid transmembrane transporter activity		-0.57		
Bradi1g47767	Inorganic H+ pyrophosphatase		-0.57		
Bradi3g48820	Microtubule binding, HEAT repeat family protein		-0.57	-0.76	
Bradi3g14260	NAD dependent epimerase/dehydratase family protein		-0.48		
Bradi3g51280	Tetracycline:hydrogen antiporter activity		-0.37	1.07	
Bradi4g16327	BSD domain-containing protein		-0.34		
Bradi1g48070	Calcium-dependent cysteine-type endopeptidase activity	0.31	0.55	0.93	0.30
Bradi1g29060	Cellulose synthase, (BdCesA5)	0.42			
Bradi2g24910	Hydrogen:amino acid symporter	0.41			
Bradi3g55270	Zinc Finger Fyve Domain Containing Protein	0.45		-0.93	0.73
Bradi1g50970	Zinc Finger Fyve Domain Containing Protein	0.51		-0.99	0.62
Bradi2g36700	Agenet domain	0.52		1.17	0.44
Bradi1g55620	Early Light-Induced Protein 1, Chloroplastic-Related	0.59		1.36	0.79
Bradi1g27800	Rho guanyl-nucleotide exchange factor activity	0.62		-1.80	0.74
Bradi2g48690	MADS transcription factor, BdMAD28	0.95	1.38		
Bradi2g25490	EamA-like transporter family	1.89	2.74	4.04	1.47
Bradi1g78100	Arsenite secondary active transmembrane transporter	1.90	2.34		1.23
Bradi2g36730	MYB family transcription factor, SWAM4		2.78		
Bradi4g04662	Leucine-Rich repeat-containing Protein	8.95	8.78	8.98	5.41

CHAPTER 4

CONCLUSIONS

Plant cell walls are complex structures that contain essential components for plant growth, cell shape, and cell differentiation. The organization and biosynthesis of the matrix of cellulose, lignin, and hemicellulose within the cell wall require an orchestrated mechanism of transcription factors that regulate this molecular process. Functional genetic studies in the eudicot plant *Arabidopsis thaliana* have identified an extensive collection of transcription factors involved in the biosynthesis of cell wall components that share protein domains and are grouped in protein families, work in a transcriptional regulatory network, and are the foundation of the functional characterization of several transcription factors in other plant species.

Several transcription factors of the NAC and MYB protein families have been identified in *A. thaliana*, and the functional characterization of their orthologous genes in grasses would provide information about commonalities in their molecular mechanism. Evidence suggests that genes controlling this process may be different between eudicots and monocots, and little is known about the transcriptional regulation of secondary cell wall thickening in the monocot *Brachypodium distachyon*. *GNRF* (*GRASS NAC REPRESSOR OF FLOWERING*), *SWAM1* (*SECONDARY WALL ASSOCIATED MYB1*), and *SWAM4* were identified for functional characterization in this study.

Phylogenetic analysis identified *GNRF* as the *A. thaliana SND2* orthologous transcription factor in the grass *Brachypodium distachyon*. Plants that overexpressed *GNRF (GNRF-OE)*

remained in a prolonged juvenile stage. Five *GNRF* mutant alleles (*gnrf-1, gnrf-2, gnrf-3, gnrf-4,* and *gnrf-5*) from a TILLING (Targeting Induced Local Lesion IN Genome) collection and a mutant line (*gnrf-6*) carrying a T-DNA insertion in the *GNRF 5'UTR* were identified. *GNRF-OE* and *gnrf* mutants stems were subjected to lignin quantification, cell wall thickness measurements, Q-RT-PCR, and RNA-seq analysis. Since mutant alleles with non-synonymous mutations are heavily mutagenized and were not complemented, the results obtained with those reagents were used comparatively. Cell wall and transcriptomic analysis revealed that *GNRF* is a repressor of *SWAM1*, a MYB activator of cell wall thickening, and represses genes encoding cellulose, lignin, and xylan biosynthetic enzymes. *GNRF* was found to function as a pleiotropic repressor of cell wall biosynthesis, flowering and transport proteins. Protein-DNA interactions by a luciferase-compatible yeast-one-hybrid assay and a DNA affinity purification sequencing (DAP-seq) assay were used to identify the *GNRF* binding site (CT/GTA/G/CA/TNNNNT/G/CAA/CA/T/GA/TA/T). Collectively, these data indicate that *GNRF* has a pleotropic function in the repression of cell wall biosynthesis and flowering. *GNRF* forms a negative feedback loop with *SWAM1* to regulate secondary wall thickening in *B. distachyon*.

