Qualitative GUS and targeted yeast one-hybrid analyses of the *Eucalyptus grandis SND2* promoter region

by

Jonathan Botha

Submitted in partial fulfilment of the requirements for the degree

Magister Scientiae

In the Faculty of Natural and Agricultural Sciences

Department of Genetics

University of Pretoria

Pretoria

July 2013

Under the supervision of Prof. Alexander A. Myburg and

co-supervision of Dr. Christine Maritz-Olivier and

Mr Eshchar Mizrachi

DECLARATION

I, the undersigned, hereby declare that the dissertation submitted herewith for the degree M.Sc. to the University of Pretoria, contains my own independent work and has not been submitted for any degree at any other university.

_____.

Jonathan Botha

August 2013

PREFACE

The secondary cell walls of woody plant species are a major source of cellulose. This biopolymer is among the most abundant on the planet and is very useful to industry, having many applications, such as in the paper industry and as a possible source of fermentable sugar for biofuel production. Therefore, much emphasis is placed on understanding the formation of woody biomass (mainly comprising of secondary cell walls) as a possible feedstock for cellulose extraction. Wood fibres are of particular interest as they are most useful for the paper industry and have thick, cellulose rich secondary cell walls. Transcriptional regulation of secondary cell wall biosynthesis is facilitated through a complex network of proteins and genes operating in a semi-hierarchical manner. In xylem, each cell type is controlled by a "master switch", a transcription factor (TF) which is able to activate the entire secondary cell wall biosynthetic program for that particular cell type. In *Arabidopsis* xylem fibres, SECONDARY WALL ASSOCIATED NAC DOMAIN PROTEIN1 (SND1) is the master regulator for secondary cell wall deposition. Overexpression of SND1 results in the upregulation of a number TFs important for secondary cell wall biosynthesis. Three of the upregulated TFs were able to induce expression of a reporter gene under the control of a cellulose synthase (*AtCesA8*) promoter, indicating these TFs may be involved in cellulose biosynthesis. Of these, only SECONDARY WALL ASSOCIATED NAC DOMAIN PROTEIN2 (SND2) is not a direct target of SND1. While much work has been done on the regulatory network in the model plant *Arabidopsis*, there are still major gaps in our understanding. Comparatively little work has been done to study this network in hardwood crop tree species such as *Eucalyptus grandis*. Properly characterising and understanding this network will allow us to manipulate hardwood crops to obtain trees with commercially desirable traits and will also provide a platform for future studies in the field of transcriptional regulation of wood formation.

The aim of this MSc is to determine the expression pattern of the 1.5 kb *EgrSND2* promoter fragment in *Arabidopsis* through qualitative GUS analysis and to determine which TFs

important for secondary cell wall biosynthesis bind directly to this promoter fragment *in vivo,* through the use of Yeast One-Hybrid analyses. The promoter fragment was obtained by amplifying and cloning a 1.5 kb region directly upstream of the start codon of an ortholog of *SND2* that was identified in *Eucalyptus grandis.* This will allow us to not only make progress in understanding the transcriptional network of secondary cell wall biosynthesis in trees, but also elucidate new connections and functional relationships in the *Arabidopsis* network.

Chapter 1 is a review of the recent literature in regards to transcriptional regulation, promoters and wood development. In this chapter, the development of xylem (xylogenesis) is discussed as well as the transcriptional regulation thereof. The focus is on the transcription factors thought to be important for secondary cell wall biosynthesis. The two main classes of transcription factors known to be involved in this network are also reviewed, namely the MYB and NAC TF families. The basic structure of plant promoters is discussed, as well as certain core *cis*-regulatory elements. The *in silico* identification and annotation of *cis*-elements and experimental identification and verification are also discussed. Lastly, some of the more complex layers of eukaryotic transcriptional regulation are discussed, such as chromatin conformation and epigenetic modifications.

Chapter 2 discusses the results of the qualitative GUS and Yeast One-Hybrid analyses. In this chapter, the *Arabidopsis* tissues in which the 1.5 kb *EgrSND2* promoter fragment is active at various ages are identified. Direct interactions between the 1.5 kb *EgrSND2* promoter fragment and TFs important for secondary cell wall biosynthesis are also indentified *in vivo.* Based on the main findings of the chapter, the role and placement of *EgrSND2* in the transcriptional network controlling secondary cell wall biosynthesis in *E. grandis* are discussed.

Concluding remarks are included at the end of the dissertation. In this section, the results are put into context of the published literature, and are discussed with regards to their value

to the academic field and to industry. Possible improvements and current shortcomings in the field are also discussed.

ACKNOWLEDGEMENTS

I would like to acknowledge the following people, organisations and institutes for the assistance rendered to me in the completion of this study:

- To Prof. Zander Myburg for his support, insight and guidance during this study; for the opportunities and exposure to the scientific community he has given me and for setting an excellent example as a top researcher.
- Dr. Christine Maritz-Olivier for her exceptional insight, advice, perspective and review of this study. Though often quite different to that of Prof. Myburg, your input was highly valued.
- Eshchar Mizrachi for keeping me on track, reviewing the first drafts (Ain't nobody got time fo dat) and offering in equal measure insight, stern words and encouragement when needed, all while completing a degree of his own.
- Nicky Creux, I cannot put into words how much you have helped me in this degree. Through long working hours, dying plants and yeast and the ire of supervisors, your mentorship, advice and friendship have been invaluable. I believe in the future you will be an outstanding researcher (though you could be considered one already) and an even better supervisor. Also, you instituted lab wine.
- Minique de Castro for giving me promoters with cut sites and for always being so warm and enthusiastic about everything you do, whether that be advising an honours student on how to clone and sequence properly or meeting someone for a *wyntjie*.
- Steven Hussey for providing me with coding sequences and his valuable insight and critical review of my writing. Thank you for patiently viewing my scratchings and always offering harsh, yet constructive criticism, even when my points were somewhat... nebulous.

- All the staff and students of FMG, in particular, Lab Z. Nicky, Steve, Rits and Des, it was awesome to share a space with you guys. I can only hope that wherever I go I can find a group as great as you lot.
- Annelize Pienaar, my mentorship student. Thanks for all the sterile glucose, help when I needed extra hands and for not giving up on THAT PCR. I'm sure you will do excellently in all your plans for the future.
- The Department of Genetics and Forestry and Agricultural Biotechnology Institute at the University of Pretoria for providing solid infrastructure and top of the range facilities, and for giving me a proper introduction to real science.
- National Research Foundation (NRF) for funding supplied by the Scarce Skills scholarship.
- The Human Resources and Technology for Industry Programme (THRIP), Sappi and Mondi for providing plant material used in these studies and funding contributions to the project
- My friends, particularly Stew, Terence and Marcio, for all the celebrations and commiserations and for reminding me that there is a life outside of thesis.
- Caryn for coffee and procrastination.
- My Family. My mother for all her financial and emotional support, as well as for being a great example for an aspiring researcher. Without your tolerance, motivation and understanding, I would not have been able to complete this degree. Also, my brother Justin, my sister in-law Vanessa and my little nephew and niece, Zander and Talise. Thanks for all for your genuine interest and curiosity, for listening even when you had no idea what I was talking about and for showing me that there is more to live for than work.
- Last and never least, to Nora. You alone got to see the cracks, and always helped me shore them back up. I am eternally grateful for your love and support.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

CHAPTER 2

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

LIST OF APPENDICES

APPENDIX 1: SUPPLEMENTARY TABLES AND FIGURES

EgrKNAT7

EgrMYB46

EgrMYB52/54-A

EgrMYB52/54-B

EgrMYB83

EgrMYB85

EgrNST1

EgrSND1

EgrVND6

EgrVND7

EgrZF1

EgrZF2

APPENDIX 3: UNFORMATTED YEAST PICTURES (Please see attached CD)

Figure 1: Yeast one-hybrid analyses of the 1.5 kb *EgrSND2* promoter fragment.

Figure 2: Yeast one-hybrid analyses of the 500 bp *EgrSND2* promoter fragment.

Figure 3: Repeated Yeast one-hybrid analyses for both the 1.5 kb and 500 bp *EgrSND2* promoter fragments.

Figure 4: Yeast one-hybrid analyses of EgrMYB52/54-A and EgrKNAT7 cotransformants against the 500 bp *EgrSND2* promoter fragment.

APPENDIX 4: VECTOR MAPS (Please see attached CD)

pMDC162

pCR8/GW/TOPO

pHIS2.1

pDEST-GADT7

CHAPTER 1

LITERATURE REVIEW

PROMOTER-PROTEIN INTERACTIONS OF SECONDARY CELL WALL BIOSYNTHESIS IN PLANTS

1.1 Introduction

Vascular plants first evolved approximately 700 million years ago (Mya; Heckman et al. 2001; Hedges et al. 2004). When they made the transition to land about 400 Mya (Kenrick and Crane 1997; Proost et al. 2011) an efficient system of transport was needed to replace water lost through transpiration (Sperry 2003). The incidence of xylem (wood) allowed for the transport of water to the upper parts of the plant while resisting the high negative pressures associated with the cohesion-tension action of this kind of transport, and also lent structural support to the plant (Lucas et al. 2013). Woody plants have many advantages. For example, wood is dead and requires no maintenance and therefore it is a biologically economical form of water transport (Sperry 2003). Due to an improved tolerance for structural stress in woody plants, higher levels of transpiration and CO_2 diffusion are achieved (Sperry 2003). Lastly, the increase in structural support leads to larger plants, which means greater access to light and water supplies through tall shoots and large root systems, respectively (Raven and Edwards 2001; Lucas et al. 2013).

Wood consists of mature xylem and lends structure and support to large, mostly perennial plants. Primary xylem and phloem are derived from the procambium, one of the primary meristems that develop from the shoot apical meristem (SAM; Schuetz et al. 2013). Secondary growth is generally seen in gymnosperms and dicotyledonous angiosperms, but not in monocots (Spicer and Groover 2010). The primary lateral meristem which gives rise to secondary vascular tissues in plant stems and roots is known as the vascular cambium and is derived from the procambium (Schuetz et al. 2013). Secondary xylem is the main component of wood and develops from cambial initial cells (also known as xylem mother cells) in the vascular cambium. There are two types of cambial initial cells, namely the fusiform and ray initial cells. Fusiform initials give rise to vertically orientated cells such as the xylem vessels, while the ray initials give rise to cells used for lateral transport in the stem, such as ray parenchyma (Lucas et al. 2013). The fusiform initials give rise to xylem

cells which subsequently differentiate into a mature tissue that is called wood when present in the stems of large perennials. The derived cells expand, and while this is happening the primary cell wall further expands. After this the secondary cell wall (SCW) is deposited on the inside of the primary cell wall. The cell then undergoes programmed cell death and the SCW is left behind as the main structural and chemical component of wood (Plomion et al. 2001; Dejardin et al. 2010). The compound middle lamella (CML), which consists of the middle lamella and primary cell wall, surrounds xylem cells. The SCW consists of three layers, S1-S3, (Timell 1986) with the S1 layer immediately inside of the primary cell wall and the S3 layer adjacent to the plasma membrane (Plomion et al. 2001; Dejardin et al. 2010). In xylem cells it is the SCW that is usually thickened and enriched with cellulose and lignin.

The differentiation of cambial cells to xylem cells is controlled by the interactions of the products of the *REVOLUTA/INTERFASCICULAR FIBRELESS 1* (*REV/IFL1*) (Zhong et al. 1997; Emery et al. 2003; Prigge et al. 2005), *CORONA/ATHB15* (*CNA/ATHB15*), *PHABULOSA/ATHB14* (*PHB/ATHB14*), *PHAVOLUTA/ATHB9* (*PHV/ATHB9*) and *ATHB8* genes (Prigge et al. 2005), which have been shown to be responsible for embryo patterning in *Arabidopsis thaliana*. Further studies in *Arabidopsis* have demonstrated that the differentiation of xylem mother cells to xylem fibres is dependent on the interactions between the NAC SECONDARY CELL WALL THICKENING PROMOTING FACTOR 1 (NST1) and SECONDARY WALL ASSOCIATED NAC DOMAIN PROTEIN1 (SND1)/NST3 proteins, which act as master regulatory switches for SCW biosynthesis (Mitsuda et al. 2007). They directly and indirectly affect the promoters of the genes coding for SND2, SND3, MYB20, MYB42, MYB43 MYB52, MYB54, MYB85, MYB103 and KNAT7 transcription factors (TFs) and upregulate their expression (Zhong et al. 2008). These TFs appear to play a prominent role in SCW biosynthesis and SND2, SND3 and MYB103, in particular, are proposed to play a role in regulating cellulose biosynthesis (Zhong et al. 2008).

Cellulose is synthesised by a complex of cellulose synthase (CesA) proteins in the cell membrane. The *CesA4, -7* and *-8* genes are responsible for cellulose synthesis in SCWs of

Arabidopsis (Taylor et al. 2003). The CesA proteins can be found in the cell membrane on the cytoplasmic side in a structure known as a rosette (Kimura et al. 1999). Cellulose biosynthesis, the regulation thereof, CesA proteins and the structure of cellulose itself have been studied in detail (Somerville 2006; Taylor 2008), though knowledge and understanding of this area is far from complete (Guerriero et al. 2010). In terms of regulation, there are still many protein-protein and protein-DNA interactions that need to be characterised in order to complete the picture. In the transcriptional network of SCW biosynthesis in *Arabidopsis*, the interactions between the *CesA* genes, the *CesA*-associated SND2, SND3 and MYB103 TFs and the other TFs in the network are still not clear.

The TFs in the transcriptional network of SCW biosynthesis (a cascade of interacting TFs that govern the transcription of SCW related biosynthetic genes) mainly belong to two protein families, the NAC (e.g. NST1, NST3/SND1, SND2 and SND3; Zhong et al. 2006; Zhong et al. 2007b; Zhong et al. 2008) and MYB (e.g. MYB46, MYB83 and MYB103; Zhong et al. 2007a; Zhong et al. 2008; Zhong and Ye 2012) families. NAC TFs play a role in many different biological processes in the plant (Olsen et al. 2005), while the role of MYB proteins are normally the regulation of developmental processes (Stracke et al. 2001; Yanhui et al. 2006). Some of these proteins mediate the transcriptional regulation of SCW biosynthesis through various regulatory interactions.

Though regulatory interactions are often discussed in terms of single proteins recognising and binding to a specific *cis*-element, in truth, it is far more complex. Regulation not only takes place on the level of protein-DNA interactions, but also on the levels of protein-protein interactions (cooperativity), histone modifications and chromatin conformation and long-range interaction (Lelli et al. 2012). These are just some of the many layers of transcriptional control. The Encyclopaedia of DNA Elements (ENCODE) project seeks to assign functions to all non-coding regions of the genome (Feingold et al. 2004; Dunham et al. 2012), and has also highlighted how chromatin conformation and epigenetic modifications can affect gene expression (Thurman et al. 2012).

Eucalyptus spp are commercially important trees. They have become a large fibre crop of tropic and subtropic areas because of their ability to provide pulp, paper and other wood by-products that are not at the expense of the local tree species (Grattapaglia and Kirst 2008; Grattapaglia et al. 2009). In this literature review, the formation of xylem as well as the genetic regulation thereof will be investigated. The transcriptional regulation of SCW biosynthesis in particular will be discussed, as well as the structure and biosynthesis of cellulose. The TFs involved in the regulation of these processes (NAC TFs and MYB TFs) will be investigated. Promoters and the *cis*-regulatory elements will be reviewed, as well as experimental and computational techniques used to identify these elements. Lastly, some of the more complex mechanisms of transcriptional regulatory interactions will be discussed. All genes and proteins discussed in this review are from *Arabidopsis thaliana*, unless specifically stated otherwise.

1.2 Wood formation

Wood formation begins in the apical meristem of the plant. The shoot apical meristem (SAM) gives rise to procambium (Ye 2002). From the procambium, primary xylem, primary phloem and vascular cambial initial cells develop. The vascular cambial initials, along with the interfascicular cambial initial cells, form the vascular cambium, also known as the secondary meristem (Figure 1; Ye 2002). The vascular cambium is a lateral meristem which facilitates growth of a plant laterally, as opposed to the apical meristem which is responsible for vertical growth. The vascular cambium eventually gives rise to xylem or phloem mother cells. It is from the mother cells that the vascular tissues — xylem and phloem (Figure 1) — are finally derived (Plomion et al. 2001; Dejardin et al. 2010; Schuetz et al. 2013).

Xylem is extremely important for the proper functioning of a plant. It provides a channel to transport water from the roots to the upper parts of the plant where it can be used for various physiological processes. Xylem also lends structural support to the plant. A number of cell types can be found in xylem. These include the tracheary elements, such as protoxylem and

metaxylem vessels, which transport water through the plant; xylem fibres, which provide structural support; and xylem parenchyma cells, which provide various compounds and polymers necessary for the development of tracheary elements and fibres, as well as a conduit of transport between the cells (Figure 1; Schuetz et al. 2013). Through the process of programmed cell death (PCD) the xylem vessels themselves are hollow and dead which allows them to transport water efficiently. Protoxylem vessels are formed relatively early in xylem development: before elongation has occurred in the surrounding tissues (Esau 1960). Protoxylem vessels have spiral and annular thickenings of their cell walls, and are generally destroyed during elongation of the surrounding tissues (Esau 1960). Metaxylem vessels, however, mature after the surrounding tissues have elongated (Esau 1960). They are not destroyed, and therefore become the primary conduits for water transport in the plant. Their cell walls are characterised by pitted and reticulate thickenings (Esau 1960). The majority of wood is composed of xylem fibres which mainly provide structural support; this serves to highlight their importance in the plant.

A xylem cell must undergo five main steps in order to become a mature cell: namely cell division, cell expansion, cell wall thickening, PCD and lignification (Plomion et al. 2001; Dejardin et al. 2010). The first step is cell division: the cell is derived from the cambium, which divides to give rise to a xylem mother cell. In this step the primary cell wall is formed. The next step is elongation, where the cell and primary cell wall elongate, resulting a full sized xylem cell. Following elongation is cell wall thickening, during which the SCW is deposited. The next steps are lignification and PCD, during which the cell is waterproofed and made hollow so that it may effectively perform its function of water transportation. During cell wall thickening, cellulose, hemicelluloses and lignin are deposited (Mellerowicz et al. 2001; Plomion et al. 2001; Mellerowicz and Sundberg 2008; Dejardin et al. 2010; Schuetz et al. 2013). There is a complex network of genetic regulation in the xylem cells which ensures the unique composition of the SCW in terms of cellulose, hemicellulose and lignin

(Mellerowicz and Sundberg 2008; Dejardin et al. 2010; Du and Groover 2010; Schuetz et al. 2013).

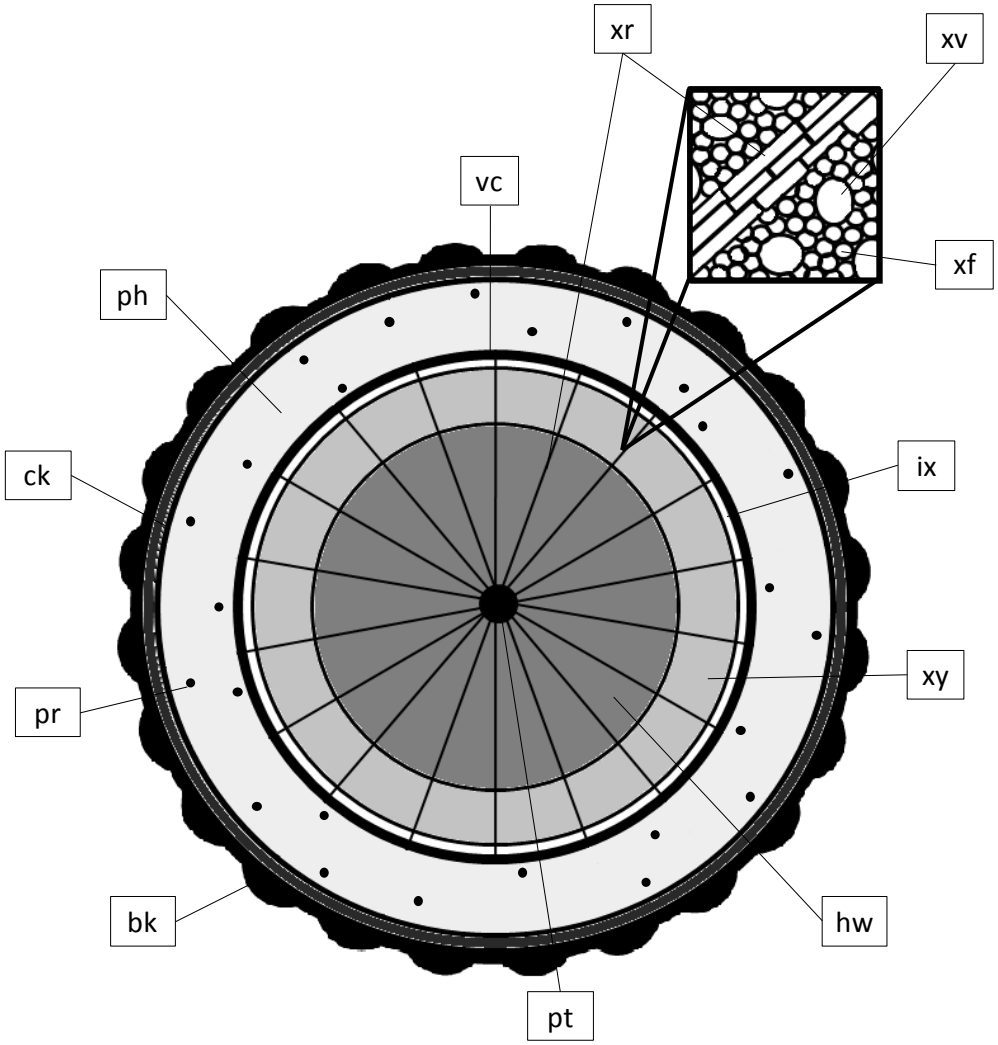


Figure 1: Schematic representation of a cross section through a typical woody tree stem. vc = vascular cambium, xr = xylem ray, xv = xylem vessel, xf = xylem fibre, ix = immature xylem, xy = mature xylem, hw = heartwood, pt = pith, ph = phloem, ck = cork cambium, pr = phloem ray and bk = bark.

1.3 Transcriptional regulation of xylem development

The transcriptional regulation of xylem development is complex and includes hormonal, transcriptional and protein signalling cascades. For the purposes of this review the transcriptional regulation of xylem development from the cambial initial cells to the xylem fibre cells will be the focus. Each step is controlled by a set of TFs, which in turn regulate

genes which code for other TFs. This continues in a cascade of transcriptional control involving both direct and indirect interactions, which ultimately results in both spatial and temporal regulation of developmental steps important for wood development in the plant. Direct interactions are described as those where there is experimental evidence for an interaction between a TF and the promoter of a gene of interest, while indirect interactions are those where a particular TF may have an effect on gene expression, but no experimental evidence of direct interaction between the TF and promoter of the gene has been observed.

The procambium and vascular cambium cells give rise to vascular tissues. If an error were to occur in the development of these cells, it would be seen in the abnormal arrangement of the developing vascular tissues. Studies by Scarpella and Meijer (2004) identified a number of proteins that appeared to be involved in vascular development. Among these it was noted that when the *MONOPTEROS/AUXIN RESPONSE FACTOR 5 (MP/ARF5)* gene was knocked out, reduced, misaligned and discontinuous vascularisation occurred (Scarpella and Meijer 2004). This indicates that the *MP/ARF5* is responsible for maintenance of the procambium and cambial cells. It is hypothesised that the *BEDENLOS/INDOLACETIC ACID-INDUCED PROTEIN 12 (BDL/IAA12)* has a similar role, as mutation of this gene also led to abnormal vasculature in *Arabidopsis* plants (Hamann et al. 1999; Hamann et al. 2002). It has been suggested that these two proteins have a complex interaction, with *BDL/IAA12* suppressing the activity of *MP/ARF5* when no auxin signal is present and *BDL/IAA12* being degraded by *TRANSPORT INHIBITOR RESPONSE 1 (TIR1)* when an auxin signal is present (Badescu and Napier 2006). In this way, vasculature development is controlled by these two genes. Additionally *MP/ARF5* has a number of downstream targets, designated the *TARGET OF MONOPTEROS (TMO)* genes, which are thought to play a role in embryo patterning and tissue differentiation in *Arabidopsis* roots (Schlereth et al. 2010).

Xylem and phloem differentiation are also affected by a number of TFs. These factors fall into 2 main gene classes. The first is the class III homeodomain-leucine zipper (HD-ZIPIII) TFs which include *REVOLUTA/INTERFASCICULAR FIBRELESS 1 (REV/IFL1),*

PHABULOSA/ARABIDOPSIS THALIANA HOMEOBOX14 (PHB/ATBH14), PHAVOLUTA/ATHB9 (PHV/ATHB9; Prigge et al. 2005) and ATHB8 (Baima et al. 2001). The second class contains proteins of the GARP-type TF family (Riechmann et al. 2000); specifically three KANADI genes, *KAN1-KAN3* in this case (Eshed et al. 2001; Emery et al. 2003). It was found that loss of function mutations in the KAN genes and gain of function mutations in the HD-ZIPIII genes both had the effect of producing amphivasal vascular bundles; a phenomenon where xylem surrounds the phloem in the vascular bundles (McHale and Koning 2004; Zhong and Ye 2004). This indicates that the HD-ZIPIII genes positively regulate xylem and negatively regulate phloem formation, while the KAN genes negatively regulate xylem formation and positively regulate the development of phloem.

Detailed genetic analysis of the REV/IFL1, PHB/ATBH14, PHV/ATHB9, CNA/ATHB15 (Prigge et al. 2005) and ATHB8 (Baima et al. 2001) showed that these proteins play overlapping roles in regulating the size and shape of vascular bundles in developing plants. Since their roles overlap, it is hypothesised that they are partially redundant in function (Prigge et al. 2005). It was seen that defective fibre mutation, observed in the REV/IFL mutants, was partially suppressed in triple mutant lines containing defective *rev-6 corona, cna/athb15* and *athb8* genes, which indicates that these genes may function antagonistically to regulate the development of xylem and phloem (Prigge et al. 2005). Gain of function mutations in the HD-ZIPIII genes is thought to be a result of mutations in micro RNAs (miRNAs), specifically miRNA165 and miRNA166 (Mallory et al. 2004b). Mutation of these sequences leads to increased mRNA stability of the HD-ZIPIII genes and thus an overexpression of their products which, in turn, result in mutant phenotypes showing abnormal vasculature (Mallory et al. 2004b; Demura and Fukuda 2007). Additionally, the HD-ZIPIII genes have been implicated in proto- and metaxylem vessel specification in *Arabidopsis* roots (Carlsbecker et al. 2010).

Differentiation into xylem cells is controlled to a large degree by the NAM/ATAF/CUC (NAC) family of genes, which are often found to be active in developing wood and differentiating

tracheary elements and are thought to play a role in their development (Zhong et al. 2010b; Mitsuda et al. 2005; Mitsuda et al. 2007; Zhong et al. 2007b; Zhao et al. 2008; Kato et al. 2009; Morishita et al. 2009; Peng et al. 2009; Zhong et al. 2010a). A study conducted by Kubo et al. (2005) identified a range of genes that were highly expressed in developing vascular tissues by making use of a microarray analysis. Two NAC domain proteins were of particular interest, namely, VASCULAR-RELATED NAC-DOMAIN (VND) -6 and VND7. Dominant repression of these two proteins resulted in plants showing a reduced proto- and metaxylem phenotype (Kubo et al. 2005). Overexpression of the *VND6* and *VND7* genes under a the control of a *35SCaMV* promoter resulted in transdifferentiation of many different cell types into metaxylem-like and protoxylem-like cells, respectively (Yamaguchi et al. 2010). Fluorescence studies indicated that these proteins are mainly expressed in the nucleus, thus providing evidence that they are TFs (Kubo et al. 2005). From these studies it can be inferred that VND6 and VND7 are master switches for the development of metaxylem and protoxylem vessels respectively.

It is believed that the master switches for xylem fibre development are NAC SECONDARY WALL THICKENING PROMOTING FACTOR 1 (NST1) and NST3 (Mitsuda et al. 2007). It is noted that NST2 may also play a role in SCW development, but it appears to be highly expressed in anther cells only (Mitsuda et al. 2005). NST3 was also described by Zhong et al. (2006), but was designated as SND1, and will be referred to as such hereafter. Overexpression of NST1 resulted in the induction of ectopic SCWs (Mitsuda et al. 2005), as did overexpression of SND1 (Zhong et al. 2006; Mitsuda et al. 2007). Repression of NST1 and SND1 resulted in reduced SCWs, as these proteins appear to have overlapping, redundant functions (Mitsuda et al. 2005; Zhong et al. 2006; Mitsuda et al. 2007; Zhong et al. 2007b; Zhong et al. 2008). NST1 and SND1 appear to be very similar to VND6 and VND7, as they belong to a similar subfamily of genes (Kubo et al. 2005; Mitsuda et al. 2007), providing further evidence that NST1 and SND1 are master switches for SCW biosynthesis in xylem fibre cells.

It is important to note that while much research has been performed on the regulation of wood development, it is only a small piece of a complex picture. There are many additional TFs and interacting partners which play a role that still need to be identified and characterised. In summary, differentiation from the procambium to cambial cells is regulated by MP/ARF5 (Scarpella and Meijer 2004). Differentiation of the cambial cells to phloem mother cells is regulated by the KANADI proteins, KAN1-3 (Eshed et al. 2001; Emery et al. 2003), while differentiation to xylem mother cells is regulated by the HD-ZIPIII proteins CNA/ATB15, REV/IFL1, PHB/ATHB14, PHV/ATHB9 (Prigge et al. 2005) and ATHB8 (Baima et al. 2001). The differentiation of xylem mother cells to metaxylem vessels and protoxylem vessels are regulated by VND6 and VND7 (Kubo et al. 2005).

Differentiation to xylem fibres is controlled by NST1, NST2 and SND1 (Mitsuda et al. 2005; Zhong et al. 2006; Mitsuda et al. 2007; Zhong et al. 2007b; Zhong et al. 2008). NST1 and SND1 function in a partially redundant manner as master regulators of the transcriptional cascade of SCW biosynthesis in xylem fibres (Zhong et al. 2007b). Overexpression of these TFs in *Arabidopsis* results in the upregulation of many other TFs, as well as a thickening of the SCWs of xylem fibres. (Mitsuda et al. 2005; Zhong et al. 2006; Demura and Fukuda 2007; Mitsuda et al. 2007; Zhong et al. 2007b; Zhong et al. 2008). They interact either directly or indirectly with genes encoding other TFs involved specifically in SCW biosynthesis.

1.4 Transcription factors

1.4.1 NAC transcription factors

NAC (NAM, ATAF1,2, CUC2) TFs are one of the largest plant-specific families of TF (Riechmann et al. 2000; Ernst et al. 2004; Gutierrez et al. 2004; Olsen et al. 2005). The family is named after three of the earliest identified and best characterised NAC-domain proteins. These are NAM, ATAF1,2 and CUC2. NAM (*no apical meristem*), identified in

petunias, was the first described NAC (Souer et al. 1996). It was found that loss of function mutations in the *NAM* gene resulted in plants lacking a SAM (Souer et al. 1996). Soon after this, CUC2 (*cup-shaped cotyledon*) was identified as playing a role in cotyledon separation in *Arabidopsis* and was found to be homologous to NAM. In fact, double mutants of both the *CUC1* and *CUC2* genes resulted in a phenotype identical to that induced by mutant *nam* (Aida et al. 1997). Lastly, ATAF1 and -2 were identified in parallel with the CUC2 proteins as having a homologous domain. The homology between these proteins led to the identification and naming of the NAC protein domain (PF018490; Aida et al. 1997).

The NAC domain, characteristic of NAC TFs, is located in the N-terminal region of NAC TFs and is highly conserved (Aida et al. 1997; Ernst et al. 2004; Olsen et al. 2004). There is no such conservation in the C-terminal region, though it is believed that there are subfamily-specific motifs in this area. For example, the VND/NST/SND1 subfamily of NAC proteins have two highly conserved motifs, the LP-box and WQ-box, located in the C-terminal region (Ko et al. 2007). Also, a frequent feature of NAC proteins is amino acid repeats at the C-terminal region which are rich in proline and glutamine, serine and arginine or any acidic residues (Souer et al. 1996; Aida et al. 1997; Ooka et al. 2003; Ernst et al. 2004; Olsen et al. 2004). The NAC domain has DNA-binding activity and recognises the core DNA motif CACG, while the C-terminal region behaves as an activation domain. NAC proteins also form homo- and heterodimers, and it is thought that the NAC domain is the focus of dimerisation in the NAC family (Xie et al. 2000; Olsen et al. 2004), though there is evidence of the C-terminal region also having an effect on dimerisation (Hegedus et al. 2003). This dimerisation allows for a combinatorial specificity approach to regulating many different pathways and processes.

The functions of NAC proteins are diverse, playing major roles in many different biological processes including flower and embryo development, formation of lateral roots, auxin signalling pathways and biotic and abiotic stress among many other things (Xie et al. 2000; Hegedus et al. 2003; Mallory et al. 2004a; Morishita et al. 2009; Peng et al. 2009; Ohashi-Ito

et al. 2010), as well as regulation of cell differentiation and SCW deposition (Mitsuda et al. 2005; Zhong et al. 2006; Mitsuda et al. 2007; Ohashi-Ito et al. 2010; Yamaguchi et al. 2010; Zhong et al. 2010a; Hussey et al. 2011; Zhong et al. 2011). Olsen et al. (2005) has comprehensively reviewed their other functions in plants. Of particular interest to this study is the role of NAC TFs in the regulation of SCW biosynthesis in plants (Yamaguchi and Demura 2010). A number of TFs were upregulated in a microarray study of *Zinnia elegans* mesophyll cells transdifferentiating into tracheary elements (Demura et al. 2002). One of these, named Z567, had a putative NAC domain. This experiment was replicated in *Arabidopsis*, in which microarray analysis was performed on suspension cells transdifferentiating into xylem vessel elements (Kubo et al. 2005). From this study, four NAC-domain TFs were identified that were significantly similar to Z567. A further search of the *Arabidopsis* genome identified three more genes that were similar to Z567. These seven genes were designated *VASCULAR-RELATED NAC DOMAIN PROTEIN 1* (VND1) to VND7 (Kubo et al. 2005). Overexpression of *VND6* and *VND7* led to transdifferentiation into metaxylem vessel-like and protoxylem vessel-like cells, respectively, and dominant repression of *VND6* and *VND7* inhibited the formation of xylem vessels. However, T-DNA knock-outs and RNAi induced silencing of *VND6* and *VND7* individually had no effect on vessel formation (Kubo et al. 2005). This, along with a partial overlap of expression of VND6 and VND7 in certain tissues, suggests that these genes are partially redundant master regulators of SCW biosynthesis in metaxylem and protoxylem vessel cells, respectively (Yamaguchi et al. 2010).

NST1, NST2 and SND1/NST3/ANAC012 make up the rest of the VND/NST1/SND1 subfamily. NST1 and NST2 were first identified in *Arabidopsis*, when dominant repression and overexpression studies showed that these two proteins were involved in SCW biosynthesis and anther dehiscence in *Arabidopsis* (Mitsuda et al. 2005). The proteins were similar to the VND TFs in that only double knockouts of both genes resulted in a phenotype, indicating redundancy. However, unlike the VND TFs, NST1 and NST2 only seemed to

activate genes involved in SCW biosynthesis, and none involved in PCD (Mitsuda et al. 2005). Shortly after this, SND1 was identified as being highly expressed in the interfascicular fibres and xylary cells of *Arabidopsis* inflorescence stems, and was able to not only reduce the SCWs of fibre cells through dominant repression, but also induce heavy ectopic deposition of SCWs in normally non-schlerenchymatous cells through overexpression (Zhong et al. 2006). At almost the same time, SND1 was investigated in two other studies and was implicated in the biosynthesis of SCWs in interfascicular fibre cells. In these two studies, SND1 was designated as ANAC012 (Ko et al. 2007) and NST3 (Mitsuda et al. 2007). In this review the SND1 naming convention will be used. It is interesting to note that the study performed by Ko et al. (2007) stated that SND1 had repressor activity, while the other two stated that it was an activator. The repression seen by Ko et al. (2007) may have been due to the dramatic overexpression of the TF, which can sometimes result in a negative effect (Gill and Ptashne 1988).

It was shown that SND1 is able to upregulate the expression of a number of TFs involved in SCW biosynthesis in fibre cells (Zhong et al. 2008). This includes two additional NAC TFs, SND2 and SND3, which were able to activate expression of a reporter gene under the control of an *AtCesA8* promoter, indicating that they may be involved in cellulose biosynthesis. A number of other NAC TFs have been implicated in SCW biosynthesis in various tissues of *Arabidopsis.* These include SOMBRERO (SMB), BEARSKIN1 (BRN1) and BRN2, which are thought to mimic the functions of the VND/NST/SND1 family in the root cap (Bennett et al. 2010), and XYLEM NAC DOMAIN 1 (XND1), which has been identified as a transcriptional repressor in programmed cell death and SCW biosynthesis (Zhao et al. 2005a; Zhao et al. 2008). NAC TFs function in a transcriptional network with many other classes of TFs, for example, the MYB family of TFs.

1.4.2 MYB transcription factors

Another family of proteins essential to many processes in plants are MYB TFs. MYB TFs are found in all eukaryotes. They are characterised by the presence of the MYB DNA binding domain (PF00249) in the N-terminal region of the protein. In general, the MYB domain is defined by the presence of up to four imperfect amino acid sequence repeats of approximately 52 amino acids each, with each repeat forming three α-helices (Ogata et al. 1996; Du et al. 2009). They can be divided into four main classes based on the number of repeats present in their MYB domain, and these classes contain those proteins with one, two, three or four repeats respectively. The main three amino acid repeats are designated R1, R2 and R3 after their discovery in c-MYB (Paz-Ares et al. 1987), and all additional repeats are then named for their similarity to these repeats. If MYB TFs have two or more repeats, they tend to bind as monomers, with the multiple repeats behaving as covalently linked dimers (Ogata et al. 1995). If only one repeat is present, then the protein tends to bind as either homo- or heterodimers (Lu et al. 2009). Plants are the only organisms where all four MYB classes are found, indicating that they may have the highest diversity of MYB TFs (Stracke et al. 2001; Yanhui et al. 2006; Matus et al. 2008; Wilkins et al. 2009).

The first MYB identified was in avian myeloblastosis virus and was designated v-MYB (Paz-Ares et al. 1987). Soon after, c-MYB, A-MYB and B-MYB were discovered in many other organisms (Weston 1998). This led to the discovery of homologs of these proteins in other organisms such as insects, fungi, slime moulds (Borevitz et al. 2000) and eventually higher plants such as *Arabidopsis,* maize, rice, petunia, snapdragon, grapevine, poplar and apple (Dubos et al. 2010). The first MYB protein identified in higher plants was C1 (Paz-Ares et al. 1987), a protein important for anthocyanin biosynthesis in maize. Since then, many MYB TFs have been identified in plants, and have been shown to have roles in a diverse array of biological processes within them (Jin and Martin 1999; Du et al. 2009; Dubos et al. 2010; Czemmel et al. 2012; El-kereamy et al. 2012; Naz et al. 2013; Tominaga-Wada et al. 2013).

A number of MYB TFs have been shown to play a role in SCW biosynthesis, namely MYB20, MYB42, MYB43, MYB52, MYB54, MYB69, MYB85, MYB103 (Zhong et al. 2008), MYB46 (Zhong et al. 2007a; Ko et al. 2009) and MYB83 (McCarthy et al. 2009). Of these, MYB46, MYB52, MYB54, MYB83, MYB85 and MYB103 will be discussed later in this review. MYB20 is a direct target of SND1 (Ko et al. 2007; Zhong et al. 2008). It has also been shown to be expressed in the xylem vessels of *Arabidopsis* roots (Nakano et al. 2010). It is not regulated by MYB46 or MYB83, mid-level master regulators of SCW biosynthesis in xylem (McCarthy et al. 2009), indicating it may only have a role in very specific cell types. However, it has been shown to be downregulated under salt and drought stress conditions in *Arabidopsis* (Zeller et al. 2009). As of yet, very little is known about MYB42 and MYB43. They are both induced by SND1 (Zhong et al. 2008) and MYB83 (McCarthy et al. 2009). However, the promoter of *MYB43* contains an ACTYP motif (a possible target element for MYB46-like TFs in *P. trichocarpa*), an indicator that it is more likely involved in general xylem development, rather than in a specific pathway such as lignin biosythesis (Winzell et al. 2010). Lastly, MYB69, apart from being induced by SND1 (Zhong et al. 2008), is also hypothesised to have a role in the negative regulation of lignin biosynthesis (Shen et al. 2009). In addition, its ortholog in *Populus, PtrMYB26*, can activate the promoters of lignin biosynthetic genes (Zhong et al. 2011). MYB69 was also found to be downregulated in MYB103 mutant plant lines (Öhman et al. 2013). All this indicates a possible role in lignin biosynthesis for MYB69.

There are many different classes of TFs that play a role in plant growth and development. This is also true for SCW biosynthesis. The MYB and NAC families are two of the largest families in plants, fulfilling functions in a diverse array of mechanisms and processes. This means that a large portion of the SCW transcriptional network is regulated by NAC and MYB TFs. SND1, NST1, NST2 (Mitsuda et al. 2005; Mitsuda et al. 2007; Zhong et al. 2007a; Zhong et al. 2007b; Zhong et al. 2008), VND6 and VND7 (Ohashi-Ito et al. 2010; Yamaguchi et al. 2010; Yamaguchi et al. 2011) are master transcriptional switches of SCW biosynthesis

in *Arabidopsis*, and MYB20, MYB42, MYB43, MYB52, MYB54, MYB69, MYB85, MYB103 (Zhong et al. 2008), MYB46 (Zhong et al. 2007a; Ko et al. 2009) and MYB83 (McCarthy et al. 2009) are regulators in the network. It is important to study all of these TFs to understand the transcriptional network, but some of these TFs are worth studying as targets for modification in order to direct the SCW biosynthetic program.

1.5 Transcriptional regulation of SCW biosynthesis

The biosynthesis of plant SCWs is regulated in a cascade-like fashion (Zhong et al. 2006). In this review, the focus will be on TFs known to be most closely associated with the biosynthesis of SCWs of xylem fibre cells. A schematic diagram showing the interactions of these TFs was constructed (Figure 2). These include SND1 and NST1 which are the functionally redundant master regulators of SCW biosynthesis in xylem fibre cells (Mitsuda et al. 2007). When these TFs were downregulated using RNA interference (RNAi), eleven other TFs showed a decrease in expression (Zhong et al. 2008). Among these were SND2, SND3, MYB103, MYB85, MYB52, MYB54 and KNAT7 (Figure 2). Additionally, MYB46 and MYB83 have also been shown to be involved SCW biosynthesis (Zhong et al. 2007a; McCarthy et al. 2009; Kim et al. 2012; Zhong and Ye 2012). Two targets of MYB46, namely AT1G66810 (also known as AtC3H14) and AT1G72220, may also play a role in SCW biosynthesis (Ko et al. 2009). All the TFs discussed in this review appear to activate expression, with exception of KNAT7 (Li et al. 2012), and more recently MYB52 (Cassan-Wang et al. 2013), which show evidence of a repressive function.

MYB46 and MYB83 have been characterised as mid-level master switches of SCW biosynthesis. This mean that they are able to activate all pathways required for the SCW to be deposited (Zhong et al. 2007a; Zhong et al. 2008; McCarthy et al. 2009), but are situated below, bound directly by, and are subject to regulation from the top level regulators in the transcriptional cascade such as NST1, NST2, SND1, VND6 and VND7 (Zhong et al. 2007a; McCarthy et al. 2009; Figure 2). MYB46 and MYB83 are both expressed in interfascicular

fibres and xylem vessels of *Arabidopsis* (Zhong et al. 2007a; McCarthy et al. 2009). Double knockouts of MYB46 and MYB83 performed in *Arabidopsis* resulted in a loss of SCWs in xylem vessel cells, while single knockouts showed no phenotype (McCarthy et al. 2009). MYB46 has been shown to upregulate MYB85 and KNAT7 significantly (Figure 2), two TFs that have been associated with lignin biosynthesis (Zhong et al. 2008; Li et al. 2012), suggesting a role for MYB46 in lignin biosynthesis. However, MYB46 was still able to induce low-level expression of SND2, SND3 and MYB103 (Zhong et al. 2007a), three TFs able to activate the *CesA8* promoter when subjected to a β-glucuronidase (GUS) reporter gene analysis (Zhong et al. 2008), indicating it can affect the other biosynthetic pathways too. MYB83 is also able to induce expression of these TFs (McCarthy et al. 2009). In *Arabidopsis* roots, MYB46 and MYB83 were able to upregulate expression of MYB52, MYB85 and MYB103 (Figure 2), and were expressed in interfascicular fibres and xylem vessels (Nakano et al. 2010). The similarity in targets, lack of phenotype from single knockouts and expression patterns of MYB46 and MYB83 therefore suggests that they behave in a redundant manner (McCarthy et al. 2009).

MYB52 and MYB54 have been associated with the biosynthesis of lignin, cellulose and xylan (Zhong et al. 2008; Ko et al. 2009). They are indirect targets of SND1 and also of the intermediate master regulators MYB83 (McCarthy et al. 2009) and MYB46, as well as the potential mid-level master regulator TF ATC3H14 (Figure 2; Ko et al. 2009). While dominant repression of MYB52 and MYB54 led to slightly thinner SCW in xylem fibres, overexpression did not result in any significant thickening of SCWs (Zhong et al. 2008). MYB52 has an expression pattern almost exactly the same as that of the *4CL1* activating MYB85 (Zhong et al. 2008), suggesting that it may have a role in lignin biosynthesis (Nakano et al. 2010). Additionally, *Arabidopsis myb52* mutants were shown to be hyper-lignified, indicating that MYB52 may repress lignin biosynthesis (Cassan-Wang et al. 2013). Through bioinformatic studies, MYB54 has been implicated in repression of lignin biosynthesis (Shen et al. 2009). While both are highly expressed in the bottom of *Arabidopsis* inflorescence stems, MYB52 is

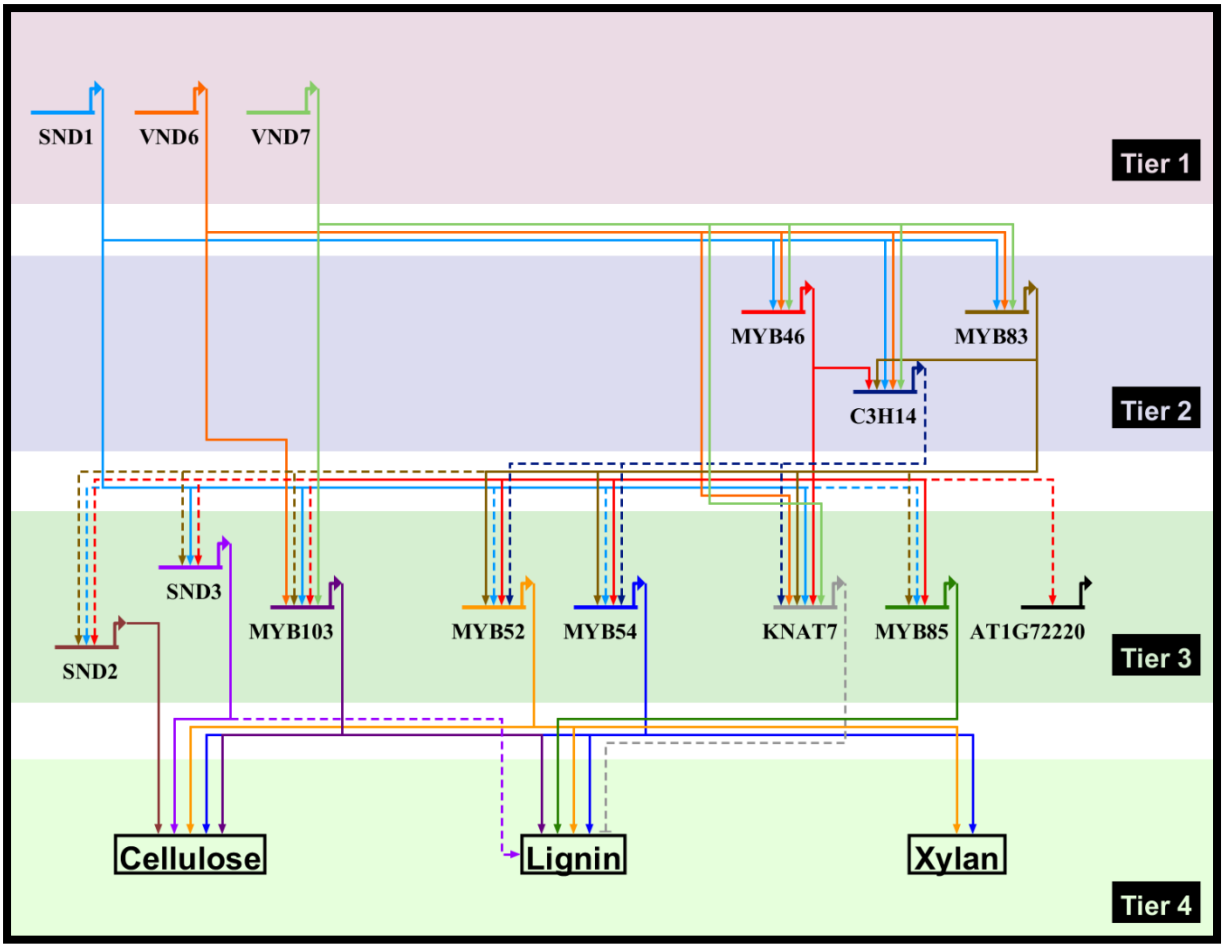


Figure 2: Simplified schematic representation of the transcriptional network of SCW biosynthesis drawn with Biotapestry (Longabaugh 2012) and adapted from *Arabidopsis* data (Zhong et al. 2007a; Zhong et al. 2008; Ko et al. 2009; McCarthy et al. 2009; Nakano et al. 2010; Ohashi-Ito et al. 2010; Yamaguchi et al. 2010; Yamaguchi et al. 2011; Li et al. 2012; Zhong and Ye 2012; Öhman et al. 2013). Different tiers of the transcriptional hierarchy are shown, with Tier 1 being the top tier and Tier 4 being the bottom tier. The genes in each tier are displayed in the block representing that tier. Interactions between genes are shown by coloured edges connecting the genes. Direct interactions are represented by a solid edge. Indirect or uncharacterised interactions are represented by a dashed edge. Arrowheads indicate enhancement/upregulation while T-bars indicate repression/downregulation.

more highly expressed in interfascicular fibres than xylem vessels, while the opposite is true for MYB54 (Zhong et al. 2008). This indicates that while MYB52 and MYB54 might function mostly redundantly, they may have more specific roles in particular cell types.

MYB85 is upregulated by SND1 (Zhong et al. 2008) and is a direct target of MYB46 (Zhong et al. 2007a) but only an indirect target of MYB83 (Zhong and Ye 2012), Figure 2). Qualitative GUS analysis showed that MYB85 was expressed mostly in the xylem and interfascicular fibres and xylem vessels in *Arabidopsis* inflorescence stems (Zhong et al. 2008). In addition, dominant repression of this TF led to reduced SCWs in fibres and deformed vessel elements (Zhong et al. 2008). Of the set of TFs tested in the study, it was the only TF that, when overexpressed, resulted in ectopic deposition of lignin in epidermal and cortical cells. This, along with its ability to activate the *4CL1* promoter (Zhong et al. 2008), strongly implicates it in the biosynthesis of phenylpropanoids, compounds important for lignin production (Boerjan et al. 2003; Yang et al. 2011). In *Arabidopsis* roots, MYB85 is also expressed in xylem vessels at later stages of development (Nakano et al. 2010); which is where lignin is being deposited in the SCW (Mellerowicz et al. 2001; Plomion et al. 2001; Dejardin et al. 2010; Schuetz et al. 2013).

KNAT7 is expressed in the metaxylem, phloem, cambium and cortical cells of *Arabidopsis* inflorescence stems. It is a direct target of NST1, SND1, VND6, VND7 (Zhong et al. 2008) and MYB46 (Zhong and Ye 2012; Figure 2), showing that it is important for SCW biosynthesis in many different cell types. Knockout of KNAT7 in *Arabidopsis* results in a phenotype of thickened SCWs with high lignin content in interfascicular fibres, while overexpression results in thinner interfascicular fibre SCWs (Li et al. 2012). KNAT7 knockout plants also showed a slight upregulation of *CesA4* and a stronger upregulation of *CesA7* and *CesA8.* This implies that KNAT7 may be a negative regulator of lignin biosynthesis (Li et al. 2012) but not of cellulose or xylan biosynthesis. In further support of transcriptional repressor activity, KNAT7 has been shown to interact with known repressive complexes, such as those of OFP1 and OFP4 (Li et al. 2011), and MYB75 (Bhargava et al. 2010).

MYB103 is a direct target of NST1, NST2, VND6, VND7 and SND1 (Zhong et al. 2008; Yamaguchi et al. 2010), Figure 2). It was initially thought to be involved in cellulose biosynthesis of the interfascicular fibres of *Arabidopsis* inflorescence stems (Zhong et al.

2008). Since then it has also been shown to be expressed in metaxylem vessels in *Arabidopsis* roots (Nakano et al. 2010). It is also a downstream target of MYB46 and MYB83 (Figure 2), indicating a role in many different cell types (Nakano et al. 2010). It has recently been implicated in lignin biosynthesis (Öhman et al. 2013). MYB20, MYB63, MYB69, SND2 and SND3 are also down-regulated in MYB103 mutants, indicating many other possible functions (Öhman et al. 2013). RNAi mediated knockout of MYB103 — which represses *MYB103* only — in *Arabidopsis* yielded no change in cell wall thickness, while dominant repression — which represses all the targets of MYB103 — and overexpression of MYB103 resulted in thinner and thicker SCWs, respectively, in interfascicular fibres (Zhong et al. 2008). This would indicate that MYB103 functions redundantly with an as of yet undiscovered TF. MYB103 is also one of three TFs that were identified as being able to activate the *CesA8* promoter *in vitro,* the other two in the group being SND2 and SND3 (Zhong et al. 2008).