SWAM4 is an ortholog of *OsMYB61b* in rice and is closely related to *AtMYB50* and *AtMYB61* in *A. thaliana*. *SWAM4* relative transcript abundance is higher in stems, compared to roots and leaves and a co-expression analysis showed that *SWAM4* is clustered with putative cell-wall-associated genes. Overexpression of *SWAM4* (*SWAM4-OE*), dominant repressor lines *(SWAM4-DR*) and two mutant alleles (*swam4-1* and *swam4-2*) from a TILLING (Targeting Induced Local Lesion IN Genome) collection were included in the study. *SWAM4-DR* lines were

dramatically shorter than control plants and showed significantly lower stem and leaf biomass. *SWAM4-OE, SWAM4-DR*, and *swam4* mutant stems were subjected to lignin quantification, cell wall thickness measurements, Q-RT-PCR, and RNA-seq analysis. Similar to *SWAM1-OE* plants, *SWAM4-OE* plants did not show upregulation of *SWAM4* as has been expected. However, cell wall analysis revealed that *SWAM4* may be associated with the regulation of lignin due to a reduction of lignin content and cell wall thickening of interfascicular fibers in *SWAM4-DR* plants. Transcriptomic analysis revealed putative targets of *SWAM4* in *SWAM4-DR* samples that include cellulose and lignin biosynthetic genes (*CESA4*, *CESA7*, *CESA8*, *CAD* and *COMT*), glycosyltransferases, glycosylhydrolases, laccases, and other proteins associated with cell wall formation. Collectively, these data suggest that *SWAM4* could be a regulator of lignin and cellulose biosynthesis in *B. distachyon*.

GNRF, *SWAM1, and* potentially *SWAM4* were found to be transcriptional regulators of secondary cell wall formation in *B. distachyon* (Figure 4.1). Further analyses are required to confirm binding interactions *in planta*. This work provides an insight to reveal details of cell wall biosynthesis in grasses that may be either unique to grasses or comparable to other plant species to understand this process better.

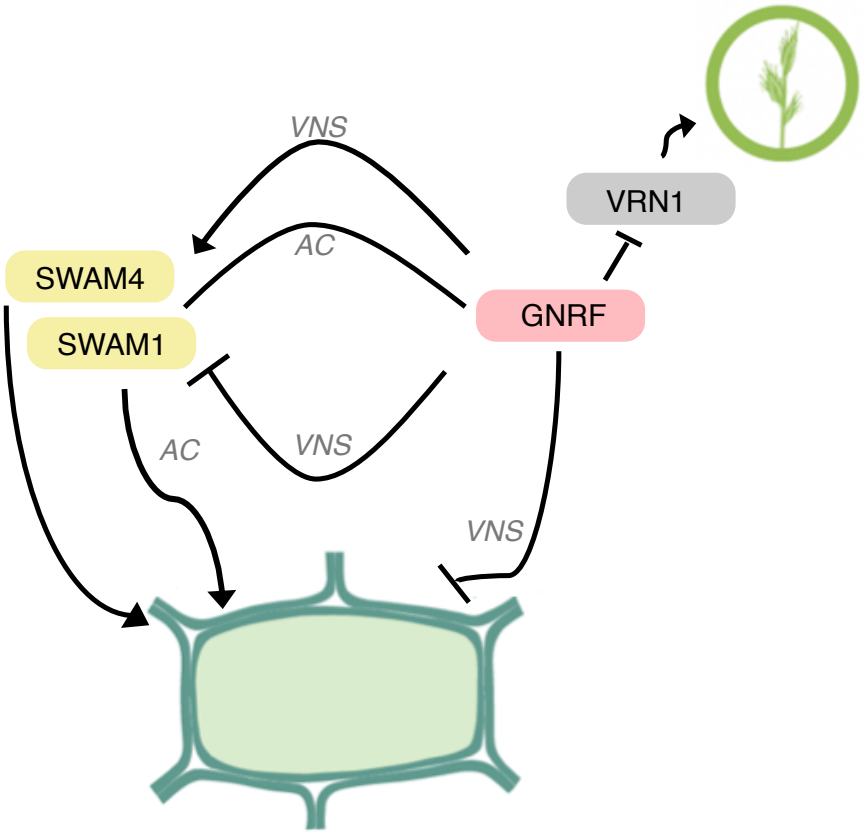


Figure 5.1 **Model of GNRF, SWAM4 and SWAM1 regulatory functions**. *GNRF* is a repressor of *SWAM1*, a MYB activator of cell wall biosynthesis; hence, GNRF and SWAM1 regulate cellulose, lignin, and xylan biosynthesis by their interaction in a negative feedback loop. In addition, the GNRF is a pleiotropic repressor of cell wall biosynthesis, flowering and transport proteins. Similar to *SWAM1*, *SWAM4* regulates lignin, cellulose, xylan biosynthetic genes and possibly other transcription factors (*GNRF*, *KNOB7*) in *B. distachyon.* In addition, the GNRF binding site is a VNS (VND, NAC, SND) binding associated motif, and SWAM4 and SWAM1 bind to a AC-type element.