While SND2 and SND3 are both able to activate the *AtCesA8* promoter, a direct target analysis using an estrogen-inducible system showed that only SND3 was directly bound by SND1 (Zhong et al. 2008). However, SND2 is still downregulated in response to knockout of SND1 and NST1 (Zhong et al. 2008; Figure 2), suggesting that while they may be involved in the same processes, they occupy different levels of the transcriptional cascade. SND2 and SND3 appear to be active in both interfascicular fibres and xylem vessels, with much higher expression in the fibre cells, but the extent to which they are expressed varies greatly, with SND3 being expressed at almost four times the levels as that of SND2 (Zhong et al. 2008). SND2 and SND3 are both upregulated by MYB46 and MYB83 (Figure 2); though SND3 is upregulated more by MYB83 than MYB46, while SND2 is upregulated by about the same amount by these two TFs (Zhong et al. 2007a; McCarthy et al. 2009). In a bioinformatic study, SND3 was seen to be co-expressed with tissue-specific lignin biosynthetic genes, while SND2 was co-expressed with genes responsible for the biosynthesis of many different SCW polysaccharides (Shen et al. 2009). This points to SND3 having a role in lignin

biosynthesis, while the role of SND2 is still unclear. This, combined with the knowledge that SND2 is not directly bound by SND1 like the other *CesA8*-activating TFs (Zhong et al. 2008), makes it an interesting target to study.

Additionally, in a study identifying the direct targets of MYB46, two zinc finger proteins were identified, namely AT1G66810 and AT1G72220 (Ko et al. 2009). The first, designated as AtC3H14 by Ko et al. (2009) is a direct target of SND1 in addition to that of MYB46 (Figure 2), and is thought to be another potential mid-level master regulator (Ko et al. 2009). Other studies indicate the contrary, showing that AtC3H14 by itself is not sufficient to activate SCW biosynthesis (Zhong and Ye 2012). AtC3H14 has also been implicated in stress responses (Wang et al. 2008). Conversely, all that is known about AT1G72220 is that it is a downstream target of MYB46 (Ko et al. 2009; Figure 2) and is upregulated in response to overexpression of *AtSND2* (Hussey et al. 2011).

It has been proposed that the transcriptional network of SCW biosynthesis is conserved between higher plant species (Zhong et al. 2010b). There is much evidence to support this from studies conducted in *Populus*. For example, PtrWND2B & PtrWND6B were identified as functional orthologs of NST1 and SND1 in *Populus* (Zhong et al. 2010a). They were able to complement *Arabidopsis nst1 snd1* double mutants, and were shown to activate the expression of a number of *Arabidopsis* genes involved in SCW biosynthesis and that are known to be targets of AtSND1, such as *AtSND2*, *AtSND3, AtKNAT7, AtMYB52, AtMYB83, AtMYB85* and *AtMYB103* when overexpressed *in planta* (Zhong et al. 2010a). PtrWND2B & PtrWND6B were also shown to interact directly with the promoters of AtMYB46, AtSND3, AtMYB103 and AtKNAT7 using an estrogen-inducible system (Zhong et al. 2010a). In addition to this, other PtrWND TFs (namely PtrWND1B, PtrWND3B, PtrWND4B and PtrWND5B) were able to activate the promoters of the *Populus* orthologs of *Arabidopsis* SCW associated genes, such as AtSND2, AtSND3, AtMYB85, AtMYB46, AtMYB52, AtMYB54 and AtMYB83, among others (Zhong and Ye 2010). Furthermore, PtrMYB3 and PtrMYB20 were identified as functional orthologs of AtMYB46 and AtMYB83, respectively,

which are intermediate master regulators of SCW biosynthesis (McCarthy et al. 2010). The *Populus* orthologs were able to rescue *Arabidopsis myb46 myb83* double mutants, and were able to upregulate the expression of a number of SCW biosynthetic genes (AtCesA4, -7 and -8, AtIRX8 and -9, AtFRA8, At4CL1 and AtCCoaAOMT1) and SCW associated TFs (AtKNAT7, AtMYB42, AtMYB43, AtMYB52, AtMYB54, AtMYB58, AtMYB63 and AtMYB85; McCarthy et al. 2010). PtrMYB3 and PtrMYB20 were also able to activate the *Populus* SCW biosynthetic genes PtrCesA8, PtrGT43BP and PtrCCoaAOMT1P, and were themselves activated by PtrWND2B and PtrWND6B (McCarthy et al. 2010), the *Populus* orthologs of AtSND1 and AtNST1 (Zhong et al. 2010a). In *Eucalyptus gunnii,* two TFs have been identified as master switches of SCW biosynthesis. EgMYB1 (Legay et al. 2007) has been identified as a repressor of SCW biosynthesis, as overexpression of this TF in both *Arabidopsis* and *Populus* lead to plants with reduced SCWs (Legay et al. 2010). Conversely, EgMYB2 (Goicoechea et al. 2005) overexpression lead to xylem cells with thickened SCWs in *Nicotiana* plants (De Micco et al. 2012). Considering that EgMYB2 has already been identified as a possible ortholog of AtMYB46 and AtMYB83 (McCarthy et al. 2010) and given the conserved patterns of expression, binding and function seen between the *Populus* and *Arabidopsis* transcriptional networks of SCW biosynthesis, it is expected that the network will also be highly conserved in other hardwoods such as *E. grandis,* and therefore study of the network in a commercially important species such as *E. grandis* would be highly valuable.

Although much research has been performed on the transcriptional network governing SCW biosynthesis, large gaps still remain in our understanding. It has become evident that SCW biosynthesis is regulated by a network of tightly regulated TFs (Figure 2). However, many of the interactions between TFs and genes in the various biosynthetic pathways involved have not been identified, and to what extent TFs interact with other TFs in the network, or exactly how many TFs are involved in the network is still unknown. These are only a few of the many questions that still need to be answered. Techniques such as ChIP-seq and high-throughput Yeast Two-Hybrid screens may help us answer some of these questions. More

research needs to be performed on this transcriptional network in order to properly elucidate the role of each TF. More research also needs to be performed in woody species such as *Populus* and *Eucalyptus*, as this is where the research pertaining to wood formation may be applied the most effectively.

1.5.1 Cellulose biosynthesis

Cellulose is one of the most abundant biopolymers on the planet (Joshi and Mansfield 2007). The synthesis of cellulose is carried out by cellulose synthase (CESA) proteins. These proteins are located in the plasma membrane of vascular plants as symmetrical rosettes of six globular complexes (Kimura et al. 1999). Previously it was thought that each of these complexes has six subunits, and each subunit synthesizes a β-1,4-glucan chain, which then co-crystallises with the other chains in the subunit, and then the chains from the other five subunits, to produce a 36-glucan chain microfibril (Herth 1983; Kimura et al. 1999; Taylor 2008). However, a review by Guerriero et al. (2010) has pointed out how much is uncertain with regards to cellulose biosynthesis that needs to be investigated further for clarification (Guerriero et al. 2010).

The only known cellulose synthase proteins in higher plants are the CESA proteins. It was discovered that the *Arabidopsis* genome contains ten *CesA* genes (Holland et al. 2000; Richmond 2000). *Eucalyptus* has eleven (Ranik et al. unpublished), maize has twelve (Appenzeller et al. 2004), barley has eight (Burton et al. 2004) and *Populus* has at least seventeen (Kumar et al. 2009). The *CesA* genes in green algae were shown to be very similar to those in higher plants and have conserved intron sequences, indicating that the CesA genes are similar among all higher plants (Richmond 2000). *CesA4, -7 and -8* were identified in a study searching for novel cellulose deficient mutants as the *CesA* genes responsible for cellulose biosynthesis in the SCW and were found to be co dependant on each other for their function (Taylor et al. 2003). Additionally, co-expression analyses of *CesA4, -7 and -8* identified a number of TFs that were potentially regulating cellulose

biosynthesis in the SCWs of plant cells (Brown et al. 2005; Persson et al. 2005). Among these were SND2, KNAT7 (Persson et al. 2005) and AT1G72220 (Brown et al. 2005). In *E. grandis* the SCW *CesAs* are *CesA1, -2 and -3* (Ranik and Myburg 2006).

1.6 Promoters

Promoters are the primary mechanism through which genes are regulated. They contain core regulatory elements, as well as *cis*-elements, and provide a site for the binding of various regulatory proteins. A protein binding to a promoter can result in either upregulation, or repression of the expression of a gene (Hermsen et al. 2006). This happens through the protein recognising and binding to *cis*-elements. *Cis*-elements are DNA sequence regulatory motifs which allow proteins to bind to them to exert an effect. *Cis*-elements can be detected in many ways, using both experimental and *in silico* techniques. Experimental techniques include the use of ChIP-seq (Park 2009) to identify potential *cis*-elements and make use of techniques such as yeast one-hybrid (Y1-H; Wanke and Harter 2009) and electrophoretic mobility shift assays (EMSAs; Semenza and Wang 1992) to validate them. *In silico* methods use various algorithms and bioinformatic analysis, including several internet-based tools, which allow for the identification of novel regulatory motifs based on overrepresented motifs in sequence data (Priest et al. 2009; Ramirez and Basu 2009).

1.6.1 Promoter structure

The promoter structure can be divided into two main components, as can be seen in Figure 3. These are the proximal and distal promoters. In the proximal promoter, *cis*-elements to which the basal transcriptional machinery binds in order to initiate transcription (also known as core elements), are found. This may also encompass the transcriptional start site (TSS) and the 5'-untranslated region (5'UTR) of the gene. Some of these are key, such as the TATA box (Breathnach and Chambon 1981), the initiator (Inr) element (Smale and Baltimore 1989) located upstream of the TSS, and the Downstream Promoter Element (DPE; Burke

and Kadonaga 1997), which is typically located in the 5'-UTR. The distal promoter contains enhancers and repressors, non-core elements to which TFs bind to exert an effect. It should also be noted that while a distinction is made between these two promoter regions, they are not necessarily separate from each other. They may have a large degree of overlap, so that enhancers and repressors may be found anywhere in the proximal promoter, even overlapping into the 5'UTR and introns. However, it is still useful to distinguish between the proximal and distal promoters as the core elements are normally only found in the proximal region. Some of the core elements are known to have fixed positions (e.g. TATA box and DPE), but most other elements are not yet characterised well enough to determine whether their positions are fixed or not. In addition to the TATA box, Inr and DPE, there are many other core promoter elements (Juven-Gershon et al. 2008).

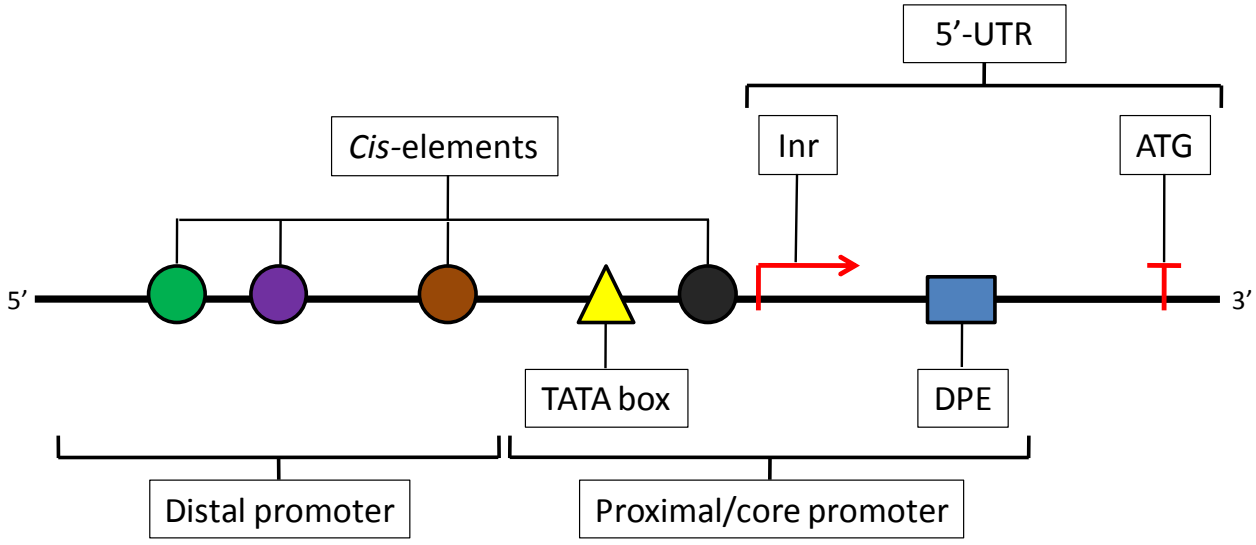


Figure 3: Schematic diagram of a simple promoter. The DNA sequence is represented by a bold black line. The *cis*-elements are represented by circles of different colours. The TATA box, Inr and DPE are represented by a yellow triangle, red right-angled arrow and blue rectangle, respectively. The translational start site (ATG) is represented by a red T-bar. The regions of the 5'-UTR, Proximal and Distal promoters are indicated on the diagram.

1.6.2 Core promoter elements

There are no core regulatory elements that are common to all promoters. However, there are some which occur more frequently than others (Juven-Gershon and Kadonaga 2010). The Initiator (Inr) element and TATA box, for example, occur frequently in many different promoter types and are two of the best studied and characterised motifs. The DPE will not be discussed in this review, as so far no DPE or DPE-like sequence has been identified in plant promoters (Molina and Grotewold 2005).

It is believed that the TATA box is one of the most ancient motifs, conserved from *Archaea* (Hausner et al. 1991). It is fairly common in eukaryotes and it is present in about one third of plant promoters (Molina and Grotewold 2005). It has a consensus sequence of TATAWAAR which is located at approximately -32 in *Arabidopsis* promoters (Molina and Grotewold 2005). It is the recognition site for a class of proteins known as TATA binding proteins (TBPs), which associate with TFIID in the transcriptional initiation complex (Roeder 1996). TATA boxes are often found in promoters involved in specific processes. For example, a TATA box motif is found in many plant genes that are light responsive (Kiran et al. 2006).

The (Inr) element is probably the most common element in core promoters. It is believed to have a consensus of TCAKTY in *Drosophila* (Burke and Kadonaga 1996), YYANWYY in humans (Burke and Kadonaga 1997) and YTCANTYY (Nakamura et al. 2002). In *Arabidopsis* and rice, an additional sequence has been identified which appears to have Inr activity, known as the YR motif (Yamamoto et al. 2007). The Inr is normally the site of initiation of transcription for focused promoters — promoters which contain only one TSS, or a number of TSSs in a narrow region of nucleotides — (Juven-Gershon et al. 2008), as TFIID has been strongly associated with this particular element. The A in the consensus sequence is normally designated as the +1 position (i.e. the first transcribed base). There are other less well-known core promoter elements. For a review of these see the paper by Juven-Gershon and colleagues (2008).

1.6.3 Secondary cell wall biosynthesis related *cis*-regulatory elements

A number of motifs important for SCW biosynthesis have been identified. For example, AC elements important for lignin biosynthesis have been identified in poplar (Winzell et al. 2010). Also, two MYB binding sites called SMRE (ACCWAMY; Zhong and Ye 2012) and MYB46RE (RKTWGGTR; Kim et al. 2012) have been identified. MYB46RE is believed to be a binding site of MYB46 and thus an indicator of genes regulated by this protein. However, the SMRE element is highly similar to the MYB46RE and is also bound by MYB83. Therefore, it is likely that MYB46RE and SMRE are the same element, and could be a general MYB binding site. Recently the SNBE element (WNNYBTNNNNNNNAMGNHW; Zhong et al. 2011b; McCarthy et al. 2011) was identified. At first it was believed that this motif was exclusively an SND1 binding site, however, a number of the NAC master regulators have been shown to bind to it (Zhong et al. 2011b; McCarthy et al. 2011). The presence of this motif in the promoters of genes associated with SCW biosynthesis, as well as being a target to master regulators of SCW biosynthesis such as VND6, VND7, NST1, NST2 and SND1, points to an important role of the SNBE element in SCW biosynthesis (Zhong et al. 2011b; McCarthy et al. 2011).

1.7 *Cis*-elements identification and validation

TF proteins recognise cis-elements/regulatory motifs with varying specificity, bind to them, and then exert a regulatory effect on the gene associated with the element (Istrail and Davidson 2005). There are many different *cis*-elements, and there is still much to be discovered about their composition and exactly how they function. Because of the diversity and sheer number of *cis*-regulatory elements, they can be very difficult to identify and characterise. For this reason, different techniques have been developed to accurately identify and validate *cis*-elements in genomic DNA sequence. These techniques can be classified into experimental and *in silico* techniques. Experimental techniques deal with the identification of a motif *in vivo* by identifying the binding site of a specific protein. The *in silico*

methods make use of bioinformatic algorithms and statistical analysis in order to predict and infer the location and composition of DNA-binding motifs. Techniques such as Y1-H, EMSA and ChIP-seq will be discussed, as well as online cis-element repositories and various online *cis*-element prediction tools.

1.7.1 Experimental techniques

Yeast one-hybrid

Y1-H is a technique used to study interactions of proteins with DNA (Ouwerkerk and Meijer 2001). cDNA libraries are constructed from the tissue of interest. The libraries are then expressed in yeast as a fusion protein, with the DNA binding domain (BD) of the proteins to be investigated being bound to an activation domain (AD). The construct coding for the fusion protein (prey) is transformed into a yeast cell along with a reporter gene under the control of a promoter containing the sequence of interest (bait). If there is a positive interaction and the BD recognises the bait sequence, it will bind and the AD will initiate expression of the reporter gene that signals a positive interaction. In this way, a protein may be discovered which binds to a certain DNA sequence, and a potential motif to which a protein binds can be evaluated (Wanke and Harter 2009). While a particular DNA motif may be easily resolved with Y1-H by using tandem repeats of the motif as bait (Wanke and Harter 2009), resolution on a single base level will be difficult. In this case, another technique such as EMSA (discussed later in this review) may be better suited. However, when identifying novel TFs that bind a DNA fragment, very large fragments may be screened, in both *Caenorhabditis elegans* and *Arabidopsis,* with reports of up to 3 kb of promoter sequence being used successfully (Deplancke et al. 2004; Brady et al. 2011; Gaudinier et al. 2011).

Due to the *in vivo* nature of the Y1-H assay, it is susceptible to false positives (Vidalain et al. 2004). There are two main types of false positives in Y1-H assays. The first is autoactivation, where the bait sequence is recognised by an endogenous yeast protein and expression of the reporter gene is activated (Lopato et al. 2006). The second is leaky reporter gene

expression, where non-specific factors activate expression of the reporter gene (Lopato et al. 2006). Autoactivation depends on the bait sequence being used. It cannot be reduced but the level of autoactivation of a bait sequence can be determined using the appropriate controls (Lopato et al. 2006). Leaky expression can be countered by the addition of a competitive inhibitor of the reporter gene (e.g. 3-aminotriazole for the inhibition of the *HIS3* gene; Lopato et al. 2006). This ensures that only TFs that interact with the bait sequence will activate reporter gene expression. False negatives also occur in Y1-H. Many TFs require additional protein partners, factors or modifications in order to interact with DNA. These may not be present in the yeast cell and so will produce a negative result (Chen and Shin 2008). Also, if the analysis is too stringent (i.e. high levels of competitive inhibitor), TFs that normally have weak interactions may be missed (Chen and Shin 2008). Lastly, the expressed prey protein may be toxic to the yeast cell and cause it to die (Coates and Hall 2003; Chen and Shin 2008). The abundance of false positives and false negatives in Y1-H analyses highlight a need for validation. This may be done with techniques such as EMSA (Kerr 1995; Hellman and Fried 2007), ChIP-seq (Park 2009) and transactivation analyses in plant protoplasts (Abe et al. 1997; Wehner et al. 2011).

Y1-H has been in use for many years. Wei et al. (1999) made use of the technique to identify a transcriptional regulator of the egg hatching enzyme in sea urchin embryos. *cDNA-GAL4* activation domain fusions were constructed using the RNA from sea urchin eggs. A promoter construct was created using the native *SpHE* promoter sequence from -324 to -143, which ensured that the possible *cis*-elements that any proteins may bind to were present. A DNA binding protein designated SpEts4 was identified and found to upregulate the *SpHE* gene (Wei et al. 1999). Y1-H is still in use today as an effective technique of characterising promoter protein interactions. For example, a high efficiency Y1-H and Yeast two-hybrid system was developed (Mitsuda et al. 2010). By building a library of approximately 1500 full-length cDNA preys and submitting a single bait sequence to Y1-H analysis with this library, a number of protein interactions with the promoter can be quickly and efficiently characterised.

It has also become common practice to use a library of TFs specific to a particular process or environment. For example, Pruneda-Paz et al. (2009) used a library of circadian clock associated TFs in a Y1-H analysis to determine which proteins were responsible for regulation of the *CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)* and *LATE ELONGATED HYPOCOTYL (LHY)* genes.

Modifications of yeast one-hybrid

It is possible to modify the Y1-H system. For example, Gateway cloning (www.invitrogen.com) is popular because of its high efficiency, and one can clone many DNA fragments into different vectors because it makes use of homologous recombination to facilitate cloning rather than the traditional restriction enzyme digestion and ligation (Hartley et al. 2000). A modified system making use of the Gateway technology (Curtis and Grossniklaus 2003) was effectively demonstrated by Deplancke et al. (2004). DNA bait-reporter gene constructs were generated through cloning, and then integrated into the genome of a yeast strain being used, into a marker locus using homologous recombination. The bait-reporter gene constructs also contained a wild-type marker, making them easy to select for (Deplancke et al. 2004). Another modification of the Y1-H system involves using yeast mating types to increase efficiency (Lopato et al. 2006). By transforming the bait and prey constructs into separate yeast mating-types, a Y1-H screen can be achieved by mating the different strains. This is more efficient because like mating-types cannot mate, so the mixing of a bait and prey is guaranteed with every mating.

Yeast one–hybrid analysis may also be used in tandem with other techniques. In a previous study (Kim et al. 2007), a yeast-one hybrid screen was used in conjunction with yeast two-hybrid screen in a dual reporter system. The first screening step made use of Y1-H analysis and indicated a positive result by means of growth on a histidine deficient medium. The second screening step made use of yeast two-hybrid and indicated a positive result by blue yeast colonies growing on a medium containing X-gal. In this way it is possible to detect protein complexes binding to DNA sequences.

A modification of Y1-H, known as Reverse One-Hybrid, can be used to identify proteins which do not interact with the bait sequence. In this assay, the standard *HIS3* reporter gene is replaced with a toxic gene. This means that any interaction of a prey will activate the reporter gene and kill the cell, and only those deficient in interaction will grow. This is particularly useful for identifying mutations in proteins which disrupt binding ability and dissociator molecules which may have some therapeutic value (Vidal et al. 1996).

Electrophoretic mobility shift assay

EMSA is a technique which allows characterisation of the interaction between a specific protein and a DNA fragment to which, it is suspected, this protein binds (Kerr 1995; Hellman and Fried 2007). What makes this technique powerful is that one may test one protein against many binding sites, or test more than one protein against a single site (Kerr 1995). However, EMSA requires expression of a protein, which is a laborious process. The protein of interest is incubated with the DNA being investigated, and is then analysed by gel electrophoresis. The DNA may be labelled radioactively or with a non-radioactive florescent dye. If the protein binds to the DNA sequence, the progress of the DNA through the gel will be retarded due to the size added by the protein, as compared to unbound DNA (Hellman and Fried 2007). In this way protein-DNA interactions can be confirmed and regulatory motifs may be characterised. The specificity of an interaction can also be determined by making use of competitive EMSA assays where unlabelled probe may be added to compete with the labelled probe for protein binding (Hellman and Fried 2007).

Two well known modifications of the EMSA technique are Capillary EMSA (CEMSA) and Immunodepletion EMSA (IDEMSA). CEMSA is performed by performing EMSA with Capillary Electrophoresis, rather than conventional agarose gel electrophoresis. This allows for more rapid separation and the use of fluorescent dyes for detection instead of radioactive materials, and is far more sensitive than traditional EMSA (Xian et al. 1996; Foulds and Etzkorn 1998). IDEMSA was used to elucidate the proteins which were playing a role in forming complexes that bind to DNA. By removing a particular protein from an extract and

then submitting the extract to EMSA and supershift EMSA analyses (an EMSA where a protein specific antibody is also added to the assay to increase the size of the complexes which creates a larger shift), it was possible to determine whether a particular protein was required for interaction with a DNA sequence (Dyer and Herzog 1995).

ChIP-seq

ChIP-seq is a technique which allows for the rapid genome-wide identification of binding sites of a particular protein (Park 2009). It is a modification of ChIP, or Chromatin Immunoprecipitation. ChIP was predominantly used to identify binding sites in protein DNA complexes and had higher resolution than other techniques such as DNA footprinting assays (Solomon et al. 1988). In a typical ChIP experiment, proteins and chromatin are cross-linked *in vivo* using either formaldehyde or UV-light (Pruss and Bavykin 1997). The cross-linked protein and DNA are then sonicated to shear the chromatin into fragments not larger than 1 kb. The excess debris is removed, and the chromatin-protein complexes are then pulled down or precipitated using antibodies against either epitopes on the protein of interest or tags present on the expressed protein. The captured complexes are collected and washed to remove any non-specifically bound chromatin. The cross-links are then reversed, freeing the chromatin (Pruss and Bavykin 1997). This chromatin can be analysed using various techniques to determine the binding site of the protein of interest (Dey et al. 2012).

The next step of ChIP was to incorporate microarrays in order to determine binding sites across genomes. This modification of ChIP is known as ChIP-chip (Buck and Lieb 2004). In these analyses, the enriched chromatin from the ChIP assay is PCR amplified and labelled with fluorescent dye. Genomic DNA prepared from the immunoprecipitation input sample is labelled with a different fluorescent dye and used as a reference. Both these samples are then hybridised to a slide containing probes which represent the entire genome. By looking at the fluorescence patterns, it is possible to identify regions of the genome that are enriched by immunoprecipitation (Ren et al. 2000).

Most recently, next-generation sequencing technologies have influenced ChIP. By performing ChIP, submitting the enriched chromatin to next-generation sequencing and analysing the returned data against a reference genome, it is possible to identify regions of the genome which show enrichment for ChIP signal, giving a relatively accurate representation of the binding sites of the protein across an entire genome (Zhang et al. 2008). ChIP-seq is advantageous over ChIP-chip as it is able to reveal a peak of binding activity in the genome; it is not subject to noise from hybridisation present in microarrays; and it is easier to quantify, as ChIP-chip fluorescence signals may not behave in a linear manner and can reach saturation. ChIP-seq is also superior as one is not limited by what is present on the probe array. Rather, the whole genome is available for analysis (Park 2009). Due to its reliance on next generation sequencing, there are some disadvantages. For example, ChIP-seq is subject to sequencing error and is biased towards chromatin fragments with high GC content due to the preparation and amplification of the sequence libraries used in the analysis (Hillier et al. 2008; Quail et al. 2008; Chen et al. 2013). Additionally, it should be noted that even though a protein may display many ChIP-seq peaks across a genome, there is no guarantee that all of these are functional binding sites for that protein (Zhang et al. 2008), so experimental validation of sites of interest would be useful.

1.7.2 *In silico* resources

Cis-*element repositories*

There are many web-based tools and applications that are available for *cis*-element discovery and characterisation. These are divided into *cis*-element discovery tools and *cis*-element databases. The databases are usually searchable and provide links to literature pertaining to previously described regulatory motifs. A few examples of these databases can be seen in Table 1. However, these websites are only repositories of information and they lack any real means to perform *de novo* prediction of protein binding sites *in silico.*

Web-based cis-*element discovery*

There are a number of online *cis*-element discovery tools available for use. These tools generally favour one of three main methods in their approach to motif discovery. The first method is known as the Enumerative method, and is based on the identification of overrepresented motifs in specified DNA sequences. A motif is identified and a consensus sequence of the motif is constructed by combining different versions of the sequence motif that have been indentified (D'Haeseleer 2006). An example of a tool using this approach is Weeder (Pavesi et al. 2004; Pavesi et al. 2007). The second method of motif discovery is known as Deterministic optimization. In this method, a motif is identified, and the probability that it is a real motif and not background sequence is calculated. A consensus sequence is then constructed from the weighted averages across the probabilities calculated for each potential motif. This process is repeated for every possible motif identified in the sequence (D'Haeseleer 2006). An example of a tool using this approach is MEME (Bailey et al. 2006; Bailey et al. 2009). The third method is known as Probabilistic optimization. This method makes use of Gibbs sampling and is similar to Deterministic optimization in that it identifies candidate motifs against the distribution of the background sequence. However, it differs, in that, instead of taking weighted averages, a weighted sample is taken. The sample is created from randomly selected sites and each new candidate motif is probabilistically scored against the sample. The motif is either included or discarded, and the process repeats itself with the new weighted sample set (D'Haeseleer 2006). An example of a tool using this method is Motifsampler (Thijs et al. 2001; Thijs et al. 2002).

The best method to use for motif discovery is debatable. Each method has its own advantages and drawbacks. Therefore it is generally thought that the best way to perform a motif discovery analysis would be to use a combination of the above mentioned techniques (D'Haeseleer 2006). This can be achieved by using one of many programs or online tools which allow comparison of the results from other motif finding programs. This is known as a consensus ensemble approach and allows for the analysis of the overlap between many *cis*-

Table 1: *Cis*-element databases available on the internet.

Database name[a]	Web address[b]	Description[c]	Primary reference[d]
AGRIS	*Arabidopsis*.med.ohio-state.edu	Contains AtcisDB, a database of both predicted and experimentally validated *Arabidopsis* regulatory elements	Palaniswamy et al. 2006
AtCOECis	bioinformatics.psb.ugent.be/ATCOECIS	Database that uses *Arabidopsis cis*-elements to draw coexpression networks	Vandepoele et al. 2009
Athamap	www.athamap.de	Database of TF and small RNA binding sites in *Arabidopsis*	Steffens et al. 2005
Athena	www.bioinformatics2.wsu.edu/Athena	Database of *Arabidopsis cis*-elements with visualisation tool	O'Connor et al. 2005
CisView	lgsun.grc.nia.nih.gov/cisview	Database of mouse cis-elements	Sharov et al. 2006
DATF	datf.cbi.pku.edu.cn	Database of *Arabidopsis* TFs and their binding sites	Guo et al. 2005
DoOP	doop.abc.hu	Database of plant and chordate promoters used to find conserved *cis*-elements	Barta et al. 2005
EPD	epd.vital-it.ch	Database of eukaryotic promoter data	Praz et al. 2002
JASPAR	jaspar.genereg.net	Database of eukaryotic TF binding sites	Chang et al. 2008
PLACE	www.dna.affrc.go.jp/PLACE	Database of plant *cis*-elements	Higo et al. 1999
Plant Promoter DB	ppdb.agr.gifu-u.ac.jp	Database of rice, moss and *Arabidopsis* promoter data	Yamamoto and Obokata 2008
PlantCARE	bioinformatics.psb.ugent.be/webtools/plantcare/html/	Database of plant *cis*-elements	Rombauts et al. 1999; Lescot et al. 2002
PlantProm DB	www.softberry.com/berry.phtml?topic=plantprom...plantprom	Database of plant promoter data	Shahmuradov et al. 2003
REDfly	redfly.ccr.buffalo.edu	*Drosophila* cis-element database	Gallo et al. 2006
RiceSRTFDB	ricefrend.dna.affrc.go.jp	Rice TF and *cis*-element database	Priya and Jain 2013
TAIR	www.*Arabidopsis*.org	Contains links to various *Arabidopsis* promoter databases and tools	Swarbreck et al. 2008
Transfac	www.gene-regulation.com/pub/databases.html	Database of eukaryotic TFs and their binding sites	Matys et al. 2006
TRED	rulai.cshl.edu/TRED	Mammalian *cis*- and *trans*- regulatory element database	Zhao et al. 2005b

[a]The name of the database
[b]The web address at which the database may be found
[c]A short description of the information that the database contains
[d]The primary reference in which the database was reported in literature

element prediction datasets, thus increasing confidence in the predictions (Hu et al. 2005; MacIsaac and Fraenkel 2006; Tompa et al. 2005). Another thing to consider is sensitivity versus specificity. If the criteria for *cis*-element identification are very stringent, few hits with high confidence will be obtained (i.e. high specificity and low sensitivity) (GuhaThakurta 2006). If the criteria are not stringent, many hits with low confidence will be obtained (i.e. low specificity and high sensitivity; GuhaThakurta 2006). In the first approach you may miss many real *cis*-elements that are not as statistically significant as others, while in the second approach these less significant *cis*-elements will be identified, but many false positives will be obtained as well. Whether to take a specific or sensitive approach is largely subjective, and will vary between experiments.

1.8 The role of the promoter in transcriptional regulation

Understanding the interactions between TFs and promoter sequences is only part of the complexity of Eukaryotic gene expression regulation. A TF binds to a specific DNA motif and exerts an effect on its target (Istrail and Davidson 2005). However, we are becoming increasingly aware of how complex this interaction really is. This subject is extensive, so only a few basic principles will be discussed in this section. For a recent review on this subject, see Lelli et al. (2012). Trees are multicellular organisms which require complex control to ensure proper spatial and temporal expression in different cell types at different levels of development. Taking this into consideration, the dynamic nature of the cellular environment and especially that of the chromatin itself is important for the proper regulation of transcription.

In the nucleus, DNA is found in the form of chromatin. The most basic unit of chromatin is known as the nucleosome. This consists of 147 bp of DNA wrapped around a histone octamer (Li and Reinberg 2011) consisting of two H2A, H2B, H3 and H4 subunits each (Luger et al. 1997). The DNA that is wrapped around histones is less active and it is more difficult for TFs and polymerases to access. Generally, chromatin is found in one of two

states, either it has an active, open conformation and is known as euchromatin, or it has a tight inactive conformation and is known as heterochromatin (Jenuwein and Allis 2001; Felsenfeld and Groudine 2003). Post-translational modification of histones can affect the accessibility of chromatin (Felsenfeld and Groudine 2003). For example, histone complexes which have been monomethylated at the 4th lysine residue of H3 (H3K4me1) are characteristic of active chromatin where TFs are likely to bind. Histone complexes dimethylated at the 4th lysine residue of H3 (H3K4me2) are associated with enhancers, promoters and TSSs, and histone complexes trimethylated at the 4th lysine residue of H3 (H3K4me3) are believed to indicative of promoters and TSSs of genes that are being actively transcribed (Barski et al. 2007; Heintzman et al. 2007; Rando and Chang 2009). It is also important to note that active chromatin does not necessarily indicate high expression, as H3K4me1 can be associated with repression as well as activation (Barski et al. 2007; Heintzman et al. 2007; Rando and Chang 2009). Additionally, histone complexes that have been trimethylated at the 27th lysine residue (H3K27me3) have been associated with repressed gene function (Pauler et al. 2009). These factors affect the conformation of chromatin and, subsequently, the ability for regulatory proteins to bind.

The interactions between various proteins and DNA are as complicated as that of chromatin. Due to multiple gene and sometimes whole-genome duplications in an organism's evolutionary history there are many families of proteins that have similar domains and DNA recognition specificities, resulting in functional redundancy. A good example of this are NST1 and NST2 genes in *Arabidopsis*, where both proteins need to be knocked out for any phenotype to be visible (Mitsuda et al. 2005). Apart from this, proteins can also work together to bind DNA. This phenomenon is known as cooperativity. There are three main types of cooperativity:

The first is known as classic cooperativity. This is when the direct interactions between proteins allow them to increase their binding affinity. This is largely the case for proteins which function as homo- or heterodimers, such as the previously discussed NAC or MYB

proteins (Olsen et al. 2005; Dubos et al. 2010). There is also a variation of this known as latent specificity, in which the direct interaction between proteins not only leads to an increase in affinity but also a change in specificity, so that together the proteins involved may bind different targets than they would individually (Slattery et al. 2011).

The second is known as enhanceosome or modular cooperativity. This is the phenomenon of a large number of proteins binding to DNA in a complex that does not seem to rely on direct interactions between proteins. A clear example of this cooperation can be seen with the *cis*-regulatory module (CRM) for viral inducible expression of IFN-β in humans (Thanos and Maniatis 1995; Panne 2008). This enhanceosome consists of eight proteins binding to a 55 bp enhancer (Panne 2008). If one protein is removed, the expression of IFN-β is lost (Thanos and Maniatis 1995). However, there is a lack of protein-protein interactions in this complex, and that the complexes themselves can be highly flexible in arrangement (Pan and Nussinov 2011), indicating that DNA might be the organising factor (Panne 2008).

The third type of cooperativity is known as collaborative competition. This type of cooperation is different from the others, as while the other two take place on regions of DNA not bound in nucleosomes (i.e. "naked" DNA), collaborative competition occurs in the vicinity of nucleosomes. In this cooperation, many different TFs can bind with much higher affinity than single TFs, and are able to compete with the nucleosome for binding space (Miller and Widom 2003).

The manner in which proteins interact with each other and chromatin, as well as the state of the chromatin itself, are important but it should be noted that the spatial distribution of chromatin also affects regulation. It has been shown that chromosomes tend to occupy specific locations in the nucleus (Zorn et al. 1979). These locations are dynamic, and constantly changing, making it difficult to study these phenomena (Cavalli 2007). However, the advent and advancement of technologies, such as chromosome conformation capture (3C; Dekker et al. 2002) and chromosome conformation capture carbon-copy (5C; Dostie et

al. 2006), are allowing for great progress in this regard (de Wit and de Laat 2012). In the meantime, some patterns are becoming obvious. For example, those regions of chromatin which are rich in genes tend to be found near the centre of the nucleus, while those regions poor in genes are found closer to the periphery (Dostie and Bickmore 2012). Also active genes are normally found near the surface of chromosome localities, while inactive genes are normally found towards the interior (Sexton et al. 2012). Highly expressed genes can sometimes be found together in transcriptional factories (Razin et al. 2011), indicating that genes that are colocalised may also be coregulated. However, this raises the question of how these genes are colocalised in the first place, especially if they are on different chromosomes.

Chromosomes can interact with each other. CRMs can be far from the targets which they regulate. This is possible due to two phenomena, namely looping for interactions which are *cis* (Krivega and Dean 2012), and transvection (where a promoter may interact with an enhancer on a separate chromosome) for those interactions which are *trans* (Duncan 2002). Also, recently, the ENCODE project has provided much evidence for long range interactions between sections of chromatin, showing that one enhancer may be used for the activation of many different genes (Thurman et al. 2012).

1.9 Conclusion

Much research has been done regarding the transcriptional regulation of xylem development and SCW deposition in *Arabidopsis*, and at least some elements of this network have been displayed in woody species such as *Populus,* indicating that the regulation of SCW biosynthesis might be very similar to that of a herbaceous plant such as *Arabidopsis.* However, comparatively little research has gone into the transcriptional network of SCW biosynthesis in *Eucalyptus.* If we are to make full use of the resources available to us through hardwood crops, this will need to be remedied. Even in *Arabidopsis,* there is still much uncertainty about the transcriptional regulation of xylem development. Many TFs

(mostly NAC TFs and MYB TFs) have been identified as playing a role in transcriptional regulation of SCW formation, but their exact functions and the mechanisms through which they work are still unclear. Also, the extent to which these proteins interact with each other is largely unknown. Most NAC TFs and some MYB TFs bind as either homo- or heterodimers so it is likely that these interactions will play an important role in the network. In order to properly understand and dissect the transcriptional network regulating SCW biosynthesis, these TFs will need to be characterised.

However, the TFs are not the only part of this network. The promoters of the genes involved in transcriptional regulation of SCW biosynthesis need to be analysed and studied, as they are a vitally important component. *Cis*-elements that are associated with lignin biosynthesis have been identified (Winzell et al. 2010), but so far no elements specific to cellulose and hemicellulose biosynthesis have been identified. With the advent of resources such as TRANSFAC (Matys et al. 2006) and other online *cis*-element prediction tools (see section 1.7.2), it is becoming easier to identify and characterise *cis*-elements. However, these online resources still require much work before their accuracy can be improved to approach that of experimental methods.

Over the past decade, the experimental techniques used to characterise binding between a protein and DNA have advanced considerably. The advent of next generation sequencing has allowed for high throughput analyses that were not possible in the past. Complementary information from technologies such as EMSAs, Y1-H and ChIP-seq are improving our resolution. However, very few techniques account for the complexity of the native environment in which these transcriptional network interactions occur. In an *in situ* interaction, there are many proteins, cofactors and environmental changes which will affect binding, and this is severely underrepresented in the current techniques. Apart from a new technique designed to simulate these conditions, a combination of current techniques may go some way towards mitigating this problem.

Transcriptional regulation is an incredibly complex process. We tend to think of transcriptional regulation in terms of a single protein binding to a single DNA sequence, while in reality, this is far from correct. Apart from the interactions between proteins and DNA, and the cooperation of different proteins leading to different affinities and specificities of proteins for sequences, there are the additional layers of regulation in the form of the states of the histones in the nucleosome, and the conformation of the chromatin itself. There are also many other layers not mentioned in this review, such as epigenetic regulation and that mediated by non-coding RNA (Jones-Rhoades et al. 2006; Ponting et al. 2009). Therefore, in order to properly understand transcriptional regulation in the future, we will have to view these various components of regulation in an integrated and holistic manner.

Therefore, the aim of this study is to characterise a piece of this extremely complex and important network in *E. grandis*, with a focus on the *EgrSND2* promoter. The tissues (if any) in which the *EgrSND2* promoter induces expression in *Arabidopsis* will be identified. *E. grandis* orthologs of TFs important for SCW biosynthesis in *Arabidopsis* will also be identified, and their ability to interact with the *EgrSND2* promoter will be evaluated. To do this, qualitative GUS analyses will be performed on a 1.5 kb *EgrSND2* promoter fragment in six-week old *A. thaliana* plants. Putative *E. grandis* orthologs of *Arabidopsis* SCW TFs will be identified using phylogenetic analyses and the abilities of orthologs to bind to the *EgrSND2* promoter will be elucidated using Y1-H analyses. By determining in which *Arabidopsis* tissues the *EgrSND2* promoter is active and identifying any direct connections between this promoter and other TFs in the transcriptional network of SCW biosynthesis in *E. grandis*, we will provide a stepping stone for further research. This will provide the groundwork needed to understand the transcriptional network on a protein-promoter level, and eventually integrate it into the wider field of transcriptional regulation of wood formation. This work will also lead to insight into promoter-protein interactions and, as such, may also form a basis for study in the field of synthetic promoters, which would have applications in many different fields.

1.10 References

Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D. and Shinozaki, K. (1997) Role of Arabidopsis MYC and MYB homologs in drought-and abscisic acid-regulated gene expression. *The Plant Cell Online* 9: 1859-1868.

Aida, M., Ishida, T., Fukaki, H., Fujisawa, H. and Tasaka, M. (1997) Genes involved in organ separation in *Arabidopsis*: an analysis of the cup-shaped cotyledon mutant. *Plant Cell* 9: 841-857.

Appenzeller, L., Doblin, M., Barreiro, R., Wang, H., Niu, X., Kollipara, K., et al. (2004) Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (*CesA*) gene family. *Cellulose* 11: 287-299.

Badescu, G.O. and Napier, R.M. (2006) Receptors for auxin: will it all end in TIRs? *Trends in Plant Science* 11: 217-223.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., et al. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* 37: W202-208.

Bailey, T.L., Williams, N., Misleh, C. and Li, W.W. (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research* 34: W369-373.

Baima, S., Possenti, M., Matteucci, A., Wisman, E., Altamura, M.M., Ruberti, I., et al. (2001) The *Arabidopsis* ATHB-8 HD-zip protein acts as a differentiation-promoting transcription factor of the vascular meristems. *Plant Physiology* 126: 643-655.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., et al. (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129: 823-837.

Barta, E., Sebestyen, E., Palfy, T.B., Toth, G., Ortutay, C.P. and Patthy, L. (2005) DoOP: Databases of Orthologous Promoters, collections of clusters of orthologous upstream sequences from chordates and plants. *Nucleic Acids Research* 33: D86-90.

Bennett, T., van den Toorn, A., Sanchez-Perez, G.F., Campilho, A., Willemsen, V., Snel, B., et al. (2010) SOMBRERO, BEARSKIN1, and BEARSKIN2 regulate root cap maturation in *Arabidopsis*. *Plant Cell* 22: 640-654.

Bhargava, A., Mansfield, S.D., Hall, H.C., Douglas, C.J. and Ellis, B.E. (2010) MYB75 functions in regulation of secondary cell wall formation in the *Arabidopsis* inflorescence stem. *Plant Physiology* 154: 1428-1438.

Boerjan, W., Ralph, J. and Baucher, M. (2003) Lignin biosynthesis. *Annual Review of Plant Biology* 54: 519-546.

Borevitz, J.O., Xia, Y., Blount, J., Dixon, R.A. and Lamb, C. (2000) Activation Tagging Identifies a Conserved MYB Regulator of Phenylpropanoid Biosynthesis. *The Plant Cell Online* 12: 2383-2393.

Brady, S.M., Zhang, L., Megraw, M., Martinez, N.J., Jiang, E., Yi, C.S., et al. (2011) A stele-enriched gene regulatory network in the Arabidopsis root. *Molecular systems biology* 7: 459.

Breathnach, R. and Chambon, P. (1981) Organization and expression of eukaryotic split genes coding for proteins. *Annual Review of Biochemistry* 50: 349-383.

Brown, D.M., Zeef, L.A., Ellis, J., Goodacre, R. and Turner, S.R. (2005) Identification of novel genes in *Arabidopsis* involved in secondary cell wall formation using expression profiling and reverse genetics. *The Plant Cell Online* 17: 2281-2295.

Buck, M.J. and Lieb, J.D. (2004) ChIP-chip: Considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* 83: 349-360.

Burke, T.W. and Kadonaga, J.T. (1996) *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes & Development* 10: 711-724.

Burke, T.W. and Kadonaga, J.T. (1997) The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes & Development* 11: 3020-3031.

Burton, R.A., Shirley, N.J., King, B.J., Harvey, A.J. and Fincher, G.B. (2004) The *CesA* gene family of barley. Quantitative analysis of transcripts reveals two groups of co-expressed genes. *Plant Physiology* 134: 224-236.

Carlsbecker, A., Lee, J.-Y., Roberts, C.J., Dettmer, J., Lehesranta, S., Zhou, J., et al. (2010) Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. *Nature* 465: 316-321.

Cassan-Wang, H., Goué, N., Saidi, M.N., Legay, S., Sivadon, P., Goffner, D., et al. (2013) Identification of novel transcription factors regulating secondary cell wall formation in *Arabidopsis*. *Frontiers in plant science* 4: 189.

Cavalli, G. (2007) Chromosome kissing. *Current Opinion in Genetics & Development* 17: 443-450.

Chang, W.C., Lee, T.Y., Huang, H.D., Huang, H.Y. and Pan, R.L. (2008) PlantPAN: Plant promoter analysis navigator, for identifying combinatorial *cis*-regulatory elements with distance constraint in plant gene groups. *BMC Genomics* 9: 561.

Chen, G. and Shin, J.A. (2008) AhR/Arnt: XRE interaction: Turning false negatives into true positives in the modified yeast one-hybrid assay. *Analytical biochemistry* 382: 101-106.

Chen, Y.-C., Liu, T., Yu, C.-H., Chiang, T.-Y. and Hwang, C.-C. (2013) Effects of GC Bias in Next-Generation-Sequencing Data on *De Novo* Genome Assembly. *PLoS One* 8: e62856.

Coates, P.J. and Hall, P.A. (2003) The yeast two-hybrid system for identifying protein–protein interactions. *The Journal of Pathology* 199: 4-7.

Curtis, M.D. and Grossniklaus, U. (2003) A gateway cloning vector set for high-throughput functional analysis of genes *in planta*. *Plant Physiology* 133: 462-469.

Czemmel, S., Heppel, S.C. and Bogs, J. (2012) R2R3 MYB transcription factors: key regulators of the flavonoid biosynthetic pathway in grapevine. *Protoplasma* 249 Suppl 2: S109-118.

D'Haeseleer, P. (2006) How does DNA sequence motif discovery work? *Nature Biotechnology* 24: 959-961.

De Micco, V., Ruel, K., Joseleau, J.-P., Grima-Pettenati, J. and Aronne, G. (2012) Xylem anatomy and cell wall ultrastructure of *Nicotiana tabacum* after lignin genetic modification through transcriptional activator EGMYB2. *IAWA Journal-International Association of Wood Anatomists* 33: 269-286.

de Wit, E. and de Laat, W. (2012) A decade of 3C technologies: insights into nuclear organization. *Genes & Development* 26: 11-24.

Dejardin, A., Laurans, F., Arnaud, D., Breton, C., Pilate, G. and Leple, J.C. (2010) Wood formation in Angiosperms. *Comptes Rendus Biologies* 333: 325-334.

Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002) Capturing chromosome conformation. *Science* 295: 1306-1311.Demura, T. and Fukuda, H. (2007) Transcriptional regulation in wood formation. *Trends in Plant Science* 12: 64-70.

Demura, T., Tashiro, G., Horiguchi, G., Kishimoto, N., Kubo, M., Matsuoka, N., et al. (2002) Visualization by comprehensive microarray analysis of gene expression programs during transdifferentiation of mesophyll cells into xylem cells. *Proceedings of the National Academy of Sciences* 99: 15794-15799.

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

Dey, B., Thukral, S., Krishnan, S., Chakrobarty, M., Gupta, S., Manghani, C., et al. (2012) DNA-protein interactions: methods for detection and analysis. *Molecular and Cellular Biochemistry* 365: 279-299.

Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., et al. (2006) Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome research* 16: 1299-1309.

Dostie, J. and Bickmore, W.A. (2012) Chromosome organization in the nucleus – charting new territory across the Hi-Cs. *Current Opinion in Genetics & Development* 22: 125-131.

Du, H., Zhang, L., Liu, L., Tang, X.F., Yang, W.J., Wu, Y.M., et al. (2009) Biochemical and molecular characterization of the plant MYB transcription factor family. *Biochemistry (Mosc)* 74: 1-11.

Du, J. and Groover, A. (2010) Transcriptional regulation of secondary growth and wood formation. *Journal of Integrative Plant Biology* 52: 17-27.

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.

Duncan, I.W. (2002) Transvection effects in *Drosophila*. *Annual Review of Genetics* 36: 521-556.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., et al. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57-74.

Dyer, R.B. and Herzog, N.K. (1995) Immunodepletion EMSA: a novel method to identify proteins in a protein-DNA complex. *Nucleic Acids Research* 23: 3345-3346.

El-kereamy, A., Bi, Y.-M., Ranathunge, K., Beatty, P.H., Good, A.G. and Rothstein, S.J. (2012) The Rice R2R3-MYB Transcription Factor OsMYB55 Is involved in the tolerance to high temperature and modulates amino acid metabolism. *PLoS One* 7: e52030.

Emery, J.F., Floyd, S.K., Alvarez, J., Eshed, Y., Hawker, N.P., Izhaki, A., et al. (2003) Radial patterning of *Arabidopsis* shoots by class III HD-ZIP and KANADI genes. *Current Biology* 13: 1768-1774.

Ernst, H.A., Olsen, A.N., Skriver, K., Larsen, S. and Leggio, L.L. (2004) Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors. *EMBO reports* 5: 297-303.

Esau, K. (1960) Anatomy of seed plants. *Soil Science* 90: 149.

Eshed, Y., Baum, S.F., Perea, J.V. and Bowman, J.L. (2001) Establishment of polarity in lateral organs of plants. *Current Biology* 11: 1251-1260.

Feingold, E., Good, P., Guyer, M., Kamholz, S., Liefer, L., Wetterstrand, K., et al. (2004) The ENCODE (ENCyclopedia of DNA elements) project. *Science* 306: 636-640.

Felsenfeld, G. and Groudine, M. (2003) Controlling the double helix. *Nature* 421: 448-453.

Foulds, G.J. and Etzkorn, F.A. (1998) A capillary electrophoresis mobility shift assay for protein—DNA binding affinities free in solution. *Nucleic Acids Research* 26: 4304-4305.

Gallo, S.M., Li, L., Hu, Z. and Halfon, M.S. (2006) REDfly: a regulatory element database for *Drosophila*. *Bioinformatics* 22: 381-383.

Gaudinier, A., Zhang, L., Reece-Hoyes, J.S., Taylor-Teeples, M., Pu, L., Liu, Z., et al. (2011) Enhanced Y1H assays for *Arabidopsis*. *Nature methods* 8: 1053-1055.

Gill, G. and Ptashne, M. (1988) Negative effect of the transcriptional activator GAL4. *Nature* 334: 721-724.

Goicoechea, M., Lacombe, E., Legay, S., Mihaljevic, S., Rech, P., Jauneau, A., et al. (2005) EgMYB2, a new transcriptional activator from *Eucalyptus* xylem, regulates secondary cell wall formation and lignin biosynthesis. *The Plant Journal* 43: 553-567.

Grattapaglia, D. and Kirst, M. (2008) *Eucalyptus* applied genomics: from gene sequences to breeding tools. *New Phytologist* 179: 911-929.

Grattapaglia, D., Plomion, C., Kirst, M. and Sederoff, R.R. (2009) Genomics of growth traits in forest trees. *Current Opinion in Plant Biology* 12: 148-156.

Guerriero, G., Fugelstad, J. and Bulone, V. (2010) What do we really know about cellulose biosynthesis in higher plants? *Journal of Integrative Plant Biology* 52: 161-175.

GuhaThakurta, D. (2006) Computational identification of transcriptional regulatory elements in DNA sequence. *Nucleic Acids Research* 34: 3585-3598.

Guo, A., He, K., Liu, D., Bai, S., Gu, X., Wei, L., et al. (2005) DATF: a database of *Arabidopsis* transcription factors. *Bioinformatics* 21: 2568-2569.

Gutierrez, R.A., Green, P.J., Keegstra, K. and Ohlrogge, J.B. (2004) Phylogenetic profiling of the *Arabidopsis thaliana* proteome: what proteins distinguish plants from other organisms? *Genome Biology* 5: R53.

Hamann, T., Benkova, E., Baurle, I., Kientz, M. and Jurgens, G. (2002) The *Arabidopsis BODENLOS* gene encodes an auxin response protein inhibiting MONOPTEROS-mediated embryo patterning. *Genes & Development* 16: 1610-1615.

Hamann, T., Mayer, U. and Jurgens, G. (1999) The auxin-insensitive *bodenlos* mutation affects primary root formation and apical-basal patterning in the *Arabidopsis* embryo. *Development* 126: 1387-1395.

Hartley, J.L., Temple, G.F. and Brasch, M.A. (2000) DNA cloning using in vitro site-specific recombination. *Genome Research* 10: 1788-1795.