REFERENCES

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Arvidsson, S., Kwasniewski, M., Riano-Pachon, D., and Mueller-Roeber, B. (2008). QuantPrime - a flexible tool for reliable high-throughput primer design for quantitative PCR. Bmc Bioinformatics *9*, 465.

Barrett, L.W., Fletcher, S., and Wilton, S.D. (2012). Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. Cell Mol Life Sci *69*, 3613-3634.

Bashline, L., Lei, L., Li, S., and Gu, Y. (2014). Cell wall, cytoskeleton, and cell expansion in higher plants. Molecular plant *7*, 586-600.

Betekhtin, A., Rojek, M., Nowak, K., Pinski, A., Milewska-Hendel, A., Kurczynska, E., Doonan, J.H., and Hasterok, R. (2018). Cell Wall Epitopes and Endoploidy as Reporters of Embryogenic Potential in Brachypodium Distachyon Callus Culture. Int J Mol Sci *19*.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.

Bomal, C., Bedon, F., Caron, S., Mansfield, S.D., Levasseur, C., Cooke, J.E., Blais, S., Tremblay, L., Morency, M.J., Pavy, N.*, et al.* (2008). Involvement of Pinus taeda MYB1 and MYB8 in phenylpropanoid metabolism and secondary cell wall biogenesis: a comparative in planta analysis. J Exp Bot *59*, 3925-3939.

Bonaldi, K., Li, Z., Kang, S.E., Breton, G., and Pruneda-Paz, J.L. (2017). Novel cell surface luciferase reporter for high-throughput yeast one-hybrid screens. Nucleic acids research *45*, e157.

Bragg, J.N., Wu, J., Gordon, S.P., Guttman, M.E., Thilmony, R., Lazo, G.R., Gu, Y.Q., and Vogel, J.P. (2012). Generation and characterization of the Western Regional Research Center Brachypodium T-DNA insertional mutant collection. Plos One *7*, e41916.

Cai, B., Li, C.H., and Huang, J. (2014). Systematic identification of cell-wall related genes in Populus based on analysis of functional modules in co-expression network. Plos One *9*, e95176.

Chen, A., and Dubcovsky, J. (2012). Wheat TILLING mutants show that the vernalization gene VRN1 down-regulates the flowering repressor VRN2 in leaves but is not essential for flowering. Plos Genet *8*, e1003134.

Cosgrove, D.J. (2005). Growth of the plant cell wall. Nature reviews Molecular cell biology *6*, 850-861.
Cosgrove, D.J., and Jarvis, M.C. (2012). Comparative structure and biomechanics of plant primary and secondary cell walls. Frontiers in plant science *3*, 204.

Dalmais, M., Antelme, S., Ho-Yue-Kuang, S., Wang, Y., Darracq, O., d'Yvoire, M.B., Cezard, L., Legee, F., Blondet, E., Oria, N.*, et al.* (2013). A TILLING Platform for Functional Genomics in. Plos One *8*, e65503.

Deplancke, B., Dupuy, D., Vidal, M., and Walhout, A.J. (2004). A gateway-compatible yeast one-hybrid system. Genome Res *14*, 2093-2101.

Deplancke, B., Vermeirssen, V., Arda, H.E., Martinez, N.J., and Walhout, A.J. (2006). Gateway-compatible yeast one-hybrid screens. CSH Protoc *2006*.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.
Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in Arabidopsis. Trends in plant science *15*, 573-581.

Farrokhi, N., Burton, R.A., Brownfield, L., Hrmova, M., Wilson, S.M., Bacic, A., and Fincher, G.B. (2006). Plant cell wall biosynthesis: genetic, biochemical and functional genomics approaches to the identification of key genes. Plant biotechnology journal *4*, 145-167.

Ferreira, S.S., Hotta, C.T., Poelking, V.G., Leite, D.C., Buckeridge, M.S., Loureiro, M.E., Barbosa, M.H., Carneiro, M.S., and Souza, G.M. (2016). Co-expression network analysis reveals transcription factors associated to cell wall biosynthesis in sugarcane. Plant molecular biology *91*, 15-35.

Fornale, S., Shi, X., Chai, C., Encina, A., Irar, S., Capellades, M., Fuguet, E., Torres, J.L., Rovira, P., Puigdomenech, P.*, et al.* (2010). ZmMYB31 directly represses maize lignin genes and redirects the phenylpropanoid metabolic flux. The Plant journal : for cell and molecular biology *64*, 633-644.

Foster, C.E., Martin, T.M., and Pauly, M. (2010). Comprehensive compositional analysis of plant cell walls (Lignocellulosic biomass) part I: lignin. J Vis Exp.

Girin, T., David, L.C., Chardin, C., Sibout, R., Krapp, A., Ferrario-Mery, S., and Daniel-Vedele, F. (2014). Brachypodium: a promising hub between model species and cereals. J Exp Bot *65*, 5683-5696.

Grant, E.H., Fujino, T., Beers, E.P., and Brunner, A.M. (2010). Characterization of NAC domain transcription factors implicated in control of vascular cell differentiation in Arabidopsis and Populus. Planta *232*, 337-352.
Guo, K., Zou, W., Feng, Y., Zhang, M., Zhang, J., Tu, F., Xie, G., Wang, L., Wang, Y., Klie, S.*, et al.* (2014). An integrated genomic and metabolomic framework for cell wall biology in rice. BMC Genomics *15*, 596.

Handakumbura, P.P., Brow, K., Whitney, I.P., Zhao, K., Sanguinet, K.A., Lee, S.J., Olins, J., Romero-Gamboa, S.P., Harrington, M.J., Bascom, C.J.*, et al.* (2018). SECONDARY WALL ASSOCIATED MYB1 is a positive regulator of secondary cell wall thickening in Brachypodium distachyon and is not found in the Brassicaceae. The Plant journal : for cell and molecular biology *96*, 532-545.

Handakumbura, P.P., and Hazen, S.P. (2012). Transcriptional Regulation of Grass Secondary Cell Wall Biosynthesis: Playing Catch-Up with Arabidopsis thaliana. Frontiers in plant science *3*, 74.

Handakumbura, P.P., Matos, D.A., Osmont, K.S., Harrington, M.J., Heo, K., Kafle, K., Kim, S.H., Baskin, T.I., and Hazen, S.P. (2013). Perturbation of Brachypodium distachyon CELLULOSE SYNTHASE A4 or 7 results in abnormal cell walls. BMC plant biology *13*.

Hatfield, R., and Vermerris, W. (2001). Lignin formation in plants. The dilemma of linkage specificity. Plant Physiol *126*, 1351-1357.

He, J.B., Zhao, X.H., Du, P.Z., Zeng, W., Beahan, C.T., Wang, Y.Q., Li, H.L., Bacic, A., and Wu, A.M. (2018). KNAT7 positively regulates xylan biosynthesis by directly activating IRX9 expression in Arabidopsis. J Integr Plant Biol *60*, 514-528.

Hirano, K., Aya, K., Morinaka, Y., Nagamatsu, S., Sato, Y., Antonio, B.A., Namiki, N., Nagamura, Y., and Matsuoka, M. (2013a). Survey of genes involved in rice secondary cell wall formation through a co-expression network. Plant Cell Physiol *54*, 1803-1821.

Hirano, K., Kondo, M., Aya, K., Miyao, A., Sato, Y., Antonio, B.A., Namiki, N., Nagamura, Y., and Matsuoka, M. (2013b). Identification of Transcription Factors Involved in Rice Secondary Cell Wall Formation. Plant Cell Physiol *54*, 1791-1802.

Hiratsu, K., Matsui, K., Koyama, T., and Ohme-Takagi, M. (2003). Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in Arabidopsis. The Plant journal : for cell and molecular biology *34*, 733-739.

Hong, S.-Y., Seo, P., Yang, M.-S., Xiang, F., and Park, C.-M. (2008). Exploring valid reference genes for gene expression studies in Brachypodium distachyon by real-time PCR. BMC plant biology *8*, 112.

Hong, S.Y., Park, J.H., Cho, S.H., Yang, M.S., and Park, C.M. (2011). Phenological growth stages of Brachypodium distachyon: codification and description. Weed Res *51*, 612-620.

Hu, R., Qi, G., Kong, Y., Kong, D., Gao, Q., and Zhou, G. (2010). Comprehensive analysis of NAC domain transcription factor gene family in Populus trichocarpa. BMC plant biology *10*, 145.

Huang, D., Wang, S., Zhang, B., Shang-Guan, K., Shi, Y., Zhang, D., Liu, X., Wu, K., Xu, Z., Fu, X.*, et al.* (2015). A Gibberellin-Mediated DELLA-NAC Signaling Cascade Regulates Cellulose Synthesis in Rice. Plant Cell *27*, 1681-1696.

Huang, J., Guo, Y., Sun, Q., Zeng, W., Li, J., Li, X., and Xu, W. (2018). Genome-Wide Identification of R2R3-MYB Transcription Factors Regulating Secondary Cell Wall Thickening in Cotton Fiber Development. Plant Cell Physiol.