Hausner, W., Frey, G. and Thomm, M. (1991) Control regions of an archaeal gene: A TATA box and an initiator element promote cell-free transcription of the *tRNAVal* gene of *Methanococcus vannielii*. *Journal of Molecular Biology* 222: 495-508.

Heckman, D.S., Geiser, D.M., Eidell, B.R., Stauffer, R.L., Kardos, N.L. and Hedges, S.B. (2001) Molecular evidence for the early colonization of land by fungi and plants. *Science* 293: 1129-1133.

Hedges, S.B., Blair, J.E., Venturi, M.L. and Shoe, J.L. (2004) A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evolutionary Biology* 4: 2.

Hegedus, D., Yu, M., Baldwin, D., Gruber, M., Sharpe, A., Parkin, I., et al. (2003) Molecular characterization of *Brassica napus* NAC domain transcriptional activators induced in response to biotic and abiotic stress. *Plant Molecular Biology* 53: 383-397.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., et al. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics* 39: 311-318.

Hellman, L.M. and Fried, M.G. (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature Protocols* 2: 1849-1861.

Hermsen, R., Tans, S. and ten Wolde, P.R. (2006) Transcriptional regulation by competing transcription factor modules. *PLoS Computional Biology* 2: e164.

Herth, W. (1983) Arrays of plasma-membrane "rosettes" involved in cellulose microfibril formation of *Spirogyra*. *Planta* 159: 347-356.

Higo, K., Ugawa, Y., Iwamoto, M. and Korenaga, T. (1999) Plant *cis*-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Research* 27: 297-300.

Hillier, L.W., Marth, G.T., Quinlan, A.R., Dooling, D., Fewell, G., Barnett, D., et al. (2008) Whole-genome sequencing and variant discovery in *C. elegans*. *Nature methods* 5: 183-188.

Holland, N., Holland, D., Helentjaris, T., Dhugga, K.S., Xoconostle-Cazares, B. and Delmer, D.P. (2000) A comparative analysis of the plant *cellulose synthase* (*CesA*) gene family. *Plant Physiology* 123: 1313-1324.

Hu, J., Li, B. and Kihara, D. (2005) Limitations and potentials of current motif discovery algorithms. *Nucleic Acids Research* 33: 4899-4913.

Hussey, S., Mizrachi, E., Spokevicius, A., Bossinger, G., Berger, D. and Myburg, A. (2011) *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: 173.

Istrail, S. and Davidson, E.H. (2005) Logic functions of the genomic *cis*-regulatory code. *Proceedings of the National Academy of Sciences* 102: 4954-4959.

Jenuwein, T. and Allis, C.D. (2001) Translating the Histone Code. *Science* 293: 1074-1080.

Jin, H. and Martin, C. (1999) Multifunctionality and diversity within the plant MYB-gene family. *Plant Molecular Biology* 41: 577-585.

Jones-Rhoades, M.W., Bartel, D.P. and Bartel, B. (2006) MicroRNAs and their regulatory roles in plants. *Annual Review of Plant Biology* 57: 19-53.

Joshi, C.P. and Mansfield, S.D. (2007) The cellulose paradox--simple molecule, complex biosynthesis. *Current Opinion in Plant Biology* 10: 220-226.

Juven-Gershon, T., Hsu, J.Y., Theisen, J.W. and Kadonaga, J.T. (2008) The RNA polymerase II core promoter - the gateway to transcription. *Current Opinion in Cell Biology* 20: 253-259.

Juven-Gershon, T. and Kadonaga, J.T. (2010) Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Developmental Biology* 339: 225-229.

Kato, H., Motomura, T., Komeda, Y., Saito, T. and Kato, A. (2009) Overexpression of the NAC transcription factor family gene *ANAC036* results in a dwarf phenotype in *Arabidopsis thaliana*. *Journal of Plant Physiology*.

Kenrick, P. and Crane, P.R. (1997) The origin and early evolution of plants on land. *Nature* 389: 33-39.

Kerr, L.D. (1995) Electrophoretic mobility shift assay. *Methods in Enzymology* 254: 619-632.

Kim, J.Y., Park, O.G., Lee, J.W. and Lee, Y.C. (2007) One- plus two-hybrid system, a novel yeast genetic selection for specific missense mutations disrupting protein/protein interactions. *Molecular Cellular Proteomics* 6: 1727-1740.

Kim, W.C., Ko, J.H. and Han, K.H. (2012) Identification of a *cis*-acting regulatory motif recognized by MYB46, a master transcriptional regulator of secondary wall biosynthesis. *Plant Molecular Biology* 78: 489-501.

Kimura, S., Laosinchai, W., Itoh, T., Cui, X., Linder, C.R. and Brown, R.M., Jr. (1999) Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant *vigna angularis*. *Plant Cell* 11: 2075-2086.

Kiran, K., Ansari, S.A., Srivastava, R., Lodhi, N., Chaturvedi, C.P., Sawant, S.V., et al. (2006) The TATA-box sequence in the basal promoter contributes to determining light-dependent gene expression in plants. *Plant Physiology* 142: 364-376.

Ko, J.H., Kim, W.C. and Han, K.H. (2009) Ectopic expression of MYB46 identifies transcriptional regulatory genes involved in secondary wall biosynthesis in *Arabidopsis*. *The Plant Journal* 60: 649-665.

Ko, J.H., Yang, S.H., Park, A.H., Lerouxel, O. and Han, K.H. (2007) ANAC012, a member of the plant-specific NAC transcription factor family, negatively regulates xylary fiber development in *Arabidopsis thaliana*. *The Plant Journal* 50: 1035-1048.

Krivega, I. and Dean, A. (2012) Enhancer and promoter interactions—long distance calls. *Current Opinion in Genetics & Development* 22: 79-85.

Kubo, M., Udagawa, M., Nishikubo, N., Horiguchi, G., Yamaguchi, M., Ito, J., et al. (2005) Transcription switches for protoxylem and metaxylem vessel formation. *Genes & Development* 19: 1855-1860.

Kumar, M., Thammannagowda, S., Bulone, V., Chiang, V., Han, K.-H., Joshi, C.P., et al. (2009) An update on the nomenclature for the *cellulose synthase* genes in Populus. *Trends in Plant Science* 14: 248-254.

Legay, S., Lacombe, E., Goicoechea, M., Briere, C., Séguin, A., Mackay, J., et al. (2007) Molecular characterization of *EgMYB1*, a putative transcriptional repressor of the lignin biosynthetic pathway. *Plant Science* 173: 542-549.

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

Lelli, K.M., Slattery, M. and Mann, R.S. (2012) Disentangling the many layers of eukaryotic transcriptional regulation. *Annual Review of Genetics* 46: 43-68.

Li, E., Bhargava, A., Qiang, W., Friedmann, M.C., Forneris, N., Savidge, R.A., et al. (2012) The Class II KNOX gene KNAT7 negatively regulates secondary wall formation in *Arabidopsis* and is functionally conserved in *Populus*. *New Phytologist* 194: 102-115.

Li, E., Wang, S., Liu, Y., Chen, J.G. and Douglas, C.J. (2011) OVATE FAMILY PROTEIN4 (OFP4) interaction with KNAT7 regulates secondary cell wall formation in *Arabidopsis thaliana*. *The Plant Journal* 67: 328-341.

Li, G. and Reinberg, D. (2011) Chromatin higher-order structures and gene regulation. *Current Opinion in Genetics & Development* 21: 175-186.

Longabaugh, W.J. (2012) BioTapestry: a tool to visualize the dynamic properties of gene regulatory networks. *Methods in Molecular Biology* 876*:* 359-394.

Lopato, S., Bazanova, N., Morran, S., Milligan, A.S., Shirley, N. and Langridge, P. (2006) Isolation of plant transcription factors using a modified yeast one-hybrid system. *Plant Methods* 2: 3.

Lu, S.X., Knowles, S.M., Andronis, C., Ong, M.S. and Tobin, E.M. (2009) CIRCADIAN CLOCK ASSOCIATED1 and LATE ELONGATED HYPOCOTYL Function Synergistically in the Circadian Clock of *Arabidopsis*. *Plant Physiology* 150: 834-843.

Lucas, W.J., Groover, A., Lichtenberger, R., Furuta, K., Yadav, S.R., Helariutta, Y., et al. (2013) The Plant Vascular System: Evolution, Development and Functions. *Journal of Integrative Plant Biology*.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 389: 251-260.

MacIsaac, K.D. and Fraenkel, E. (2006) Practical strategies for discovering regulatory DNA sequence motifs. *PLoS computational biology* 2: e36.

Mallory, A.C., Dugas, D.V., Bartel, D.P. and Bartel, B. (2004a) MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs. *Current Biology* 14: 1035-1046.

Mallory, A.C., Reinhart, B.J., Jones-Rhoades, M.W., Tang, G., Zamore, P.D., Barton, M.K., et al. (2004b) MicroRNA control of PHABULOSA in leaf development: Importance of pairing to the microRNA 5' region. *EMBO JOURNALournal* 23: 3356-3364.

Matus, J., Aquea, F. and Arce-Johnson, P. (2008) Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across Vitis and *Arabidopsis* genomes. *BMC Plant Biology* 8: 83.

Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., et al. (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Research* 34: D108-110.

McCarthy, R.L., Zhong, R., Fowler, S., Lyskowski, D., Piyasena, H., Carleton, K., et al. (2010) The poplar MYB transcription factors, PtrMYB3 and PtrMYB20, are involved in the regulation of secondary wall biosynthesis. *Plant Cell Physiology* 51: 1084-1090.

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 Physiology* 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 Signaling & Behaviour* 6: 1282-1285.

McHale, N.A. and Koning, R.E. (2004) MicroRNA-directed cleavage of *Nicotiana sylvestris PHAVOLUTA* mRNA regulates the vascular cambium and structure of apical meristems. *Plant Cell* 16: 1730-1740.

Mellerowicz, E.J., Baucher, M., Sundberg, B. and Boerjan, W. (2001) Unravelling cell wall formation in the woody dicot stem. *Plant Molecular Biology* 47: 239-274.

Mellerowicz, E.J. and Sundberg, B. (2008) Wood cell walls: biosynthesis, developmental dynamics and their implications for wood properties. *Current Opinion in Plant Biology* 11: 293-300.

Miller, J.A. and Widom, J. (2003) Collaborative competition mechanism for gene activation *in vivo*. *Molecular and Cellular Biology* 23: 1623-1632.

Mitsuda, N., Ikeda, M., Takada, S., Takiguchi, Y., Kondou, Y., Yoshizumi, T., et al. (2010) Efficient yeast one-/two-hybrid screening using a library composed only of transcription factors in *Arabidopsis thaliana*. *Plant and cell physiology* 51: 2145-2151.

Mitsuda, N., Iwase, A., Yamamoto, H., Yoshida, M., Seki, M., Shinozaki, K., et al. (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.

Molina, C. and Grotewold, E. (2005) Genome wide analysis of *Arabidopsis* core promoters. *BMC Genomics* 6: 25.

Morishita, T., Kojima, Y., Maruta, T., Nishizawa-Yokoi, A., Yabuta, Y. and Shigeoka, S. (2009) *Arabidopsis* NAC transcription factor, ANAC078, regulates flavonoid biosynthesis under high-light. *Plant Cell Physiology* 50: 2210-2222.

Nakamura, M., Tsunoda, T. and Obokata, J. (2002) Photosynthesis nuclear genes generally lack TATA-boxes: a tobacco photosystem I gene responds to light through an initiator. *The Plant Journal* 29: 1-10.

Nakano, Y., Nishikubo, N., Goué, N., Ohtani, M., Yamaguchi, M., Katayama, Y., et al. (2010) MYB transcription factors orchestrating the developmental program of xylem vessels in *Arabidopsis* roots. *Plant Biotechnology* 27: 267-272.

Naz, A.A., Raman, S., Martinez, C.C., Sinha, N.R., Schmitz, G. and Theres, K. (2013) Trifoliate encodes a MYB transcription factor that modulates leaf and shoot architecture in tomato. *Proceedings of the National Academy of Sciences*.

O'Connor, T.R., Dyreson, C. and Wyrick, J.J. (2005) Athena: a resource for rapid visualization and systematic analysis of *Arabidopsis* promoter sequences. *Bioinformatics* 21: 4411-4413.

Ogata, K., Kanei-Ishii, C., Sasaki, M., Hatanaka, H., Nagadoi, A., Enari, M., et al. (1996) The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. *Nature Structural and Molecular Biology* 3: 178-187.

Ogata, K., Morikawa, S., Nakamura, H., Hojo, H., Yoshimura, S., Zhang, R., et al. (1995) Comparison of the free and DNA-complexed forms of the DNA-binding domain from c-Myb. *Nature Structural and Molecular Biology* 2: 309-320.

Ohashi-Ito, K., Oda, Y. and Fukuda, H. (2010) *Arabidopsis* VASCULAR-RELATED NAC-DOMAIN6 directly regulates the genes that govern programmed cell death and secondary wall formation during xylem differentiation. *Plant Cell* 22: 3461-3473.

Öhman, D., Demedts, B., Kumar, M., Gerber, L., Gorzsás, A., Goeminne, G., et al. (2013) MYB103 is required for FERULATE-5-HYDROXYLASE expression and syringyl lignin biosynthesis in *Arabidopsis* stems. *The The Plant Journal* 73: 63-76.

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.

Olsen, A.N., Ernst, H.A., Lo Leggio, L., Johansson, E., Larsen, S. and Skriver, K. (2004) Preliminary crystallographic analysis of the NAC domain of ANAC, a member of the plant-specific NAC transcription factor family. *Acts Crystallographica Section D: Biological Crystallography* 60: 112-115.

Ooka, H., Satoh, K., Doi, K., Nagata, T., Otomo, Y., Murakami, K., et al. (2003) Comprehensive analysis of NAC family genes in Oryza sativa and *Arabidopsis thaliana*. *DNA Research* 10: 239-247.

Ouwerkerk, P.B.F. and Meijer, A.H. (2001) Yeast One-Hybrid Screening for DNA-Protein Interactions. In *Current Protocols in Molecular Biology* 12. John Wiley & Sons, Inc.

Palaniswamy, S.K., James, S., Sun, H., Lamb, R.S., Davuluri, R.V. and Grotewold, E. (2006) AGRIS and AtRegNet. a platform to link *cis*-regulatory elements and transcription factors into regulatory networks. *Plant Physiology* 140: 818-829.

Pan, Y. and Nussinov, R. (2011) The Role of Response Elements Organization in Transcription Factor Selectivity: The IFN-β Enhanceosome Example. *PLoS Computional Biology* 7: e1002077.

Panne, D. (2008) The enhanceosome. *Current Opinion in Structural Biology* 18: 236-242.

Panne, D., Maniatis, T. and Harrison, S.C. (2004) Crystal structure of ATF-2/c-Jun and IRF-3 bound to the interferon-[beta] enhancer. *EMBO JOURNAL* 23: 4384-4393.

Park, P.J. (2009) ChIP-seq: Advantages and challenges of a maturing technology. *Nature Reviews Genetics* 10: 669-680.

Pauler, F.M., Sloane, M.A., Huang, R., Regha, K., Koerner, M.V., Tamir, I., et al. (2009) H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. *Genome research* 19: 221-233.

Pavesi, G., Mereghetti, P., Mauri, G. and Pesole, G. (2004) Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Research* 32: W199-203.

Pavesi, G., Zambelli, F. and Pesole, G. (2007) WeederH: an algorithm for finding conserved regulatory motifs and regions in homologous sequences. *BMC Bioinformatics* 8: 46.

Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P.A. and Saedler, H. (1987) The regulatory *c1* locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *EMBO Journal* 6: 3553-3558.

Peng, H., Cheng, H.Y., Chen, C., Yu, X.W., Yang, J.N., Gao, W.R., et al. (2009) A NAC transcription factor gene of Chickpea (*Cicer arietinum*), *CarNAC3*, is involved in drought stress response and various developmental processes. *Journal of Plant Physiology* 166: 1934-1945.

Persson, S., Wei, H., Milne, J., Page, G.P. and Somerville, C.R. (2005) Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. *Proceedings of the National Academy of Sciences of the United States of America* 102: 8633-8638.

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

Ponting, C.P., Oliver, P.L. and Reik, W. (2009) Evolution and Functions of Long Noncoding RNAs. *Cell* 136: 629-641.

Praz, V., Périer, R., Bonnard, C. and Bucher, P. (2002) The Eukaryotic Promoter Database, EPD: new entry types and links to gene expression data. *Nucleic Acids Research* 30: 322-324.

Priest, H.D., Filichkin, S.A. and Mockler, T.C. (2009) *Cis*-regulatory elements in plant cell signaling. *Current Opinion in Plant Biology* 12: 643-649.

Prigge, M.J., Otsuga, D., Alonso, J.M., Ecker, J.R., Drews, G.N. and Clark, S.E. (2005) Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in *Arabidopsis* development. *Plant Cell* 17: 61-76.

Priya, P. and Jain, M. (2013) RiceSRTFDB: A database of rice transcription factors containing comprehensive expression, *cis*-regulatory element and mutant information to facilitate gene function analysis. *Database: the journal of biological databases and curation* 2013.

Proost, S., Pattyn, P., Gerats, T. and Van de Peer, Y. (2011) Journey through the past: 150 million years of plant genome evolution. *The Plant Journal* 66: 58-65.

Pruneda-Paz, J.L., Breton, G., Para, A. and Kay, S.A. (2009) A functional genomics approach reveals CHE as a component of the Arabidopsis circadian clock. *Science* 323: 1481-1485.

Pruss, D. and Bavykin, S.G. (1997) Chromatin studies by DNA-protein cross-linking. *Methods: A Companion to Methods in Enzymology* 12: 36-47.

Quail, M.A., Kozarewa, I., Smith, F., Scally, A., Stephens, P.J., Durbin, R., et al. (2008) A large genome center's improvements to the Illumina sequencing system. *Nature methods* 5: 1005-1010.

Ramirez, S.R. and Basu, C. (2009) Comparative analyses of plant transcription factor databases. *Current genomics* 10: 10.

Rando, O.J. and Chang, H.Y. (2009) Genome-wide views of chromatin structure. *Annual Review of Biochemistry* 78: 245-271.

Ranik, M. and Myburg, A.A. (2006) Six new *cellulose synthase* genes from *Eucalyptus* are associated with primary and secondary cell wall biosynthesis. *Tree Physiol* 26: 545-556.

Raven, J. and Edwards, D. (2001) Roots: evolutionary origins and biogeochemical significance. *Journal of experimental botany* 52: 381-401.

Razin, S.V., Gavrilov, A.A., Pichugin, A., Lipinski, M., Iarovaia, O.V. and Vassetzky, Y.S. (2011) Transcription factories in the context of the nuclear and genome organization. *Nucleic Acids Research* 39: 9085-9092.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., et al. (2000) Genome-wide location and function of DNA binding proteins. *Science* 290: 2306-2309.

Richmond, T. (2000) Higher plant cellulose synthases. *Genome Biology* 1: REVIEWS3001.

Riechmann, J., Heard, J., Martin, G., Reuber, L., Keddie, J., Adam, L., et al. (2000) *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290: 2105-2110.

Roeder, R.G. (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* 21: 327-335.

Rombauts, S., Dehais, P., Van Montagu, M. and Rouze, P. (1999) PlantCARE, a plant cis-acting regulatory element database. *Nucleic Acids Research* 27: 295-296.

Scarpella, E. and Meijer, A.H. (2004) Pattern formation in the vascular system of monocot and dicot plant species. *New Phytologist* 164: 209-242.

Schlereth, A., Möller, B., Liu, W., Kientz, M., Flipse, J., Rademacher, E.H., et al. (2010) MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. *Nature* 464: 913-916.

Schuetz, M., Smith, R. and Ellis, B. (2013) Xylem tissue specification, patterning, and differentiation mechanisms. *Journal of experimental botany* 64: 11-31.

Semenza, G.L. and Wang, G.L. (1992) A nuclear factor induced by hypoxia via *de novo* protein synthesis binds to the human erythropoietin gene enhancer at a site required for transcriptional activation. *Molecular and Cellular Biology* 12: 5447-5454.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., et al. (2012) Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* 148: 458-472.

Shahmuradov, I.A., Gammerman, A.J., Hancock, J.M., Bramley, P.M. and Solovyev, V.V. (2003) PlantProm: a database of plant promoter sequences. *Nucleic Acids Research* 31: 114-117.

Sharov, A.A., Dudekula, D.B. and Ko, M.S. (2006) CisView: a browser and database of *cis*-regulatory modules predicted in the mouse genome. *DNA Research* 13: 123-134.

Shen, H., Yin, Y., Chen, F., Xu, Y. and Dixon, R. (2009) A bioinformatic analysis of NAC genes for plant cell wall development in relation to lignocellulosic bioenergy production. *BioEnergy Research* 2: 217-232.

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., et al. (2011) Cofactor Binding Evokes Latent Differences in DNA Binding Specificity between Hox Proteins. *Cell* 147: 1270-1282.

Smale, S.T. and Baltimore, D. (1989) The "initiator" as a transcription control element. *Cell* 57: 103-113.

Solomon, M.J., Larsen, P.L. and Varshavsky, A. (1988) Mapping protein-DNA interactions *in vivo* with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell* 53: 937-947.

Somerville, C. (2006) Cellulose synthesis in higher plants. *Annual Review of Cell and Developmental Biology* 22: 53-78.

Souer, E., van Houwelingen, A., Kloos, D., Mol, J. and Koes, R. (1996) The *no apical meristem* gene of *Petunia* is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. *Cell* 85: 159-170.

Sperry, J.S. (2003) Evolution of water transport and xylem structure. *International Journal of Plant Sciences* 164: S115-S127.

Spicer, R. and Groover, A. (2010). "Evolution of development of vascular cambia and secondary growth." *New Phytologist* 186: 577-592.

Steffens, N.O., Galuschka, C., Schindler, M., Bulow, L. and Hehl, R. (2005) AthaMap web tools for database-assisted identification of combinatorial *cis*-regulatory elements and the display of highly conserved transcription factor binding sites in *Arabidopsis thaliana*. *Nucleic Acids Research* 33: W397-402.

Stracke, R., Werber, M. and Weisshaar, B. (2001) The R2R3-MYB gene family in *Arabidopsis thaliana*. *Current Opinion in Plant Biology* 4: 447-456.

Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T.Z., Garcia-Hernandez, M., Foerster, H., et al. (2008) The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Research* 36: D1009-1014.

Taylor, N.G. (2008) Cellulose biosynthesis and deposition in higher plants. *New Phytologist* 178: 239-252.

Taylor, N.G., Howells, R.M., Huttly, A.K., Vickers, K. and Turner, S.R. (2003) Interactions among three distinct CesA proteins essential for cellulose synthesis. *Proceedings of the National Academy of Sciences* 100: 1450-1455.

Thanos, D. and Maniatis, T. (1995) Virus induction of human *IFN beta* gene expression requires the assembly of an enhanceosome. *Cell* 83: 1091-1100.

Thijs, G., Lescot, M., Marchal, K., Rombauts, S., De Moor, B., Rouze, P., et al. (2001) A higher-order background model improves the detection of promoter regulatory elements by Gibbs sampling. *Bioinformatics* 17: 1113-1122.

Thijs, G., Marchal, K., Lescot, M., Rombauts, S., De Moor, B., Rouze, P., et al. (2002) A Gibbs sampling method to detect overrepresented motifs in the upstream regions of coexpressed genes. *Journal of Computational Biology* 9: 447-464.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., et al. (2012) The accessible chromatin landscape of the human genome. *Nature* 489: 75-82.

Timell, T. (1986) Chemical properties of compression wood. *Compression wood in gymnosperms* 1: 289-408.

Tominaga-Wada, R., Nukumizu, Y., Sato, S. and Wada, T. (2013) Control of plant trichome and root-hair development by a tomato *Solanum lycopersicum* R3 MYB transcription factor. *PLoS One* 8: e54019.

Tompa, M., Li, N., Bailey, T.L., Church, G.M., De Moor, B., Eskin, E., et al. (2005) Assessing computational tools for the discovery of transcription factor binding sites. *Nature Biotechnology* 23: 137-144.

Vandepoele, K., Quimbaya, M., Casneuf, T., De Veylder, L. and Van de Peer, Y. (2009) Unraveling transcriptional control in *Arabidopsis* using *cis*-regulatory elements and coexpression networks. *Plant Physiology* 150: 535-546.

Vidal, M., Brachmann, R.K., Fattaey, A., Harlow, E. and Boeke, J.D. (1996) Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. *Proceedings of the National Academy of Sciences* 93: 10315-10320.

Vidalain, P.-O., Boxem, M., Ge, H., Li, S. and Vidal, M. (2004) Increasing specificity in high-throughput yeast two-hybrid experiments. *Methods (San Diego, Calif.)* 32: 363.

Wang, D., Guo, Y., Wu, C., Yang, G., Li, Y. and Zheng, C. (2008) Genome-wide analysis of the CCCH zinc finger family in *Arabidopsis* and rice. *BMC Genomics* 9: 44.

Wanke, D. and Harter, K. (2009) Analysis of plant regulatory DNA sequences by the yeast-one-hybrid assay. *Methods in Molecular Biology* 479: 291-309.

Wehner, N., Hartmann, L., Ehlert, A., Böttner, S., Oñate-Sánchez, L. and Dröge-Laser, W. (2011) High-throughput protoplast transactivation (PTA) system for the analysis of Arabidopsis transcription factor function. *The Plant Journal* 68: 560-569.

Wei, Z., Angerer, R.C. and Angerer, L.M. (1999) Identification of a new sea urchin ets protein, SpEts4, by yeast one-hybrid screening with the hatching enzyme promoter. *Molecular and Cellular Biology* 19: 1271-1278.

Weston, K. (1998) Myb proteins in life, death and differentiation. *Current Opinion in Genetics & Development* 8: 76-81.

Wilkins, O., Nahal, H., Foong, J., Provart, N.J. and Campbell, M.M. (2009) Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. *Plant Physiology* 149: 981-993.

Winzell, A., Aspeborg, H., Wang, Y. and Ezcurra, I. (2010) Conserved CA-rich motifs in gene promoters of Pt×tMYB021-responsive secondary cell wall carbohydrate-active enzymes in *Populus*. *Biochemical and Biophysical Research Communications* 394: 848-853.

Xian, J., Harrington, M.G. and Davidson, E.H. (1996) DNA-protein binding assays from a single sea urchin egg: a high-sensitivity capillary electrophoresis method. *Proceedings of the National Academy of Sciences* 93: 86-90.

Xie, Q., Frugis, G., Colgan, D. and Chua, N.H. (2000) *Arabidopsis* NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. *Genes & Development* 14: 3024-3036.

Yamaguchi, M. and Demura, T. (2010) Transcriptional regulation of secondary wall formation controlled by NAC domain proteins. *Plant Biotechnology* 27: 237-242.

Yamaguchi, M., Goue, N., Igarashi, H., Ohtani, M., Nakano, Y., Mortimer, J.C., et al. (2010) VASCULAR-RELATED NAC-DOMAIN6 and VASCULAR-RELATED NAC-DOMAIN7 effectively induce transdifferentiation into xylem vessel elements under control of an induction system. *Plant Physiology* 153: 906-914.

Yamaguchi, M., Mitsuda, N., Ohtani, M., Ohme-Takagi, M., Kato, K. and Demura, T. (2011) VASCULAR-RELATED NAC-DOMAIN 7 directly regulates the expression of a broad range of genes for xylem vessel formation. *The The Plant Journal* 66: 579-590.

Yamamoto, Y.Y., Ichida, H., Abe, T., Suzuki, Y., Sugano, S. and Obokata, J. (2007) Differentiation of core promoter architecture between plants and mammals revealed by LDSS analysis. *Nucleic Acids Research* 35: 6219-6226.

Yamamoto, Y.Y. and Obokata, J. (2008) ppdb: a plant promoter database. *Nucleic Acids Research* 36: D977-981.

Yang, J., Chen, F., Yu, O. and Beachy, R.N. (2011) Controlled silencing of 4-coumarate:CoA ligase alters lignocellulose composition without affecting stem growth. *Plant Physiology and Biochemistry* 49: 103-109.

Yanhui, C., Xiaoyuan, Y., Kun, H., Meihua, L., Jigang, L., Zhaofeng, G., 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.

Ye, Z.-H. (2002) Vascular tissue differentiation and pattern formation in plants. *Annual Review of Plant Biology* 53: 183-202.

Zeller, G., Henz, S.R., Widmer, C.K., Sachsenberg, T., Rätsch, G., Weigel, D., et al. (2009) Stress-induced changes in the *Arabidopsis thaliana* transcriptome analyzed using whole-genome tiling arrays. *The The Plant Journal* 58: 1068-1082.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biology* 9.

Zhao, C., Avci, U., Grant, E.H., Haigler, C.H. and Beers, E.P. (2008) XND1, a member of the NAC domain family in *Arabidopsis thaliana*, negatively regulates lignocellulose synthesis and programmed cell death in xylem. *The Plant Journal* 53: 425-436.

Zhao, C., Craig, J.C., Petzold, H.E., Dickerman, A.W. and Beers, E.P. (2005a) The xylem and phloem transcriptomes from secondary tissues of the *Arabidopsis* root-hypocotyl. *Plant Physiology* 138: 803-818.

Zhao, F., Xuan, Z., Liu, L. and Zhang, M.Q. (2005b) TRED: a Transcriptional Regulatory Element Database and a platform for *in silico* gene regulation studies. *Nucleic Acids Research* 33: D103-D107.

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. (2011a) Transcriptional activation of secondary wall biosynthesis by rice and maize NAC and MYB transcription factors. *Plant Cell Physiology* 52: 1856-1871.

Zhong, R., Lee, C. and Ye, Z.-H. (2010a) Functional characterization of poplar wood-associated NAC domain transcription factors. *Plant Physiology* 152: 1044-1055.

Zhong, R., Lee, C. and Ye, Z.H. (2010b) Evolutionary conservation of the transcriptional network regulating secondary cell wall biosynthesis. *Trends in Plant Science* 15: 625-632.

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., Lee, C. and Ye, Z.-H. (2011) Dissection of the transcriptional program regulating secondary wall biosynthesis during wood formation in poplar. *Plant Physiology* 157: 1452-1468.

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., Taylor, J.J. and Ye, Z.H. (1997) Disruption of interfascicular fiber differentiation in an *Arabidopsis* mutant. *The Plant Cell Online* 9: 2159-2170.

Zhong, R. and Ye, Z.-H. (2010) The poplar PtrWNDs are transcriptional activators of secondary cell wall biosynthesis. *Plant Signaling & Behavior* 5: 469-472.

Zhong, R. and Ye, Z.H. (2004) Amphivasal vascular bundle 1, a gain-of-function mutation of the *IFL1/REV* gene, is associated with alterations in the polarity of leaves, stems and carpels. *Plant Cell Physiology* 45: 369-385.

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 Physiology* 53: 368-380.

Zorn, C., Cremer, C., Cremer, T. and Zimmer, J. (1979) Unscheduled DNA synthesis after partial UV irradiation of the cell nucleus: Distribution in interphase and metaphase. *Experimental Cell Research* 124: 111-119.

CHAPTER 2

Examining the expression pattern and interactions of the *Eucalyptus grandis SND2* 1.5 kb promoter fragment in the transcriptional network of secondary cell wall biosynthesis

J. Botha[1,2], C. Maritz-Olivier[1], E. Mizrachi[1,2], A.A. Myburg[1,2]

[1]Department of Genetics, [2]Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private bag x20, Pretoria, 0028, South Africa

This chapter has been prepared in the format of a manuscript for the research journal *Plant Cell Physiology* (PCP). I generated the transgenic plants for GUS analysis, both of the bait constructs and most of the prey constructs used in the yeast one-hybrid analysis. I performed all the GUS and yeast one-hybrid analyses. I also performed all of the data analyses and prepared this manuscript. Prof A.A. Myburg, Dr C. Maritz-Olivier and Mr E. Mizrachi provided advice, direction and supervision in the planning of the project. They also provided direction in the interpretation of the results and provided critical revision of the manuscript. All other technical assistance are acknowledged at the end of the Chapter.

2.1 Abstract

The secondary cell walls (SCWs) of cells in woody tissues in trees are a tremendous source of biomass, rich in cellulose, lignin and xylan. At SND2, a SCW-related NAC transcription factor (TF), has been identified as a possible regulator of cellulose biosynthesis in the SCWs of *Arabidopsis* interfascicular fibres. Within *Eucalyptus grandis*, a widely grown hardwood plantation species, the transcriptional network regulating SCW formation is not well characterised. In this study, we investigate the heterologous expression pattern of a 1.5 kb *E. grandis SND2* promoter fragment in *Arabidopsis*, and identify *Eucalyptus* TFs important for SCW biosynthesis that bind to it. A qualitative GUS-analysis was used to determine in which *Arabidopsis* tissues the promoter fragment was active, and a targeted yeast one-hybrid (Y1-H) analysis was performed on the 1.5 kb fragment and a 500 bp truncation to identify TFs important for SCW biosynthesis that bind to them. GUS expression of the *EgrSND2* promoter fragment was very strong in many tissues, indicating powerful promoter activity. Y1-H analyses showed that EgrMYB46 (Eucgr.B03684), EgrMYB83 (Eucgr.G03385), EgrSND2 (Eucgr.K01061), EgrSND3 (Eucgr.E03226), EgrVND6 (Eucgr.A02887), EgrZF1 (Eucgr.F02796) and EgrZF2 (Eucgr.B00047) interacted with the *EgrSND2* promoter fragment, indicating a possible role for SND2 in cell types other than interfascicular fibres. In this study, we determine that *EgrSND2* is not exclusively expressed in *Arabidopsis* interfascicular fibres, and elucidate new connections in the SCW transcriptional regulatory network. These results have added to the understanding of the *Eucalyptus* SCW regulatory network and will form the basis for further studies on the regulation of biomass development in this important plantation species.

Keywords: *Eucalyptus grandis*, SND2, cellulose biosynthesis, secondary cell wall, tree, regulation

2.2 Introduction

Secondary cell walls (SCWs) make up a large portion of plant biomass and therefore constitute a large part of the biomass on the planet. Xylem tissues contain fibre cells, which have characteristically thick SCWs (reviewed in Gorshkova et al. 2012; Schuetz et al. 2013). The biosynthesis of xylem fibre SCWs is subject to transcriptional control by a network of transcription factors (TFs) working in a cascade-like fashion (Zhong et al. 2006). They are important for many different applications, including those in the paper and chemical cellulose industries. Xylem fibres are rich in cellulose and xylan, and have relatively low levels of lignin when compared to other xylem cell types (Plomion et al. 2001). As such, they are worthy of study, seeing as plant lignocellulosic biomass is becoming an increasingly important renewable alternative to fossil fuels (Ohlrogge et al. 2009).. *Eucalyptus spp* are an economically important hardwood crop that can provide large amounts of lignocellulosic biomass in the form of xylem fibres (Grattapaglia and Kirst 2008; Grattapaglia et al. 2009). It is important to study the transcriptional regulation of SCW biosynthesis in *E. grandis* wood fibres, as in the future this may allow us to produce wood fibres with improved cellulose and lower lignin content, which we can use for paper, chemical cellulose and biofuels.

Wood formation is a complex developmental process which begins at the vascular cambium, the secondary meristem responsible for production of the vascular tissues. Through a number of complex transcriptional interactions, the cambial initial cells give rise to the xylem and phloem mother cells, and then eventually to mature xylem and phloem. This process has been extensively reviewed (Mellerowicz et al. 2001; Plomion et al. 2001; Mellerowicz and Sundberg 2008; Lucas et al. 2013; Schuetz et al. 2013). A xylem cell undergoes a number of steps on its way to maturation. In several of these steps, but specifically the cell wall thickening step, the SCW is deposited. The SCW is located on the inside of the compound middle lamella (CML), has three layers designated S1-S3 (Timell 1986) and is composed of mainly cellulose, hemicellulose and lignin (reviewed in Mellerowicz et al. 2001).

Xylem itself consists of mainly fibres, vessels and xylem parenchyma cells in rays (Demura and Fukuda 2007). The fibres are of particular interest, as their SCWs contain more cellulose and hemicellulose and less lignin than xylem vessels (Mellerowicz and Sundberg 2008). This is due to their role in the mechanical strength of the plant stem, as opposed to that of water transport for the xylem vessels (Schuetz et al. 2013).

SCW development in xylem is controlled by a hierarchy of TFs which function in a cascade of transcriptional regulation. This mode of regulation is characteristic of feed-forward loops (Alon 2007), where gene X directly interacts with gene Y, gene Y directly interacts with gene Z and gene X directly interacts with gene Z, and which are prominent in SCW biosynthesis (McCarthy et al. 2009; Winzell et al. 2010). At the top of this hierarchy are master regulator TFs which are necessary and sufficient to activate entire biosynthetic programs in individual cell types (Zhong et al. 2006; Yamaguchi et al. 2010). AtNST1 and AtNST3/AtSND1/ANAC012 are master regulators of SCW biosynthesis in xylem fibre cells (Zhong et al. 2006), while AtVND6 and AtVND7 are master regulators of protoxylem and metaxylem vessels respectively (Kubo et al. 2005). The SNBE *cis*-element has been identified as a possible binding site for the master regulators AtSND1, AtVND6 and AtVND7 and has been found in a number of their targets (Zhong et al. 2010b; McCarthy et al. 2011). In *Eucalyptus gunnii*, EgMYB1 and EgMYB2 have been identified as potential master regulators of SCW biosynthesis. EgMYB1 has been associated with repression of SCW biosynthesis (Legay et al. 2007; Legay et al. 2010), while EgMYB2 has been shown to be an activator of SCW biosynthesis (Goicoechea et al. 2005; De Micco et al. 2012).

In a study on AtSND1, a number of TFs were identified to be highly induced by AtSND1, and were shown to be important for SCW biosynthesis (Zhong et al. 2008). Of these, AtMYB46 and AtMYB83 are considered functionally redundant mid-level master regulators, able to activate biosynthetic pathways of lignin, cellulose and hemicellulose (Zhong et al. 2007a; McCarthy et al. 2009) by recognizing and binding to MYB46RE (Kim et al. 2012) or SMRE

(Zhong and Ye 2012) *cis*-elements. AtMYB52 and AtMYB54 are highly induced by, but not direct targets of, AtSND1 (Zhong et al. 2008). They are both expressed in xylem vessel and fibre cells, though to slightly different levels in each (Zhong et al. 2008), and even though they are able to induce expression of genes in all three biosynthetic pathways, they have been specifically associated with lignin biosynthesis (Zhong et al. 2008; Shen et al. 2009; Nakano et al. 2010; Cassan-Wang et al. 2013). Another two TFs, AtMYB85 and AtKNAT7 have also been implicated in lignin biosynthesis. AtMYB85 is able to activate the promoter of *4CL1,* a core lignin biosynthetic gene, and constitutive ectopic overexpression of AtMYB85 leads to ectopic deposition of lignin in epidermal and cortical cells (Zhong et al. 2008). On the other hand, overexpression of AtKNAT7 leads to thinner interfascicular fibre cell walls, while knockout leads to thicker fibre cell walls with higher lignin content and it is thus thought to be a repressor of lignin biosynthesis (Li et al. 2012).

Of the TFs identified by Zhong et al., (2008), AtSND2, AtSND3 and AtMYB103 were able to induce expression of the SCW related *AtCesA8* promoter and showed high specificity for interfascicular fibres. Recently, AtMYB103 has been shown to also activate biosynthesis of S-lignin (Ohman et al. 2012) and has been shown to be active in cell types other than interfascicular fibres (Nakano et al. 2010; Ohashi-Ito et al. 2010; Yamaguchi et al. 2010). Of the remaining two TFs, AtSND3 is a direct target of AtSND1, while AtSND2 is not (Zhong et al. 2008). Also, it is thought that AtSND2 regulates polysaccharide biosynthetic genes, but not monolignol biosynthetic genes (Zhong et al. 2008; Hussey et al. 2011). This indicates that AtSND2 may be regulated in a more complex fashion than the typical feed-forward loops seen in transcriptional regulation of SCW biosynthesis (McCarthy et al. 2009; Winzell et al. 2010).

In this study, we aimed to determine whether the cloned fragment of *EgrSND2* promoter was able to modulate reporter gene expression in *Arabidopsis*, and to identify which *Eucalyptus* TFs expressed in the native xylem bind to the *EgrSND2* promoter fragment. To achieve this,

we employed a qualitative *GUS* reporter gene analysis of the *EgrSND2* promoter in *Arabidopsis.* A panel of fifteen *Eucalyptus* homologs of the SCW-associated TFs were identified, isolated and cloned into yeast expression vectors. The panel of TFs where used in a yeast one-hybrid (Y1-H) screen against a 1.5 kb *EgrSND2* promoter fragment and a 500 bp truncation of the promoter fragment. We were able to determine that the *EgrSND2* 1.5 kb promoter fragment is a strong transcriptional activator in *Arabidopsis,* is strongly expressed in vascular tissues and interfascicular fibres but does not show an interfascicular fibre specific expression pattern in *Arabidopsis*. *E. grandis* orthologs of a number of TFs important for SCW biosynthesis including AtMYB46, AtMYB83 and AtVND6 bind to the *EgrSND2* promoter fragment *in vivo* during Y1-H analysis and may modulate EgrSND2 gene expression.

2.3 Materials and Methods

2.3.1 Transgenic *Arabidopsis* plants

EgrSND2_1.5kb promoter::GUS reporter vectors were generated by cloning the previously isolated promoter sequences into the pMDC162 Gateway™ vector (Curtis and Grossniklaus 2003), which contains a β-Glucuronidase (*GUS*) reporter gene (Jefferson et al. 1987). Fifty *Arabidopsis thaliana* Col-0 plants were grown under long (16 h) day conditions, with inflorescence stems cut back several times to encourage increased boulting, before transforming them using *Agrobacterium tumefaciens* via a dipping method (Clough and Bent 1998). Plants were grown until they produced seed (approximately six weeks); seeds were collected from individual plants (T_0) and bulked into one stock to represent the transformation of the *EgrSND2_1.5kb promoter::GUS* construct. Seed from all fifty plants was sieved to remove large debris.

T_1 seeds were planted on agar medium containing hygromycin (20 mg/ml) and cefatoxamine (100 mg/ml). Seedlings were grown on the selective plates at 22^0C under long (16 h) day conditions. After two weeks the seedlings that were not showing stunted growth were transferred to Jiffies™ (Jiffy Products International, Norway) and were grown at 22^0C under 16 h day conditions for six weeks. T_2 seeds were collected separately for each plant to preserve the transgenic events represented by the T_1 plants. Transgenic plants containing *EgrSND3_1.5kb promoter::GUS* reporter vectors were obtained in an identical manner.

2.3.2 β-glucuronidase qualitative assays

Qualitative *GUS* assays were performed at one, three and six weeks. T_2 seeds were sowed on to agar plates containing cefatoxamine (100 mg/ml) and hygromycin (20 mg/ml), and grown for 1 week at 22^0C under long (16 h) day conditions. Fifteen T_2 seedlings of each transformation event were planted separately into Jiffies and at least three one week-old plants of each transformation event were submerged in *GUS* buffer (200 mM Na_2HPO_4, 0.5 mM potassium-ferrocyanide ($K_4[Fe(CN)_6]•3H_2O$), 0.5 mM potassium-ferricyanide ($K_3[Fe(CN)_6]$), 2 mM X-gluc ($C_{14}H_{13}BrClNO_7•C_6H_{13}N$), 0.1 % Triton X-100, 20 % methanol; Van Beveren et al. 2006). Seedlings in *GUS* buffer were incubated at 55^0C for 10 minutes before being transferred to a shaking incubator and incubated at 37^0C for 24 hours while shaking at 250 rpm. After incubation, plants were de-stained by placing them in 100% ethanol for 15 minutes, then the ethanol was replaced until all chlorophyll was removed from the samples. After de-staining, the seedlings were stored in 70% ethanol. Three of the remaining T_2 plants of each transformation event were selected at three weeks for *GUS* staining. The protocol used was similar to that used for the one week seedlings. At six weeks, three more replicates of each T_2 line were selected for *GUS* analysis. For the six week *GUS* analysis the leaves (L), upper inflorescence stem (US), lower inflorescence stem (LS), roots and hypocotyl (RH), flowers and siliques (FS) and whole stems (WS) were analysed separately. A pMDC162 *1.5kb promoter::GUS* construct for *EgrSND3* was also

generated. An identical *GUS* analysis was also performed on this construct concurrent to the *EgrSND2* analysis for comparison.

In order to determine the specificity of expression of *AtSND2* and *EgrSND2*, expression data across various tissues was used. Absolute expression data for *AtSND2* in the inflorescence stem, flowers, leaf, roots, shoot tip and hypocotyl was extracted from the Genevestigator arrays (Zimmermann et al. 2004). *E. grandis* RNA-seq data (raw FPKMs; Hefer et al., in preparation) for *EgrSND2* in the immature xylem, flowers, mature leaf, phloem, roots, shoot tips and young leaf was obtained. Expression data for separate xylem and phloem tissues was available from Genevestigator. However, it was not used as it consisted of only three replicates each, could not be found in a peer reviewed publication and showed expression profiles in direct conflict with previously published data (Zhong et al. 2008). Therefore the hypocotyl and inflorescence stem were used as more robust representations of vascular tissue expression due to their increased replicates (twelve and forty five in Genevestigator, respectively) and use in peer reviewed literature (Bergmann et al. 2004; Nagpal et al. 2005; Le et al. 2010; Chen et al. 2012). From this data, heatmaps were generated showing tissue specificity and xylem to leaf expression ratios were calculated. For *AtSND2*, the sum of absolute expression across all tissues analysed was calculated and used as the total. The xylem to leaf ratio (X:L) was calculated by dividing the average percentage expression across both inflorescence stems and hypocotyl by the percentage expression from leafs. For *EgrSND2*, the mean expression across three replicates was calculated in each tissue and the sum of the calculated means was used as the total. The X:L was calculated by dividing the percentage of transcripts from immature xylem by the average of the percentage of transcripts from both mature leaf and young leaf. In the heatmaps, the expression level of *AtSND2* and *EgrSND2* in a particular tissue were represented as a percentage of the total. All unprocessed expression data for both *AtSND2* and *EgrSND2* can be viewed in Electronic Supplementary File 1 (Please see attached CD).

The presence of the promoter and the *GUS* reporter gene were confirmed in the transgenic plants using PCR screening. Leaf samples were taken from T_2 plants for DNA extraction and PCR screening, T_2 plants were also used for seed collection. Genomic DNA was extracted from the leaves of transgenic plants using the Qiagen Plant II Extract kit (Qiagen, Valencia, CA, USA). A combination of a promoter-specific forward primer and *GUS*-specific reverse primer was used (Supplementary Table 1). The PCR was performed as follows: an initial denaturation for 3 min at 94^0C was performed, then 30 cycles of the following program; 30 s at 94^0C, 30 s at 58^0C and 2 min at 72^0C. A final elongation step of 10 min at 72^0C was performed and reactions were put on hold at 4^0C. The PCR products were visualised with agarose gel electrophoresis.

2.3.3 Ortholog identification and isolation

In order to obtain orthologs of *Arabidopsis* SCW associated TFs, a number of TFs involved in SCW biosynthesis in *Arabidopsis thaliana* were identified from current literature (Table 1). Amino acid sequences were obtained from TAIR (www.*Arabidopsis*.org) for each TF and pBLAST analyses were perfomed using the amino acid sequences as query against the protein databases of *A. thaliana* on TAIR and *Eucalyptus grandis* on Phytozome (www.phytozome.net), respectively. For each amino acid sequence, the five BLAST hits with the lowest E-values from each database were used for further analyses. These sequences were subsequently aligned using the ClustalW algorithm and used to compile Neighbour-joining phylogenies using the MEGA5 program (Tamura, Peterson et al. 2011). A bootstrap test with 1000 replicates and a Jones-Taylor-Thornton model with Gamma-distributed rates among sites (JTT + G) was used. Expression data for the top *E. grandis* hits were obtained from *E. grandis* RNA-seq data (Hefer et al., in preparation) in immature xylem, mature leaf, flowers, phloem, roots, shoot tips and young leaf. The xylem to leaf expression ratio (X:L) was also calculated for the *E. grandis* hits by dividing the percentage expression from

immature xylem by the average of percentage expression across young and mature leaves. Neighbour-joining trees were used to identify orthologs in *E. grandis* for each TF (Supplementary Figure 1). Candidate orthologs were analysed for preferential expression in *E. grandis* immature xylem using RNA-seq data from three *E. grandis* trees (Hefer et al., in preparation, Supplementary Figure 2). Where it was not possible to elucidate the ortholog from the neighbour joining trees (as in the case of EgrVND6 and EgrZF2), putative orthologs were identified from *E. grandis NAM ATAF CUC (NAC)* gene family phylogenetic trees in Hussey et al. (in preparation). Names were assigned to these orthologs based on the phylogenetic relationships to their *Arabidopsis* counterparts (Table 3). The orthologs will be reassigned with more permanent names upon publication of the *Eucalyptus grandis* genome paper (Myburg et al., in preparation).

For directional cloning, primers were designed to bind at the start and stop codons of each putative ortholog (Table 2), respectively, using *Primer Designer 4* software v4.20 (Sci-Ed Software, Durham, NC). The coding sequences for the putative orthologs were PCR amplified from *E. grandis* immature xylem cDNA. Approximately 1 µg of mRNA was converted to cDNA and enriched for full-length cDNAs using the SMART™ cDNA library construction kit (Clontech, Mountain View, USA), and used as template in the PCR reactions. Final concentrations of 0.5 µM of each primer, 0.2 mM dNTPs, 1x Phusion PCR buffer and 0.5 U of Phusion Taq (New England Biolabs, Ipswich, USA) were used in a total volume of 20 µl per reaction. The PCR was performed as follows: an initial denaturation for 30 s at 94^{0}C was performed, then 30 cycles of the following program; 7 s at 94^{0}C, 15 s at designated Tm (Table 3) and 45 s at 72^{0}C. A final elongation step of 2 min at 72^{0}C was perfomed and reactions were put on hold at 4^{0}C. PCR products were analysed via agarose gel electrophoresis. Bands of the correct sizes (Table 3) were excised and purified using the *NucleoSpin*® Gel and PCR Clean-up kit (Macherey-Nagel & Co., Dűren, Germany), cloned into the pCR8 GW pTOPO Gateway™ vector (pCR8/GW/TOPO, Invitrogen, Carlsbad, CA) and transformed into chemically competent *E. coli* DH5α cells. Colony PCRs were

performed to determine the orientation of the inserts. Briefly, seven colonies were picked for each transformation and resuspended in 20 µl of sterile water. The suspension was incubated at 90^{0}C for 10 min to lyse the cells. A volume of 3 µl of suspension was used as template. Final concentrations of 0.4 µM of each primer in the pair (Table 3, Supplementary Table 1), 0.2 mM dNTPs (Fermentas, Glen-Burnie, USA), 1x ExSel PCR buffer and 0.8 U of ExSel Taq (Separation Scientific, Johannesburg, South Africa) were used in a total volume of 20 µl per reaction. The PCR was performed as follows: an initial denaturation for 2 min at 92^{0}C was performed, then 30 cycles of the following program; 20 s at 92^{0}C, 30 s at 58^{0}C and 1 min 15 s at 72^{0}C. A final elongation step of 10 min at 72^{0}C was perfomed and reactions were put on hold at 4^{0}C. The PCR products were visualised with agarose gel electrophoresis. To determine orientation, colony PCRs using three combinations of vector-specific and gene-specific primers were performed (Supplementary Table 1). Plasmids from three clones of each ortholog in the correct orientation were purified using the GeneJET^{TM} Plasmid Miniprep Kit (Fermentas) and used as template for sequence analysis (Macrogen Inc., Seoul, Korea). Sequences were aligned and a consensus sequence was generated using the CLC Bio main workbench v.6.0 (CLC Bio, Katrienbjerg, Denmark). The consensus sequences were translated into amino acid sequences and aligned to the translated CDS of the orthologs obtained from the *E. grandis* genome assembly version 1.0 on Phytozome (www.phytozome.net). The amino acid alignments were used to differentiate between allelic variants and PCR induced mutations in the amplified putative orthologs. A substitution was considered allelic if it was silent or if it was consistent with the *E. grandis* RNA-seq data. All sequence and protein alignments are given in Appendix 2.

2.3.4 Synthesis of EgrMYB46 and EgrKNAT7

Due to difficulty in amplifying *EgrMYB46* and *EgrKNAT7* from *E. grandis* cDNA, a gene synthesis approach was used. The CDS of *EgrMYB46* and *EgrKNAT7* were manually

annotated in BoGaS (http://bioinformatics.psb.ugent.be/webtools/bogas) and downloaded. BLASTx (Altschul et al. 1990) analyses were performed against the NCBI non-redundant protein sequence database. The five sequences with the lowest bit score were downloaded for each gene and the DNA binding domains were identified using BLASTx with the BLOSUM62 matrix against the same database. The sequences were translated to amino acid sequences and aligned in CLC Bio main workbench v.6.0. If the *E. grandis* amino acid sequences deviated from the consensus in the DNA binding domain regions, the change was evaluated against the RNA-seq data by browsing the transcriptome reads mapped to the *E. grandis* V1.0 genome in IGV browser (http://www.broadinstitute.org/igv). If the change was not present in the transcriptome data, the sequence was edited to match the consensus in the transcriptome data. If the change was present in the transcriptome data, the sequence was accepted. Finalised sequences were sent to Genscript (Genscript USA inc., Piscataway, NJ, USA) for gene synthesis. The synthesised genes were obtained in the pUC57-kan vector and PCR amplified using gene-specific primers as described in section 2.3.3.