Hussey, S.G., Mizrachi, E., Creux, N.M., and Myburg, A.A. (2013). Navigating the transcriptional roadmap regulating plant secondary cell wall deposition. Frontiers in plant science *4*, 325.
Hussey, S.G., Mizrachi, E., Spokevicius, A.V., Bossinger, G., Berger, D.K., and Myburg, A.A. (2011a). SND2, a NAC transcription factor gene, regulates genes involved in secondary cell wall development in Arabidopsis fibres and increases fibre cell area in Eucalyptus. BMC plant biology *11*.

Jensen, J.K., and Wilkerson, C.G. (2017). Brachypodium as an experimental system for the study of stem parenchyma biology in grasses. Plos One *12*, e0173095.

Kagale, S., and 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.

Katoh, K., Rozewicki, J., and Yamada, K.D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in bioinformatics.

Kaufmann, K., Muino, J.M., Osteras, M., Farinelli, L., Krajewski, P., and Angenent, G.C. (2010). Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). Nat Protoc *5*, 457-472.

Kebrom, T.H., McKinley, B., and Mullet, J.E. (2017). Dynamics of gene expression during development and expansion of vegetative stem internodes of bioenergy sorghum. Biotechnology for biofuels *10*, 159.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Methods *12*, 357-360.

Kim, Y., Lee, G., Jeon, E., Sohn, E.J., Lee, Y., Kang, H., Lee, D.W., Kim, D.H., and Hwang, I. (2014). The immediate upstream region of the 5'-UTR from the AUG start codon has a pronounced effect on the translational efficiency in Arabidopsis thaliana. Nucleic acids research *42*, 485-498.
Kubo, M., Udagawa, M., Nishikubo, N., Horiguchi, G., Yamaguchi, M., Ito, J., Mimura, T., Fukuda, H., and Demura, T. (2005). Transcription switches for protoxylem and metaxylem vessel formation. Genes & development *19*, 1855-1860.

Legay, S., Sivadon, P., Blervacq, A.S., Pavy, N., Baghdady, A., Tremblay, L., Levasseur, C., Ladouce, N., Lapierre, C., Seguin, A.*, et al.* (2010). EgMYB1, an R2R3 MYB transcription factor from eucalyptus negatively regulates secondary cell wall formation in Arabidopsis and poplar. New Phytol *188*, 774-786.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. Bmc Bioinformatics *12*, 323.

Li, C., Ma, X., Yu, H., Fu, Y., and Luo, K. (2018). Ectopic Expression of PtoMYB74 in Poplar and Arabidopsis Promotes Secondary Cell Wall Formation. Frontiers in plant science *9*, 1262.

Li, E., Bhargava, A., Qiang, W., Friedmann, M.C., Forneris, N., Savidge, R.A., Johnson, L.A., Mansfield, S.D., Ellis, B.E., and Douglas, C.J. (2012). The Class II KNOX gene KNAT7 negatively regulates secondary wall formation in Arabidopsis and is functionally conserved in Populus. New Phytol *194*, 102-115.

Liang, Y.K., Dubos, C., Dodd, I.C., Holroyd, G.H., Hetherington, A.M., and Campbell, M.M. (2005). AtMYB61, an R2R3-MYB transcription factor controlling stomatal aperture in Arabidopsis thaliana. Curr Biol *15*, 1201-1206.

Liu, Y.Y., You, S.J., Taylor-Teeples, M., Li, W.L., Schuetz, M., Brady, S.M., and Douglas, C.J. (2014). BEL1-LIKE HOMEODOMAIN6 and KNOTTED ARABIDOPSIS THALIANA7 Interact and Regulate Secondary Cell Wall Formation via Repression of REVOLUTA. Plant Cell *26*, 4843-4861.

Ma, X.F., Jensen, E., Alexandrov, N., Troukhan, M., Zhang, L., Thomas-Jones, S., Farrar, K., Clifton-Brown, J., Donnison, I., Swaller, T.*, et al.* (2012). High resolution genetic mapping by genome sequencing reveals genome duplication and tetraploid genetic structure of the diploid Miscanthus sinensis. Plos One *7*, e33821.

Matias-Hernandez, L., Jiang, W., Yang, K., Tang, K., Brodelius, P.E., and Pelaz, S. (2017). AaMYB1 and its orthologue AtMYB61 affect terpene metabolism and trichome development in Artemisia annua and Arabidopsis thaliana. The Plant journal : for cell and molecular biology *90*, 520-534.

Matos, D.A., Whitney, I.P., Harrington, M.J., and Hazen, S.P. (2013). Cell Walls and the Developmental Anatomy of the Brachypodium distachyon Stem Internode. Plos One *8*.

McCarthy, R.L., Zhong, R., and Ye, Z.H. (2009). MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in Arabidopsis. Plant Cell Physiol *50*, 1950-1964.
McCarthy, R.L., Zhong, R., and Ye, Z.H. (2011). Secondary wall NAC binding element (SNBE), a key cis-acting element required for target gene activation by secondary wall NAC master switches. Plant Signal Behav *6*, 1282-1285.

Mitsuda, N., Iwase, A., Yamamoto, H., Yoshida, M., Seki, M., Shinozaki, K., and Ohme-Takagi, M. (2007). NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of Arabidopsis. Plant Cell *19*, 270-280.

Mitsuda, N., Seki, M., Shinozaki, K., and Ohme-Takagi, M. (2005). The NAC transcription factors NST1 and NST2 of Arabidopsis regulate secondary wall thickenings and are required for anther dehiscence. Plant Cell *17*, 2993-3006.

Nakano, Y., Yamaguchiz, M., Endo, H., Rejab, N.A., and Ohtani, M. (2015). NAC-MYB-based transcriptional regulation of secondary cell wall biosynthesis in land plants. Frontiers in plant science *6*.

Newman, L.J., Perazza, D.E., Juda, L., and Campbell, M.M. (2004). Involvement of the R2R3-MYB, AtMYB61, in the ectopic lignification and dark-photomorphogenic components of the det3 mutant phenotype. The Plant journal : for cell and molecular biology *37*, 239-250.

Nicol, F., and Hofte, H. (1998). Plant cell expansion: scaling the wall. Current opinion in plant biology *1*, 12-17.

Nishitani, K., and Demura, T. (2015). Editorial: an emerging view of plant cell walls as an apoplastic intelligent system. Plant Cell Physiol *56*, 177-179.

Noda, S., Koshiba, T., Hattori, T., Yamaguchi, M., Suzuki, S., and Umezawa, T. (2015). The expression of a rice secondary wall-specific cellulose synthase gene, OsCesA7, is directly regulated by a rice transcription factor, OsMYB58/63. Planta.

O'Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell *166*, 1598.

Olins, J.R., Lin, L., Lee, S.J., Trabucco, G.M., MacKinnon, K.J., and Hazen, S.P. (2018). Secondary Wall Regulating NACs Differentially Bind at the Promoter at a CELLULOSE SYNTHASE A4 Cis-eQTL. Frontiers in plant science *9*, 1895.

Olsen, A.N., Ernst, H.A., Leggio, L.L., and Skriver, K. (2005). NAC transcription factors: structurally distinct, functionally diverse. Trends in plant science *10*, 79-87.

Patzlaff, A., McInnis, S., Courtenay, A., Surman, C., Newman, L.J., Smith, C., Bevan, M.W., Mansfield, S., Whetten, R.W., Sederoff, R.R*., et al.* (2003a). Characterisation of a pine MYB that regulates lignification. The Plant journal : for cell and molecular biology *36*, 743-754.
Patzlaff, A., Newman, L.J., Dubos, C., Whetten, R.W., Smith, C., McInnis, S., Bevan, M.W., Sederoff, R.R., and Campbell, M.M. (2003b). Characterisation of Pt MYB1, an R2R3-MYB from pine xylem. Plant molecular biology *53*, 597-608.

Pepke, S., Wold, B., and Mortazavi, A. (2009). Computation for ChIP-seq and RNA-seq studies. Nat Methods *6*, S22-32.

Plomion, C., Leprovost, G., and Stokes, A. (2001). Wood formation in trees. Plant Physiol *127*, 1513-1523.

Prouse, M.B., and Campbell, M.M. (2012). The interaction between MYB proteins and their target DNA binding sites. Bba-Gene Regul Mech *1819*, 67-77.

Prouse, M.B., and Campbell, M.M. (2013). Interactions between the R2R3-MYB transcription factor, AtMYB61, and target DNA binding sites. Plos One *8*, e65132.

Pruneda-Paz, J.L., Breton, G., Nagel, D.H., Kang, S.E., Bonaldi, K., Doherty, C.J., Ravelo, S., Galli, M., Ecker, J.R., and Kay, S.A. (2014). A genome-scale resource for the functional characterization of Arabidopsis transcription factors. Cell reports *8*, 622-632.

Pyo, H., Demura, T., and Fukuda, H. (2007). TERE; a novel cis-element responsible for a coordinated expression of genes related to programmed cell death and secondary wall formation during differentiation of tracheary elements. The Plant journal : for cell and molecular biology *51*, 955-965.

Raes, J., Rohde, A., Christensen, J.H., Van de Peer, Y., and Boerjan, W. (2003). Genome-wide characterization of the lignification toolbox in Arabidopsis. Plant Physiol *133*, 1051-1071.

Rao, X., Chen, X., Shen, H., Ma, Q., Li, G., Tang, Y., Pena, M., York, W., Frazier, T.P., Lenaghan, S*., et al.* (2019). Gene regulatory networks for lignin biosynthesis in switchgrass (Panicum virgatum). Plant biotechnology journal *17*, 580-593.