2.3.5 Yeast one-hybrid analyses

Bait vector construction

Bait vectors containing 1.5 kb and 500 bp *EgrSND2* promoter fragment were constructed. Gateway® pCR8/GW/TOPO plasmid (Invitrogen) containing the 1.5 kb and 500 bp promoter fragments were previously constructed. To directionally clone the 1.5 kb promoter into the bait vector, the Gateway® pCR8/GW/TOPO plasmid (Invitrogen) containing the 1.5 kb promoter sequence with MluI and SpeI restriction site overhangs was extracted from an *E. coli* DH5α clone. The promoter fragment was excised from Gateway® pCR8/GW/TOPO plasmid by digestion with MluI and SpeI restriction enzymes (Invitrogen), and was ligated into the pHIS2.1 plasmid. pHIS2.1 vector was linearised by MluI and SpeI digestion and dephosphorylated with shrimp alkaline phosphatase (Fermentas). Restriction enzyme digestions were performed by mixing 10 u of MluI and SpeI restriction enzymes each

(Fermentas), 2 µl 10x buffer Tango (Fermentas) and approximately 1 µg plasmid DNA on ice. The reaction mixture was made up to 20 µl using SABAX water (Adcock-Ingram). The reaction mixture was incubated at 37^0C for 16 hours. Digested vector was dephosphorylated by mixing MluI and SpeI double digested empty pHIS2.1 plasmid with 2 U Shrimp Alkaline Phosphatase (Fermentas) and 10 µl 10x Shrimp Alkaline Phosphatase reaction buffer on ice. The reaction mixture was made up to 100 µl using SABAX water (Adcock-Ingram) and incubated at 37^0C for 30 minutes. Promoter fragment and empty dephosphorylated vector were mixed in a ratio of three to one, respectively. Ligations were carried out by adding 2 µl 10x T4 DNA ligase reaction buffer (Fermentas) and 5 u T4 DNA ligase (Fermentas) to the reaction mixture on ice. The reaction mixture was made up to 20 µl using SABAX water (Adcock-Ingram) and incubated at 16^0C for 16 hours. The ligation mixture was transformed into chemically competent *E. coli* DH5α cells. Positive transformants were selected for on medium containing kanamycin (Sigma-Aldrich Pty. Ltd., Johannesburg, South Africa). Colony PCRs using *EgrSND2* promoter-specific forward and pHIS2.1-specific reverse primers (Supplementary Table 1) were performed as described in section 2.3.3 on eight colonies to determine if the fragment was present. Of those colonies presenting a band of the correct size, plasmid was purified from three using the GeneJET^{TM} Plasmid Miniprep Kit (Fermentas) and used as template for sequence analysis (Macrogen Inc., Seoul, Korea), using promoter specific and pHIS2.1 specific primers (Supplementary Table 1). The 500 bp bait was prepared as described above, but Gateway® pCR8/GW/TOPO plasmid containing the 500 kb promoter sequence with MluI and SpeI cut site overhangs was used as the source of the fragment. All sequence and protein alignments are given in Appendix 2.

Prey vector construction

To construct Y1H prey vectors, coding sequences of the putative orthologs were cloned from the Gateway® (pCR8/GW/TOPO) entry vector (Invitrogen) into the pDEST-GADT7 destination vector (Rossignol, Collier et al. 2007) by LR recombination (Curtis and Grossniklaus 2003) using the Gateway® LR Clonase® II Enzyme mix kit (Invitrogen). Five

clones were selected using colony PCR for confirmation with vector-specific pGAD primers (Supplementary Table 1). Colony PCRs were performed as described in section 2.3.3. Recombinant destination plasmids were purified using the GeneJET Plasmid Miniprep kit (Fermentas) and used as template for sequence analysis (Macrogen, Korea) using the pGAD and M13 primers (Supplementary Table 1). Resulting sequences were aligned and consensus sequences were generated as described in section 2.3.3. The consensus sequences were used to verify that no mutations occurred. All sequence and protein alignments can be viewed in Appendix 2.

Yeast transformations

To perform a Y1-H analysis, bait and prey plasmids needed to be transformed into yeast cells. The general transformation protocol for the yeast transformations was as follows: Approximately 50 µl of Y187 (Clontech) yeast cells were spotted onto a YPAD (2% bacto-tryptone, 1% yeast extract, 0.003% L-Adenine Hemisulphate, pH 6.5) plate. The cells were cultured for 3 days at 30^0C. The resulting lawn of cells were lightly scraped off the plate using a sterile pipette tip and resuspended in 1 ml of sterile SABAX water (Adcock-Ingram). Resuspended yeast cells were centrifuged for 5 min at 700 g and the supernatant was discarded. The pellet was resuspended in 1 ml of freshly made 100 mM LiAc and incubated at 30^0C for 5 min without agitation. Resuspended cells were again centrifuged at 700 g for 5 min and the supernatant was discarded. This was followed by the sequential addition of 240 µl 50% PEG 4000, 36 µl 1 mM LiAc, 25 µl herring testes carrier DNA (2 mg/ml) and approximately 500 ng of plasmid in a volume of 50 µl SABAX water (Adcock-Ingram). The pellet was resuspended by vortexing for at least 1 minute followed by incubation at 42^0C for 25 min. The mixture was briefly vortexed 15 minutes into incubation. Cells were collected by centrifugation at (1000 g for 5 min), the supernatant was discarded, and the cell pellet was resuspended in 200 µl sterile SABAX water (Adcock-Ingram). The entire mixture was plated onto the appropriate SDO (single dropout) media and incubated at 30^0C for 3 days. All DO media were prepared according to supplier guidelines (Clontech). SDO medium lacking

tryptophan (-Trp) was used to select yeast transformed with bait constructs and SDO medium lacking leucine (-Leu) was used to select yeast transformed with prey constructs. Double dropout (DDO) medium deficient in both leucine and tryptophan (-Leu-Trp) was used to select for yeast cotransformed with bait and prey plasmids. Transformed colonies were restreaked onto the same selective medium and incubated at 30^0C for 3 to 5 days in order to deplete the yeast cells of residual amino acids and maintain selection for the plasmids. When some prey vectors (*EgrMYB103* and *EgrMYB52/54-C* prey vectors) were transformed into yeast cells containing the bait plasmid, they yielded colonies which grew on DDO medium. However, when they were restreaked onto the same medium, they failed to grow. These prey vectors were discarded from the assay.

Yeast one-hybrid Interaction screens

In order to detect interaction between the identified TFs and the *EgrSND2* promoter fragments, a direct Y1-H screen was performed. A number of direct targets were selected to test in the Y1-H analysis (Table 3). Sequential yeast transformations were performed. Initially, bait strains were produced by transforming only the pHIS2.1 bait plasmid (Clontech) containing the *EgrSND2* promoter fragments (1.5 kb and 500 bp) into Y187 yeast cells. Positive transformation events were selected by inoculating the transformation mixture onto petri plates containing -Trp SDO medium. Colonies containing the bait plasmid were then cultured and used for transformations with prey vectors as described in section 2.3.5.3. The transformation mixtures were inoculated on petri plates containing DDO medium (-Leu, -Trp) to screen for presence of bait and prey plasmids. To test for Y1H interactions, the co-transformed yeast was plated on to petri plates containing TDO (triple dropout) medium deficient in leucine, tryptophan and histidine (-Leu-Trp-His) and containing 0 mM, 1 mM, 2 mM, 3 mM and 5 mM of 3-aminotriazole, respectively. To do this, co-transformed colonies were picked by hand and resuspended in 400 µl sterile SABAX water (Adcock-Ingram). The OD_{600} of the suspensions were determined using the Multiskan Go spectrophotometer and plate reader (Thermo Fisher Scientific, Johannesburg, South Africa). The approximate

number of cells/µl were calculated from the OD_{600} and diluted to a concentration of roughly 1000 cells/µl (OD_{600} of approximately 0.1).

Suspensions containing equal numbers of cells were placed on the petri plates in 10 µl spots and were left in the laminar flow cabinet to dry for 20 minutes. Inoculated petri plates were incubated at 30^0C for 3 to 5 days, after which they were removed and photographed. Where possible, at least six replicates were analysed (Appendix 3).

2.4 *In silico* analysis of the EgrSND2 promoter fragment

In silico analysis of the *EgrSND2* promoter fragment was performed using the Regulatory Sequence Analysis (RSA) toolkit (Thomas-Chollier et al. 2008). Three *cis*-elements associated with SCW biosynthesis were identified from literature, namely MYB46RE (RKTWGGTR; Kim et al. 2012), SMRE (ACCWAMY; Zhong and Ye 2012) and SNBE (WNNYBTNNNNNNNAMGNHW; Zhong et al. 2010b; McCarthy et al. 2011). Using the string pattern match function, the putative number and positions of each of these *cis*-elements in the 1.5 kb *EgrSND2* promoter fragment were determined.

2.4 Results

2.4.1 Generation of *1.5 kb EgrSND2 promoter::GUS* transgenic plant lines

Transgenic plant lines expressing *GUS* under the control of the 1.5 kb *EgrSND2* promoter were generated for use in the qualitative *GUS* analysis. pMDC162 plasmid containing the 1.5 kb *EgrSND2* promoter sequence directly upstream of the *β-glucuronidase* coding sequence was transformed into wild-type Col-0 *A, thaliana* plants. Five *EgrSND2 1.5 kb promoter::GUS* plant lines (D-L, as depicted in Figure 1) showing a consensus phenotype

were selected as representatives and used for analyses. All other plant lines and replicates are shown in Appendix 1. PCR amplification using a promoter-specific forward primer and a *GUS*–specific reverse primer resulted in bands of expected size, approximately 3 kb, for all the transgenic *EgrSND2 1.5 kb promoter::GUS* lines and 1.7 kb in size for the *35S CaMV promoter::GUS* lines, respectively (Figure 1). In a number of cases, such as F and J (Figure 1A) and A-C (Figure 1B) there is a second band present. The band is approximately 1.5 kb and 900 bp in size for the *EgrSND2* and *35S* constructs respectively. This is due to an imperfect extra binding site for the GUS(R) primer present in the pMDC162 vector at position 2442. This binding site is located directly upstream of where the promoter insert would be located in the vector. Amplification from this site would therefore result in a fragment that is approximately 1.2 kb smaller in size due to the exclusion of the *GUS* gene fragment from the PCR amplification. Other pale bands were also seen, but this was most likely due to non-specific binding of the primers.

2.4.2 Qualitative *GUS* analysis of the *EgrSND2* 1.5kb promoter fragment

A qualitative *GUS* analysis was performed to determine in which tissues the EgrSND2 1.5 kb promoter fragment was active. Qualitative GUS analyses were performed on whole plants at one- and three-weeks of age, and specific plant tissues at six weeks of age. Basic anatomy of the six week organs analysed are shown in Supplementary Figure 4. Genevestigator analysis revealed that the expression pattern of *AtSND2* is very specific in *A. thaliana* tissues (Figure 2). *AtSND2* is expressed specifically in the hypocotyl and inflorescence stem (Figure 2).

The one-week-old seedlings of the *EgrSND2* lines showed GUS reporter staining similar to that of the *35S::GUS* seedlings (Figure 3). It appeared that expression was strongly activated in many tissues as opposed to being specific to a particular tissue. In some lines (S2_F and S2_L) the roots of the seedlings showed less staining than the *35S::GUS*

controls. This is an indication that the expression pattern of the *EgrSND2* promoter is not identical to that of the *35S CaMV* promoter.

At three weeks old, all lines exhibited strong GUS staining resembling the *35S::GUS* control. The vasculature of the leaves stained darker than the rest of the leaves (Figure 4). Younger, smaller leaves stained darker than the older leaves, possibly due to an initial strong expression of the *EgrSND2* promoter fragment caused upon initiation of leaf tissue development, particularly vascular tissues, which are no longer as active in more mature leaves. The roots of all lines were darker at this time point, indicating high levels of expression (Figure 4). There were also dark spots on the leaves of some plants, including that of the *35S::GUS* control (Figure 4). This observation is characteristic of stained trichomes (Shangguan et al. 2008). Where wounding had occurred on the plants there tended to be a stronger *GUS* signal. This may be due to better *GUS* infiltration, resulting in dark patches of heavy *GUS* activity (S2_D and S2_E; Figure 4).

In the six-week-old *EgrSND2* plants, GUS staining was observed in all analysed tissues (Figure 5). The mature leaves showed patchy GUS staining. This is probably due to leaf senescence, where the GUS enzyme will not be active in dead tissue, and also due to increased infiltration of the *GUS* buffer in wounded sites. There was some staining in mature leaf vasculature, but it was not comparable to leaves displaying true vascular specific expression, such as those of transgenic *EgrCesA8::GUS* promoter lines (Figure 6, Creux et al, unpublished data). Looking at the root and hypocotyls, it was observed that the primary root epidermis did not seem to express *GUS*, while there was strong expression in the lateral roots and the tissues within the hypocotyls (Figure 5, column RH). *GUS* expression was moderate in the siliques, mainly at the tips and abscission zones (Figure 5, column S). In the flowers, expression was seen in the inflorescence stem, the sepals and the tip of the stamens (Figure 5, column F). Cross-sections of the upper and lower stems showed expression in all living tissues, with particularly dark staining in the xylem and interfascicular

fibres, xylem vessels, vascular cambium and phloem. Even though the cortex, epidermis and pith also showed staining, it was not as intense as the other tissues of the inflorescence stem (Figure 5, columns LS and US). In general, staining appeared to be strongest in living tissues, but more intense in fast growing tissues, such as the base of the leaves, or tissues with thick SCWs such as the vessels and fibres in xylem. The tips of whole stem segments also stained darkly, while the middle remained unstained. This was probably due to the limited capacity of GUS to infiltrate the tissues.

2.4.3 Ortholog identification and construction of Y1-H bait and prey vectors

Bait plasmids containing the 1.5 kb and 500 bp promoter fragments and prey plasmids containing the CDS of the TFs being investigated were generated for use in the Y1-H analyses. The 500 bp construct was generated to provide spatial data on where the prey proteins may be binding. PCR amplification using the promoter specific forward and pHIS2.1 reverse primers of plasmid DNA extracted from colonies containing the 1.5 kb and 500 bp bait constructs was performed. The PCR resulted in band sizes of approximately 2 kb and 1 kb respectively (Figure 7). These band sizes were what we expected to obtain from the PCR analysis. To generate prey constructs, a number of TFs were identified from literature as being implicated in or potentially important to SCW biosynthesis in plant xylem cells (Table 1). Phylogenetic analyses were performed on these TFs to determine their orthologs in *E. grandis* (Table 3, Supplementary Figure 1).

Phylogenetic analyses revealed strong support for orthologs of a number of *Arabidopsis* proteins (Supplementary Figure 1). Clear orthologs for AtKNAT7, AtMYB85, AtMYB103, AtVND7 and AT166810 were identified. EgrMYB103 and EgrKNAT7 had bootstrap support of 100% and were highly specific to *E. grandis* immature xylem in their expression profiles (Supplementary Figure 1 and 2). The EgrZF1 ortholog also had high confidence with

bootstrap support of 95% and expression specific to immature xylem (Supplementary Figure 1 and 2). EgrMYB85 only had 62% bootstrap support, but was the only amino acid sequence that grouped with AtMYB85 and was also highly specific to immature xylem. Conversely, the EgrVND7 ortholog had high homology to its *Arabidopsis* counterpart (bootstrap support of 100%), but was not highly specific to immature xylem (Supplementary Figure 1 and 2).

While there was bootstrap support for the majority of the TFs, the phylogenies were unclear, resulting in the multiple putative orthologs for the TF in question. This was the case for AtSND1, AtNST1, AtMYB46, AtMYB83, AtMYB52, AtMYB54, AtSND2 and AtSND3 (Supplementary Figure 1). AtSND1 and AtNST1 were highly similar to each other (68% similarity) and it was therefore difficult to identify putative orthologs for each. It was possible to differentiate between EgrSND1 and EgrNST1 by viewing phylogenies of the full NAC family of proteins from *E. grandis* (Hussey et al., in preparation). Likewise, it was possible to identify specific putative orthologs for both AtSND2 and AtSND3, which were also highly similar to each other (75% similarity). For AtMYB52 and AtMYB54, there were three *E. grandis* putative orthologs that showed high sequence homology to both AtMYB52 and AtMYB54 and were expressed in *E. grandis* immature xylem in a highly specific manner (Supplementary Figure 1 and 2). All three of these putative orthologs were selected and were designated EgrMYB52/54-A,-B and -C.

For the last two TFs, AtVND6 and AT1G72220, no *E. grandis* amino acid sequences clustered with the *Arabidopsis* amino acid sequences (Supplementary Figure 1). This was most likely due to a lack of resolution caused by the small size of the phylogenies. It was possible to identify a putative *E. grandis* ortholog of AtVND6 by examining phylogenies of the *E. grandis* family of NAC domain TFs (Hussey et al., in preparation). This was not possible for AT1G72220, as this protein is not part of the NAC domain family of TFs, though one of the putative orthologs in the phylogeny did display expression highly specific to immature xylem, with a xylem to leaf ratio of expression of 154.41 (Supplementary Figure 1 and 2). This putative ortholog was selected as EgrZF2.

All selected putative orthologs were highly expressed in *E. grandis* xylem, with the exception of EgrVND7 which appears to be expressed in a non-specific fashion (Supplementary Figure 1). The putative orthologs were amplified from a *E. grandis* immature xylem cDNA library using gene specific primers with the exception of *EgrMYB46* and *EgrKNAT7,* which were synthesised by Genscript (Genscript USA Inc.). The putative orthologs were then successfully LR cloned into the pDEST-GADT7 prey vector. PCR amplifications were performed on the prey vectors to confirm the presence of the inserts (Supplementary Table 1). In all cases, amplicons of expected size were obtained (Figure 7 and Table 3).

2.4.4 Y1-H analysis of the 1.5kb *EgrSND2* promoter fragment

In order to determine which SCW-related DNA-binding proteins bind to the 1.5 kb *EgrSND2* promoter fragment, a Y1-H analysis was performed using the 1.5 kb *EgrSND2* promoter fragment as bait and the putative *E. grandis* TF orthologs as prey (Figure 8 and Table 3). A number of prey vectors, mainly NAC TFs (EgrVND7, EgrSND1 and EgrNST1), transformed with extremely low efficiency. When these colonies were inoculated onto DDO medium they failed to grow. These TFs may be cytotoxic to yeast cells (Van Aken et al. 2013), preventing them from growing. Also, due to strong autoactivation of the 1.5 kb bait, binding resolution was lower, as the negative control was only inhibited at a concentration of 5 mM 3-AT (Figure 8). This is probably due to endogenous yeast proteins binding to the promoter fragment and activating the *HIS3* reporter gene. However, a number of interactions were still apparent. EgrZF1, EgrZF2, EgrVND6 and EgrMYB46 showed growth even at 3-AT concentrations of 5 mM (Figure 8). Interestingly, neither EgrMYB52/54-A or EgrKNAT7 showed an interaction, but when both TFs were co-transformed into yeast cells containing the 1.5 kb bait, they displayed a strong interaction that was stable at 3 mM 3-AT (Figure 9). These interactions may be confirmed in the future by performing a dilution series on a

selective medium with a relatively high concentration of 3-AT (e.g. -Leu-Trp-His 3 mM 3-AT TDO medium in this case), to account for autoactivation.

2.4.5 Y1-H analysis of the 500bp *EgrSND2* promoter fragment

To determine which of the TFs (Table 3) bind to the 500 bp *EgrSND2* promoter fragment, a Y1-H analysis was performed on the 500 bp *EgrSND2* promoter fragment bait (Figure 10). As mentioned previously, a number of NAC domain TFs transformed with very low efficiency. For a number of TFs, there was no growth of yeast spots at any concentration of 3-AT. These included EgrMYB52/54-A, EgrMYB52/54-B, EgrMYB85, EgrKNAT7, EgrMYB46 and EgrMYB103 (Figure 10). These TFs were most likely not interacting with the promoter fragment. EgrZF1 showed strong interaction at 5 mM 3-AT (Figure 10). Lastly, EgrMYB83, EgrVND6, EgrSND2 and EgrSND3 showed interaction with the promoter fragment, with yeast spot growth being inhibited at a 3-AT level of approximately 2 mM (Figure 10). However, replicates for analyses of EgrSND2 were poor and will need to be repeated in the future. Once again, EgrMYB52/54-A and EgrKNAT7 did not show an interaction when transformed separately into the bait lines, but showed a strong interaction when both TF preys were transformed into yeast cells containing 500 bp bait (Figure 9).

2.4.6 *In silico* analysis of the 1.5 kb *EgrSND2* promoter fragment

In silico analysis was performed to determine the location of putative binding sites in the promoter fragments using the RSA tool suite (Thomas-Chollier et al. 2008). Three *cis*-elements important in transcriptional regulation of SCW biosynthesis, and likely targets of the TFs being analysed in this study, were selected from literature. Using the string pattern matching function in the RSA tool suite, two putative MYB46RE (RKTWGGTR, at positions -627 and -1137, Figure 11A; Kim et al. 2012), three putative SMRE (ACCWAMY, at positions

-667, -1136 and -1204, Figure 11B; Zhong and Ye 2012) and five putative SNBE (WNNYBTNNNNNNNAMGNHW, at positions -525, -526, -644, -985 and -1294, Figure 11C; Zhong et al. 2010b; McCarthy et al. 2011) elements were discovered in the *EgrSND2* promoter fragment (Figure 11D).

2.5 Discussion

Transcriptional control of SCW biosynthesis in xylem has been well studied (Demura and Fukuda 2007; Zhong and Ye 2007; Zhong et al. 2008; Du and Groover 2010; Zhong et al. 2010a; Zhong et al. 2011). This process is controlled by a cascade of TFs in a hierarchical manner, and is highly conserved between plant species (Zhong et al. 2010a). However, the main focus has been on the master regulators of the network, with less research being performed on the lower level TFs of the network. *At*SND2 was identified as one of a suite of TFs that are important for SCW biosynthesis in *Arabidopsis* interfascicular fibres (Zhong et al. 2008). It is highly expressed in *Arabidopsis* xylem fibres and is upregulated by AtSND1. AtSND2 in turn can activate *AtCesA8* expression (Zhong et al. 2008), which implies that *AtSND2* is involved in cellulose biosynthesis in SCWs of *Arabidopsis* xylem fibres. By studying the *Eucalyptus grandis* functional ortholog of *AtSND2* (*EgrSND2*) in *Arabidopsis,* we are able to effectively evaluate its role in the SCW regulatory network. This will provide insight into the role of *SND2* in *Eucalyptus*, as well as other tree species, and may eventually allow for the modification of commercially important tree crops such as *Eucalyptus spp* in order to obtain desirable traits such as higher cellulose content and lower lignin content in the woody biomass. We used qualitative GUS analysis as it is a relatively cheap and well established technique to perform. We decided to use Y1-H as it does not require expression and purification of a recombinant protein, unlike other methods such as electrophoretic mobility shift assays (EMSAs). Our study on the expression patterns of *EgrSND2* using qualitative β-glucuronidase (GUS) assays and the protein interactions of its promoter using

Y1-H analyses elucidated new links in the transcriptional network of SCW biosynthesis and confirms suspected interactions.

2.5.1 The *EgrSND2* promoter fragment is not exclusively active in *Arabidopsis* interfascicular fibres

We showed through qualitative GUS assays that the heterologously expressed *EgrSND2* promoter is active in many different tissues at many different times (Figure 3-5). At no time point was expression specific to the interfascicular fibres. At first, the expression pattern appears similar to that of the *35S CaMV* promoter. Upon closer inspection, it was seen that the roots of the one- and six-week old plants showed less or no expression of GUS and that in general the GUS signal was less intense in the *EgrSND2* promoter plant lines. Regardless, this shows that the 1.5 kb *EgrSND2* promoter fragment is a strong transcriptional activator in *Arabidopsis.* Staining was most prominent at areas such as base of the mature leaves, the abscission zone and tip of the siliques, the sepals of the flowers and the developing vascular tissues in both the upper and lower stem of the six-week-old plants (Figure 5). There was also heavy staining observed in trichomes on the leaves of three-week old *EgrSND2* promoter plant lines (Figure 4). Some of this may be ascribed to the efficiency of GUS infiltration, but most of these tissues are fast growing (Poethig 1997) and are depositing thick SCWs (Werker 2000; Mitsuda et al. 2005; Demura and Fukuda 2007; Zhong and Ye 2007), indicating that the *EgrSND2* promoter fragment is most active in those tissues that are growing rapidly or cells that are actively depositing SCWs.

As *EgrSND2* is associated with cellulose biosynthesis in interfascicular fibres (Zhong et al. 2008), we would expect to see strong expression in vasculature tissues. There was some staining in the six-week-old mature leaf vasculature (Figure 5), but was not comparable to plants showing true vascular expression, such as those containing *EgrCesA8::GUS*

promoter constructs (Figure 6). This may be due to the process of vascular development being completed in mature leaves, as opposed to leaves in three-week-old plants, where vascular development is still occurring (Figure 4). In contrast, the vascular specific expression pattern of *EgrCesA8* in mature six-week-old leaves is expected, as the *CesA* genes are at the bottom of the transcriptional network, and are thus subjected to regulation by all the TFs higher in the hierarchy, allowing for more specific regulation of expression.

In the inflorescence stems of a minority of plants, there is staining in the vascular tissues only (Supplementary Figure 10, Supplementary Figure 11). With the exception of S2_M, where the *GUS* reporter gene most likely integrated in close proximity to an SCW related gene in the *Arabidopsis* genome, this is most likely not vascular-specific expression, as the expression in the leaves of these plants at six weeks does not match that of a true vascular specific line (such as those expressing GUS under the control of the *EgrCesA8* promoter, Figure 6). Rather than showing evidence of vascular specificity, this may be due to the ability of the GUS buffer to penetrate further into the stem because of the open channels provided by the vascular tissues. Therefore a microscope section taken at a point distant from the edge of the stem sample may show staining in the vascular tissues and nowhere else.

Y1-H analysis suggests that a number of proteins are binding to the *EgrSND2* promoter (Figures 8 and 10). Some of these are particularly useful for determining the specificity of *EgrSND2*. EgrVND6, the *E. grandis* ortholog of a master regulator of SCW biosynthesis in metaxylem vessels (Kubo et al. 2005; Ohashi-Ito et al. 2010; Yamaguchi et al. 2010), was found to bind to the 500 bp and 1.5 kb *EgrSND2* promoter fragments (Figure 10). This implies that *EgrSND2* may aid in SCW biosynthesis in metaxylem vessels. The *EgrSND2* promoter is also bound by EgrMYB46 (Figure 8) and EgrMYB83 (Figure 10). These two proteins function redundantly as mid-level master regulators of SCW biosynthesis (McCarthy et al. 2009; Zhong and Ye 2012). They are able to induce the pathways for lignin, cellulose and hemicellulose biosynthesis and are direct targets of the top level master regulators

AtSND1, AtNST1, AtNST2, AtVND6 and AtVND7 (Zhong et al. 2007a; Zhong et al. 2008; McCarthy et al. 2009). It is also interesting to note that *AtSND2* is not a direct target of *AtSND1* (Zhong et al. 2008), indicating that it is lower in the transcriptional hierarchy, and thus subject to regulation by additional TFs. These interactions show that *EgrSND2* is bound and possibly influenced by proteins involved in the transcriptional regulation of SCW biosynthesis of different cell types such as proto- and metaxylem vessels, as opposed to being specific to interfascicular fibres in *Arabidopsis*.