Rao, X., and Dixon, R.A. (2018). Current Models for Transcriptional Regulation of Secondary Cell Wall Biosynthesis in Grasses. Frontiers in plant science *9*, 399.

Ream, T.S., Woods, D.P., Schwartz, C.J., Sanabria, C.P., Mahoy, J.A., Walters, E.M., Kaeppler, H.F., and Amasino, R.M. (2014). Interaction of Photoperiod and Vernalization Determines Flowering Time of Brachypodium distachyon. Plant Physiol *164*, 694-709.

Robinson, O., Dylus, D., and Dessimoz, C. (2016). Phylo.io: Interactive Viewing and Comparison of Large Phylogenetic Trees on the Web. Mol Biol Evol *33*, 2163-2166.

Romano, J.M., Dubos, C., Prouse, M.B., Wilkins, O., Hong, H., Poole, M., Kang, K.Y., Li, E., Douglas, C.J., Western, T.L*., et al.* (2012). AtMYB61, an R2R3-MYB transcription factor, functions as a pleiotropic regulator via a small gene network. New Phytol *195*, 774-786.

Scheller, H.V., and Ulvskov, P. (2010). Hemicelluloses. Annual review of plant biology *61*, 263-289.

Shen, H., He, X.Z., Poovaiah, C.R., Wuddineh, W.A., Ma, J.Y., Mann, D.G.J., Wang, H.Z., Jackson, L., Tang, Y.H., Stewart, C.N*., et al.* (2012). Functional characterization of the switchgrass (Panicum virgatum) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks. New Phytol *193*, 121-136.

Sibout, R., Proost, S., Hansen, B.O., Vaid, N., Giorgi, F.M., Ho-Yue-Kuang, S., Legee, F., Cezart, L., Bouchabke-Coussa, O., Soulhat, C*., et al.* (2017). Expression atlas and comparative coexpression network analyses reveal important genes involved in the formation of lignified cell wall in Brachypodium distachyon. New Phytol *215*, 1009-1025.

Smita, S., Katiyar, A., Chinnusamy, V., Pandey, D.M., and Bansal, K.C. (2015). Transcriptional Regulatory Network Analysis of MYB Transcription Factor Family Genes in Rice. Frontiers in plant science *6*, 1157.

Spencer, V.A., Sun, J.M., Li, L., and Davie, J.R. (2003). Chromatin immunoprecipitation: a tool for studying histone acetylation and transcription factor binding. Methods *31*, 67-75.

Stracke, R., Werber, M., and Weisshaar, B. (2001). The R2R3-MYB gene family in Arabidopsis thaliana. Current opinion in plant biology *4*, 447-456.

Tian, M., Xia, Q.M., and Li, J.Y. (2007). [The secondary growth in plant and its molecular regulation]. Yi Chuan *29*, 1324-1330.

Trabucco, G.M., Matos, D.A., Lee, S.J., Saathoff, A.J., Priest, H.D., Mockler, T.C., Sarath, G., and Hazen, S.P. (2013). Functional characterization of cinnamyl alcohol dehydrogenase and caffeic acid O-methyltransferase in Brachypodium distachyon. Bmc Biotechnol *13*.

Valdivia, E.R., Herrera, M.T., Gianzo, C., Fidalgo, J., Revilla, G., Zarra, I., and Sampedro, J. (2013). Regulation of secondary wall synthesis and cell death by NAC transcription factors in the monocot Brachypodium distachyon. J Exp Bot *64*, 1333-1343.

Vogel, J. (2008). Unique aspects of the grass cell wall. Current opinion in plant biology *11*, 301-307.

Vogel, J., Garvin, D., Leong, O., and Hayden, D. (2006). *Agrobacterium*-mediated transformation and inbred line development in the model grass *Brachypodium distachyon*. Plant Cell Tiss Org *84*, 100179-100191.
Vogel, J., and Hill, T. (2008). High-efficiency Agrobacterium-mediated transformation of Brachypodium distachyon inbred line Bd21-3. Plant cell reports *27*, 471-478.

Walhout, A.J., and Vidal, M. (2001). High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. Methods *24*, 297-306.

Wang, H.H., Tang, R.J., Liu, H., Chen, H.Y., Liu, J.Y., Jiang, X.N., and Zhang, H.X. (2013). Chimeric repressor of PtSND2 severely affects wood formation in transgenic Populus. Tree physiology *33*, 878-886.

Wolf, S., Hematy, K., and Hofte, H. (2012). Growth control and cell wall signaling in plants. Annual review of plant biology *63*, 381-407.

Yang, L., Zhao, X., Yang, F., Fan, D., Jiang, Y., and Luo, K. (2016). PtrWRKY19, a novel WRKY transcription factor, contributes to the regulation of pith secondary wall formation in Populus trichocarpa. Sci Rep *6*, 18643.

Yanhui, C., Xiaoyuan, Y., Kun, H., Meihua, L., Jigang, L., Zhaofeng, G., Zhiqiang, L., Yunfei, Z., Xiaoxiao, W., Xiaoming, Q.*, et al.* (2006). The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. Plant molecular biology *60*, 107-124.

Yao, D.X., Wei, Q., Xu, W.Y., Syrenne, R.D., Yuan, J.S., and Su, Z. (2012). Comparative genomic analysis of NAC transcriptional factors to dissect the regulatory mechanisms for cell wall biosynthesis. Bmc Bioinformatics *13*.

Ye, Y., Wu, K., Chen, J., Liu, Q., Wu, Y., Liu, B., and Fu, X. (2018). OsSND2, a NAC family transcription factor, is involved in secondary cell wall biosynthesis through regulating MYBs expression in rice. Rice (N Y) *11*, 36.

Zhang, J., Xie, M., Tuskan, G.A., Muchero, W., and Chen, J.G. (2018a). Recent Advances in the Transcriptional Regulation of Secondary Cell Wall Biosynthesis in the Woody Plants. Frontiers in plant science *9*, 1535.

Zhang, L.M., Leng, C.Y., Luo, H., Wu, X.Y., Liu, Z.Q., Zhang, Y.M., Zhang, H., Xia, Y., Shang, L., Liu, C.M.*, et al.* (2018b). Sweet Sorghum Originated through Selection of Dry, a Plant-Specific NAC Transcription Factor Gene. Plant Cell *30*, 2286-2307.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Zhao, K., and Bartley, L.E. (2014). Comparative genomic analysis of the R2R3 MYB secondary cell wall regulators of Arabidopsis, poplar, rice, maize, and switchgrass. BMC plant biology *14*, 135.
Zhong, R., Cui, D., and Ye, Z.H. (2018). Secondary cell wall biosynthesis. New Phytol.

Zhong, R., Demura, T., and Ye, Z.H. (2006). SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of Arabidopsis. Plant Cell *18*, 3158-3170.

Zhong, R., Lee, C., McCarthy, R.L., Reeves, C.K., Jones, E.G., and Ye, Z.H. (2011). Transcriptional activation of secondary wall biosynthesis by rice and maize NAC and MYB transcription factors. Plant Cell Physiol *52*, 1856-1871.

Zhong, R., Lee, C., and Ye, Z.H. (2010). Global analysis of direct targets of secondary wall NAC master switches in Arabidopsis. Molecular plant *3*, 1087-1103.

Zhong, R., Lee, C., Zhou, J., McCarthy, R.L., and Ye, Z.H. (2008). A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in Arabidopsis. Plant Cell *20*, 2763-2782.

Zhong, R., McCarthy, R.L., Haghighat, M., and Ye, Z.H. (2013). The poplar MYB master switches bind to the SMRE site and activate the secondary wall biosynthetic program during wood formation. Plos One *8*, e69219.

Zhong, R., Richardson, E.A., and Ye, Z.H. (2007a). The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in Arabidopsis. Plant Cell *19*, 2776-2792.

Zhong, R., Richardson, E.A., and Ye, Z.H. (2007b). Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of Arabidopsis. Planta *225*, 1603-1611.

Zhong, R., and Ye, Z.H. (2007). Regulation of cell wall biosynthesis. Current opinion in plant biology *10*, 564-572.

Zhong, R., and Ye, Z.H. (2012). MYB46 and MYB83 bind to the SMRE sites and directly activate a suite of transcription factors and secondary wall biosynthetic genes. Plant Cell Physiol *53*, 368-380.

Zhong, R., and Ye, Z.H. (2014). Complexity of the transcriptional network controlling secondary wall biosynthesis. Plant Sci *229*, 193-207.

Zhong, R., and Ye, Z.H. (2015). Secondary cell walls: biosynthesis, patterned deposition and transcriptional regulation. Plant Cell Physiol *56*, 195-214.

Zhong, R., Yuan, Y., Spiekerman, J.J., Guley, J.T., Egbosiuba, J.C., and Ye, Z.H. (2015). Functional Characterization of NAC and MYB Transcription Factors Involved in Regulation of Biomass Production in Switchgrass (Panicum virgatum). Plos One *10*, e0134611.
Zhou, J., Lee, C., Zhong, R., and Ye, Z.H. (2009). MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in Arabidopsis. Plant Cell *21*, 248-266.

Zhou, J., Zhong, R., and Ye, Z.H. (2014). Arabidopsis NAC domain proteins, VND1 to VND5, are transcriptional regulators of secondary wall biosynthesis in vessels. Plos One *9*, e105726.