Transcriptome data from the main stem tissues of three mature *E. grandis* trees shows that *EgrSND2* is highly expressed in immature xylem as compared to other tissues (Figure 2). This is not surprising as stem immature xylem is likely to have many cells synthesising SCWs that are thicker than those found in most other tissues (Demura and Fukuda 2007). The study by Zhong et al. (2008) showed that AtSND2 is expressed mainly in the xylem interfascicular fibres and vessels of *Arabidopsis* inflorescence stems. Considering that the construct in Zhong et al. (2008) included the full CDS of *AtSND2* with 3 kb of upstream and 2 kb of downstream sequence, with the *GUS* CDS inserted in frame just before the termination codon, as opposed to the construct used in our study, which contained only 1.5 kb of sequence upstream of the *EgrSND2* CDS, inserted directly upstream of the *GUS* CDS, it is possible that our construct may not have the required *cis*-elements for fibre/vessel specific expression. Looking at the xylem to leaf expression ratio of SND2 in *Arabidopsis* and *E. grandis* (6.09 and 16.06, respectively) expression is more specific to immature xylem in *E. grandis* than it is in *Arabidopsis* (Figure 2). This supports the statement that the GUS expression pattern of the *EgrSND2* 1.5 kb promoter fragment that is seen in *Arabidopsis* is more general due to absence or inclusion of fibre specific regulatory elements. The expression patterns and protein interactions of the *EgrSND2* promoter point to the conclusions that *EgrSND2* is not specifically active in *Arabidopsis* interfascicular fibres and that the *EgrSND2* 1.5 kb promoter fragment may be expressed in cells that are not actively synthesising SCWs.

2.5.2 EgrSND2 is bound by a number of TFs involved in SCW biosynthesis

We were able to clearly identify *E. grandis* orthologs for AtKNAT7, AtMYB103, AtVND7 and AT1G66810. We were also able to identify putative orthologs for AtSND1, AtNST1, AtMYB46, AtMYB83, AtMYB52, AtMYB54 AtSND2, AtSND3, AtVND6 and At1g72220. Confidence in the validity of the orthologs may be improved by reconstructing phylogenies with larger data sets and by performing initial alignments with an algorithm better suited to protein alignment than ClustalW, such as MAFFT (Katoh et al. 2002) or MUSCLE (Edgar 2004). However, this shows that EgrSND2 forms part of an at least partially evolutionarily conserved network of transcriptional control of SCW biosynthesis in *E. grandis*, as was suspected (Zhong et al. 2010a). As such, one would expect the EgrSND2 TF to regulate other SCW genes, as well as the *EgrSND2* gene to be regulated by other proteins in the network. In this study we performed Y1-H analysis with 1.5 kb and 500 bp fragments of *EgrSND2* promoter sequence respectively (Figures 8 and 10) in order to determine which TFs (Table 3) interact directly with this promoter. A number of TFs were found to interact directly with the *EgrSND2* promoter, namely EgrMYB46, EgrMYB83, EgrSND2, EgrSND3, EgrVND6, EgrZF1 and EgrZF2 (Figures 8 and 10).

The *E. grandis* ortholog of AtVND6 was found to bind to the *EgrSND2* 500 bp and 1.5 kb promoter fragments (Figure 10). It is important to note that AtVND6 has been shown to directly bind the promoters of *AtMYB46* and *AtMYB83* in previous studies (Zhong et al. 2007a; McCarthy et al. 2009). As the *EgrSND2* promoter is bound by both EgrMYB46 (Figure 8) and EgrMYB83 (Figure 10), and given that this network contains many feed-forward loops, this interaction is not unexpected. The binding of the vessel master regulator EgrVND6 and the mid-level master switches EgrMYB46 and EgrMYB83 indicate that *EgrSND2* may also play a role in SCW biosynthesis in vessel cells. This is further supported

by the presence of putative SNBE elements in the *EgrSND2* promoter fragment (Figure 11C), the binding sites of master regulators such as AtNST1, AtVND6 and AtVND7 (McCarthy et al. 2011).

EgrMYB46 and EgrMYB83 showed interaction with the 1.5 kb and 500 bp *EgrSND2* promoter fragments, respectively (Figures 8 and 10). AtMYB46 and AtMYB83 are mid-level master switches of SCW biosynthesis in xylem fibres and vessels, able to activate lignin, cellulose and hemicellulose biosynthetic pathways (Zhong et al. 2007a; Zhong et al. 2008; McCarthy et al. 2009; Zhong and Ye 2012). AtMYB46 has been shown to slightly upregulate AtSND2 (Zhong et al. 2007a). AtMYB83 has also been shown to slightly upregulate the other two TFs apart from AtSND2 that were able to activate the *AtCesA8* promoter, namely AtSND3 and AtMYB103 (Zhong et al. 2008; McCarthy et al. 2009). Given the general GUS expression pattern of the *EgrSND2 1.5 kb* promoter fragment and the binding of EgrVND6, this interaction in *E. grandis* between EgrMYB46, EgrMYB83 and the *EgrSND2* promoter (Figures 8 and 10) is expected.

In *E. grandis,* EgrSND3 binds to the 500 bp *EgrSND2* promoter fragment (Figures 8 and 10). In *Arabidopsis, AtSND3* is a direct target of AtSND1 and is able to activate the *CesA8* promoter. Although *AtSND2* is not a direct target of AtSND1, it is still upregulated in response to its overexpression (Zhong et al. 2008). Therefore, in *E. grandis,* EgrSND3 may be one of the intermediate TFs through which EgrSND2 is regulated in interfascicular fibres by EgrSND1. This is also interesting as it indicates that EgrSND2 may be at a slightly lower level in the hierarchy than EgrSND3, which is characteristic of the feed-forward loops which are prominent in this network (Badescu and Napier 2006; Yu and Gerstein 2006; McCarthy et al. 2009). In GUS analysis, the *EgrSND3* 1.5 kb promoter fragment showed an expression pattern very similar to that of the *EgrSND2* 1.5 kb promoter fragment (Supplementary Figure 13-21). This suggests they may be active in the same tissues, thus supporting the hypothesis that EgrSND3 may be important for the regulation of *EgrSND2*. It is also

interesting to note that EgrSND2 binds to its own promoter fragment, indicating that a feedback loop, in which the product of a gene may influence its own expression directly or indirectly, may be involved in the regulation of EgrSND2. This is not surprising as autoregulation has been observed to occur frequently in plant transcriptional regulation (Schoof et al. 2000; Spitzer-Rimon et al. 2012). This also means that EgrSND2 may be subject to temporal regulation as autoregulation is able to speed or slow the response to a stimulus, dependant on if the autoregulation is negative or positive, respectively (Alon 2007). However, in this study the replicates for the interaction between EgrSND2 and the 500 bp *EgrSND2* promoter fragment were poor (Figure 10), due to a low transformation efficiency of the EgrSND2 prey, and thus a lack of transformed colonies, so the Y1-H analysis for the EgrSND2 prey will need to be repeated in the future for confirmation.

EgrZF1 and EgrZF2 are orthologs of proteins recently found to be upregulated by AtMYB46 in *Arabidopsis* (At1g66810 or AtC3H14, and At1g72220, respectively; Ko et al. 2009). EgrZF1 showed interaction with both the *EgrSND2* 500 bp and 1.5 kb promoter fragments, while EgrZF2 only showed interaction with 1.5 kb *EgrSND2* promoter fragment (Figures 8 and 10). AtC3H14 has been shown to be a direct target of AtMYB46 and AtSND1, and is thought to be another mid-level regulator of SCW biosynthesis (Ko et al. 2009). However, other studies have refuted this claim, stating that AtC3H14 on its own is not sufficient to activate SCW biosynthesis (Zhong and Ye 2012). AtC3H14 has also been implicated in stress tolerance (Wang et al. 2008). Since cell walls are often modified in response to stress (Ringli 2010; Ramirez et al. 2011), it is possible that EgrZF1 binds to and activates EgrSND2 for SCW modifications in stress responses (Wang et al. 2008). Relatively little is known of At1g72220, other than it is upregulated in *Arabidopsis* plants overexpressing AtSND2 (Hussey et al. 2011), and is a direct target of AtMYB46 (Ko et al. 2009). As such this is an interesting target for future research.

Of the interacting TFs, EgrMYB46 and EgrZF2 only showed interaction with the 1.5 kb fragment of promoter (Figures 8 and 10). This indicates that the binding sites of these TFs are located in the distal 1 kb of promoter that is not present in the 500 bp promoter constructs. Two elements have been identified in *Arabidopsis* to which AtMYB46 binds, namely the MYB46RE (RKTWGGTR; Kim et al. 2012) and the SMRE (ACCWAMY; Zhong and Ye 2012) elements. These two elements are highly similar (Zhong and Ye 2012). *In silico* analysis of the *EgrSND2* promoter with the RSA tool suite (Thomas-Chollier et al. 2008) revealed that two MYBREs (at positions -627 and -1137) and three SMREs (at positions -667, -1136 and -1204) were present in the 1.5 kb EgrSND2 promoter fragment (Figure 11A and 11B). All of these putative elements were further than 500 bp upstream from the start codon, and are thus in agreement with the Y1-H results as possible binding sites for EgrMYB46. As of yet, no *cis*-element has been identified as a possible binding site for EgrZF2, but analysis of this promoter sequence as well as that of other targets of EgrZF2 may elucidate this in the future. EgrVND6 showed interaction with the 500 bp and 1.5 kb *EgrSND2* promoter fragments (Figures 8 and 10). This means that an element to which EgrVND6 binds should be located within the first 500 bp upstream of the start codon. There were no SNBE elements detected in this region (Figure 11C), which means that EgrVND6 may recognise and interact with an alternative motif.

EgrMYB83, EgrSND2 and EgrSND3 showed interaction with the 500 bp promoter fragment, but not with the 1.5 kb promoter fragment (Figures 8 and 10). This is contradictory, as we would expect interaction with both fragments if binding took place in the first 500 bp of sequence. However, there was strong autoactivation with the 1.5 kb fragment and it is possible that weaker interactions were masked by the autoactivation. Also, the increased length of the 1.5 kb promoter may have weakened expression of the reporter gene, by promoting binding of additional yeast proteins. It is also possible that the three dimensional structure of the promoter fragment was not optimal for the effective binding of the prey

proteins. This may account for the loss of interaction of these three proteins in the 1.5 kb promoter fragment Y1-H screens.

The non-interacting proteins in the Y1-H screens were EgrKNAT7, EgrMYB52/54-A, EgrMYB52/54-B, EgrMYB85 and EgrMYB103 (Figures 8 and 10). *AtMYB85* is upregulated by AtSND1 and AtMYB46 and is expressed in xylem vessels and fibres (Zhong et al. 2007a; Zhong et al. 2008). Dominant repression of AtMYB85 was shown to reduce fibre SCW thickness and resulted in deformed vessels. Overexpression of AtMYB85 leads to ectopic deposition of lignin in epidermal and cortical cells. AtMYB85 has been shown to induce the lignin biosynthetic pathway, and activate the *At4CL1* promoter (Zhong et al. 2008). It is believed to bind to AC elements prominent in the promoters of lignin genes (Patzlaff et al. 2003). *EgrSND2* does not appear to have any of these elements in its promoter (not shown) and is not known to play a role in monolignol biosynthesis; so these findings are in agreement with the Y1-H results.

In *Arabidopsis,* AtMYB52 and AtMYB54 are able to induce genes in the cellulose, lignin and hemicellulose biosynthetic pathways and are highly induced by AtSND1 (Zhong et al. 2008). They have also been shown to be induced by a number of other TFs such as AtNST1 and AtMYB46 (Ko et al. 2007; Ko et al. 2009; McCarthy et al. 2009). AtMYB52 appears to have a specific role in the later stages of xylem vessel formation in roots. It has also been observed to have a similar GUS expression pattern to that of AtMYB85 linking it to a role in lignin biosynthesis (Nakano et al. 2010). In addition to this, AtMYB52 has also been implicated in the repression of lignin biosynthesis (Cassan-Wang et al. 2013). AtMYB54 is a close downstream target of AtSND1 and is hypothesised to be a repressor of phenylpropanoid and lignin biosynthesis (Shen et al. 2009) though it is also co-expressed with lignin biosynthetic genes (Ohashi-Ito et al. 2010). Therefore, AtMYB54 appears to have a role in lignin biosynthesis, though it is unclear exactly what that role is. Given the roles of AtMYB52 and AtMYB54 in lignin biosynthesis, we would not expect to see these TFs binding to AtSND2,

as it has so far not been convincingly implicated in monolignol biosynthesis (Zhong et al. 2008; Hussey et al. 2011). Thus the lack of interaction between EgrMYB52/54-A and EgrMYB52/54-B and the *EgrSND2* promoter fragments seen in the Y1-H analysis is expected.

AtMYB103 is a direct target of AtSND1 (Zhong et al. 2008), AtNST1, AtNST2, AtVND6 (Ohashi-Ito et al. 2010) and AtVND7 (Yamaguchi et al. 2010). It has been shown to be expressed in the interfascicular fibres and xylem vessels of *Arabidopsis* stems (Zhong et al. 2008) and in the metaxylem vessels of *Arabidopsis* roots (Nakano et al. 2010). It is also able to activate the *AtCesA8* promoter (Zhong et al. 2008) and has also been shown to affect *F5H* expression and S-lignin quantity in SCWs (Öhman et al. 2013). This implies that TFs can play different roles in different tissues of the plant. Using *E. grandis* orthologs, binding was not detected between EgrMYB103 and the *EgrSND2* promoter fragment (Figures 8 and 10). This suggests that *EgrSND2* does not play a role in the monolignol biosynthetic pathway and that, unlike EgrSND3, EgrMYB103 is probably not a channel through which EgrSND1 activates *EgrSND2*.

AtKNAT7 is a direct target of AtNST1, AtSND1, AtVND6, AtVND7 (Zhong et al. 2008) and AtMYB46 (Zhong and Ye 2012). In the inflorescence stem, it is expressed in developing metaxylem, phloem, the cambium and the cortical cells. It is believed to be a repressor, as overexpression of this protein leads to thinner interfascicular cell walls. In contrast, knockout of this transcript leads to thicker interfascicular fibre cell walls with higher lignin content (Li et al. 2012). It is interesting to note that, while the interfascicular fibre cells walls are thinner in the overexpression lines, the xylem fibres and vessels remained normal. In the knockout lines, the interfascicular fibre walls were thicker, but the vessel element cell walls were thinner. Loss of function of AtKNAT7 also led to an upregulation of *AtCesA7* and *AtCesA8*, and to a lesser extent *AtCesA4* (Li et al. 2012). It is thought that AtKNAT7 could be involved in a negative feedback mechanism controlling lignin biosynthesis (Li et al. 2012). The lack of

interaction between EgrKNAT7 and the *EgrSND2* promoter fragments suggest that the *EgrSND2* 1.5 kb promoter fragment may not play a role in KNAT7 mediated repression of lignin biosynthesis.

It is important to note that although many of these proteins do not bind to the *EgrSND2* promoter, their *Arabidopsis* orthologs may still show some small activation of *AtSND2* or some involvement in *AtSND2* associated pathways (Demura and Fukuda 2007; Zhong et al. 2007a; Zhong et al. 2008; Ko et al. 2009; Nakano et al. 2010). These discrepancies may be due to these proteins playing roles in different parts of the network (higher or lower in the hierarchy), different tissues of the plant, being at different stages of development or as parts of different protein complexes. Some of these may also be due to the fact that we are looking at an analogous network in *E. grandis* and not the network in *Arabidopsis,* on which the bulk of the literature is focused. Lastly, the *in vivo* environment of Y1-H may prevent us from seeing some true interactions, as the proteins in question may require other proteins or factors not present at the time to bind. In both the 500 bp and 1.5 kb Y1-H analyses, when EgrMYB52/54-A or EgrKNAT7 prey were screened against the baits individually, no interaction was seen. However, when these preys were screened together, there was strong interaction (Figures 8 and 10). This suggests that these proteins may need to behave cooperatively in order to bind to DNA.

2.5.3 The EgrSND2 promoter plays an important role in cellulose biosynthesis

In the Y1-H analyses, a selection of TFs were tested against the *EgrSND2* promoter (Table 3). We have seen a number of TFs binding to the *EgrSND2* promoter (Figures 8 and 10). Of these, EgrVND6, EgrMYB46 and EgrMYB83 are master switches, able to activate biosynthesis of lignin, cellulose and hemicellulose in the SCWs of various cell types (Zhong

et al. 2007a; Zhong and Ye 2007; Zhong et al. 2008; McCarthy et al. 2009; Ohashi-Ito et al. 2010; Zhong and Ye 2012). The master switch TFs give us an indication of what cell types the *EgrSND2* promoter may be active in, but not of the biological pathways in which it is functional (i.e. lignin, cellulose or hemicellulose). The absence or presence of the binding of TFs with known functions in the cellulose, lignin and hemicellulose biosynthetic pathways has provided data towards elucidating the biological pathways in which EgrSND2 is involved. For example, EgrSND3 and EgrSND2 itself were shown to bind in this study. In *Arabidopsis,* AtSND2 and AtSND3 are able to activate the *AtCesA8* promoter (Zhong et al. 2008) providing evidence for their role in cellulose biosynthesis. EgrSND2 has also been observed binding to the *EgrCesA8* promoter *in vivo* (Creux et al., unpublished data). The qualitative GUS analysis and *E. grandis* transcriptome data agree with these findings (Figures 2, 3, 4 and 5). Cellulose is synthesised in most if not all plant tissues, thus the widespread GUS staining seen in the GUS analyses is not unexpected. In the transcriptome data, *EgrSND2* is most highly expressed in immature xylem (Figure 2). In these woody tissues, SCWs are being deposited. Thus, massive cellulose biosynthesis will be taking place due to the thickness and prevalence of their SCWs, and so we would expect a TF involved in cellulose biosynthesis to be highly expressed in them. Lignin is also synthesised in these tissues, but the lack of interaction of EgrKNAT7, EgrMYB52/54-A, EgrMYB52/54-B, EgrMYB85 and EgrMYB103 with the *EgrSND2* promoter in the Y1-H analyses suggests that it is not key for lignin biosynthesis. This is further supported by a lack of interaction of AtSND2 with *4CL1* in *Arabidopsis* (Zhong et al. 2008), a gene involved in the biosynthesis of monolignols, the monomers of lignin (Boerjan et al. 2003).

2.6 Conclusion

In conclusion, we have identified the tissues in which the 1.5 kb fragment of the *EgrSND2* promoter is active using qualitative GUS analysis and we have partially characterised the role of *EgrSND2* in the transcriptional network of SCW biosynthesis in *E. grandis* using Y1-

H. We have found that the EgrSND2 1.5 kb promoter fragment is not sufficient to express the reporter gene specifically in fibre cells. Qualitative GUS shows staining in many different tissues at different time points and appears to be slightly preferential in tissues which are fast growing or those with thick SCWs. This indicates a preference for expression in cells which are actively depositing and thickening SCWs. We were able to identify putative orthologs of a number of TFs important for SCW biosynthesis in *Arabidopsis,* and in doing so, have generated a catalogue of *E. grandis* SCW-related TFs that can be used for further analyses. Using Y1-H, we were also able to test whether a number of TFs important for SCW biosynthesis were able to bind to the *EgrSND2* promoter. A number of these TFs were able to bind, indicating the importance of *EgrSND2* to SCW biosynthesis in *E. grandis*. A summary of these interactions can be seen in Figure 12. We were able to determine that *EgrSND2* plays a role in SCW biosynthesis in many different cell types due to the GUS qualitative analysis and TFs such as EgrVND6, EgrMYB46 and EgrMYB83 binding to the *EgrSND2* promoter. We were also able to determine that EgrSND2 does not directly influence lignin biosynthesis due to the absence of binding of lignin biosynthesis associated TFs such as EgrKNAT7, EgrMYB52/54-A, EgrMYB52/54-B, EgrMYB85 and EgrMYB103. Through the Y1-H analysis, we were able to identify a number of interactions not previously seen in plants and gain an understanding of the position of *EgrSND2* in the transcriptional hierarchy of SCW biosynthesis. This research is an important stepping stone to understanding transcriptional regulation of SCW biosynthesis as a whole, as well as understanding SCW biosynthesis in the xylem of tree species. With further research, this may eventually lead to the ability to manipulate the transcriptional network to obtain trees with desirable traits to a number of industries, such as the paper and chemical cellulose industries.

2.7 Acknowledgements

The author would like to thank the following people for their contributions to the work: Ms Nicky Creux for providing immature xylem cDNA, and for technical assistance in creating the EgrMYB46, EgrKNAT7, EgrMYB103, EgrSND2 and EgrSND3 prey vectors; Mr Steven Hussey for technical assistance in creating the EgrMYB103, EgrSND2 and EgrSND3 prey vectors; Ms Minique de Castro for constructing the pMDC162 *EgrSND2 1.5 kb promoter::GUS* and *EgrSND3 1.5 kb promoter::GUS* reporter vectors; and Ms Annelize Pienaar for her general assistance in performing PCRs and DNA extractions.

2.8 Tables

Table 1: *Arabidopsis* transcription factors important for secondary cell wall biosynthesis identified from literature as targets for the study.

A.thaliana ID[a]	AT accession[b]	Relevance[c]	Reference[d]
AtSND2	AT4G28500	Upregulated by SND1	Zhong, Lee et al. 2008
AtMYB103	AT1G63910	Upregulated by SND1	Zhong, Lee et al. 2008
AtSND1	AT1G32770	Master regulator of SCW biosynthesis	Zhong, Demura et al. 2006; Zhong, Richardson et al. 2007
AtSND3	AT1G28470	Upregulated by SND1	Zhong, Lee et al. 2008
AtMYB83	AT3G08500	Downstream master switch for SCW biosynthesis	McCarthy, Zhong et al. 2009
At1g66810	AT1G66810	Upregulated by MYB46	Ko, Kim et al. 2009
At1g72220	AT1G72220	Upregulated in tension wood	Mizrachi et al., unpublished
AtMYB85	AT4G22680	Upregulated by SND1	Zhong and Ye 2009
AtVND7	AT1G71930	Involved in tracheary element formation	Kubo, Udagawa et al. 2005; Zhong, Lee et al. 2010
AtMYB52	AT1G17950	Upregulated by SND1 and MYB46	Ko, Kim et al. 2009; Ohashi-Ito, Oda et al. 2010
AtMYB54	AT1G73410	Upregulated by SND1 and MYB46	Ko, Kim et al. 2009; Ohashi-Ito, Oda et al. 2010
AtMYB46	AT5G12870	Downstream master switch for SCW biosynthesis	McCarthy, Zhong et al. 2009
AtVND6	AT5G62380	Involved in tracheary element formation	Kubo, Udagawa et al. 2005; Zhong, Lee et al. 2010
AtKNAT7	AT1G62990	Upregulated by SND1, upregulated in tension wood	Zhong and Ye 2009 ; Mizrachi et al., unpublished

[a]The gene name in *Arabidopsis thaliana*

[b]The AT accession number as seen in TAIR 10

[c]The relevance of the transcription factors to xylem biosynthesis

[d]The primary references in which this protein was found to be important for SCW biosynthesis

Table 2: Primers used for the amplification of the coding sequences of the *Eucalyptus grandis* orthologs of the *Arabidopsis* transcription factors selected from literature.

Ortholog amplified[a]	Primer Name[b]	Primer Sequence (5'-3')[c]	Tm (°C)[d]	Product size (bp)[e]
EgrZF1	Egr_v1_0.017796_F	ATGGAAAAGGACGCCTCGCCTCA	57	1053
	Egr_v1_0.017796_R	TCAACGCGAAGCCAGCATCA		
EgrZF2	Egr_v1_0.014222F	ATGCTTTCTTGTCTTCCAATTC	54	1239
	Egr_v1_0.014222_R	TTAGGAAGGTTGAGCTGAATC		
EgrKNAT7	Egr_KNAT7_F	ATGCAAGAGCCGAACTTGGCCATGA	56	930
	Egr_KNAT7_R	CTACCTCTTGCGCTTGGACT		
EgrMYB83	Egr_MYB83_F	ATGAGGAAGCCAAGTGAGAC	60	987
	Egr_MYB83_R	TTAGAAATCAAGAAAAGGGAAGGAGGGCAAAT		
EgrMYB46	Egr_MYB46_F	ATGGAGTCAAAGACAAGACGACCACCACCCT	57	1419
	Egr_MYB46_R	CTATTGAATAACCTGGAAATCAGCAGAA		
EgrSND1	Egr_v1_0.014865_F	ATGAACCTGTCCATAAACGGCCAGTC	57	1200
	Egr_v1_0.014865_R	TTATACCGACAAATGACGTAATGGGTCAGA		
EgrNST1	Egr_v1_0.015754_F	GAAGAGATGGACATGAATTTGTC	57	1152
	Egr_v1_0.015754_R	AATTATACTGACAAGTGGTGCAACGG		
EgrMYB52/54-A	Egr_v1_0.023389_F	ATGTGCACCAGAGGCCACTG	57	792
	Egr_v1_0.023389_R	CTAACAAGAGCTTCTGACCGATA		
EgrMYB52/54-B	Egr_v1_0.020698_F	ATGGACAATTCCAGACCTAACAA	50	921
	Egr_v1_0.020698_R	TCAAGATGTGATTCCTACCCCAAGAAA		
EgrMYB85	Egr_MYB85_F	ATGGGGAGGCAACC	50	906
	Egr_MYB85_R	CTAAAATATCTTGTCGGTCATC		
EgrVND7	Egr_VND7_F	ATGGAGTTGGAATCGTGTGTAC	57	936
	Egr_VND7_R	CTAGGAGTCTGGAAAGCACCCTAG		
EgrVND6	VND6_F	ATGAACACCTTTTCACGTGTC	56	1047
	VND6_R	TCACTTCCAGAGTTCAATT		

[a]The *Eucalyptus grandis* ortholog amplified by the primer pair
[b]The name of the primer
[c]The primer sequence in a 5' to 3' configuration
[d]The annealing temperature used in the PCR amplification
[e]The expected product size to be obtained from PCR with the primer pair

Table 3: Summary of the *Eucalyptus grandis* orthologs selected for yeast one-hybrid analysis.

A. thaliana ID[a]	*E. grandis* ID[b]	New *E. grandis* annotation[c]	Old *E. grandis* annotation[d]
AT1G66810	EgrZF1	Eucgr.F02796	Egrandis_v1_0.017796m
AT1G72220	EgrZF2	Eucgr.B00047	Egrandis_v1_0.014222m
AtKNAT7	EgrKNAT7	Eucgr.D01935	Egrandis_v1_0.020462m
AtMYB103	EgrMYB103	Eucgr.D01819	Egrandis_v1_0.020161m
AtMYB46	EgrMYB46	Eucgr.B03684	Egrandis_v1_0.013118m
AtMYB52/54	EgrMYB52/54-A	Eucgr.K02297	Egrandis_v1_0.023389m
AtMYB52/54	EgrMYB52/54-B	Eucgr.F04277	Egrandis_v1_0.020698m
AtMYB52/54	EgrMYB52/54-C	Eucgr.F02756	Egrandis_v1_0.016634m
AtMYB83	EgrMYB83	Eucgr.G03385	Egrandis_v1_0.019189m
AtMYB85	EgrMYB85	Eucgr.D02014	Egrandis_v1_0.021039m
AtNST1	EgrNST1	Eucgr.D01671	Egrandis_v1_0.015754m
AtSND1	EgrSND1	Eucgr.E01053	Egrandis_v1_0.014865m
AtSND2	EgrSND2	Eucgr.K01061	Egrandis_v1_0.021305m
AtSND3	EgrSND3	Eucgr.E03226	Egrandis_v1_0.020239m
AtVND6	EgrVND6	Eucgr.A02887	Egrandis_v1_0.017979m
AtVND7	EgrVND7	Eucgr.F02615	Egrandis_v1_0.019772m

[a]The gene name in *Arabidopsis thaliana*

[b]The name assigned to the *Eucalyptus grandis* ortholog in this study

[c]The current annotation/gene model in *E. grandis*

[d]The annotation/gene model from the previous version of the *E. grandis* genome

2.9 Figures

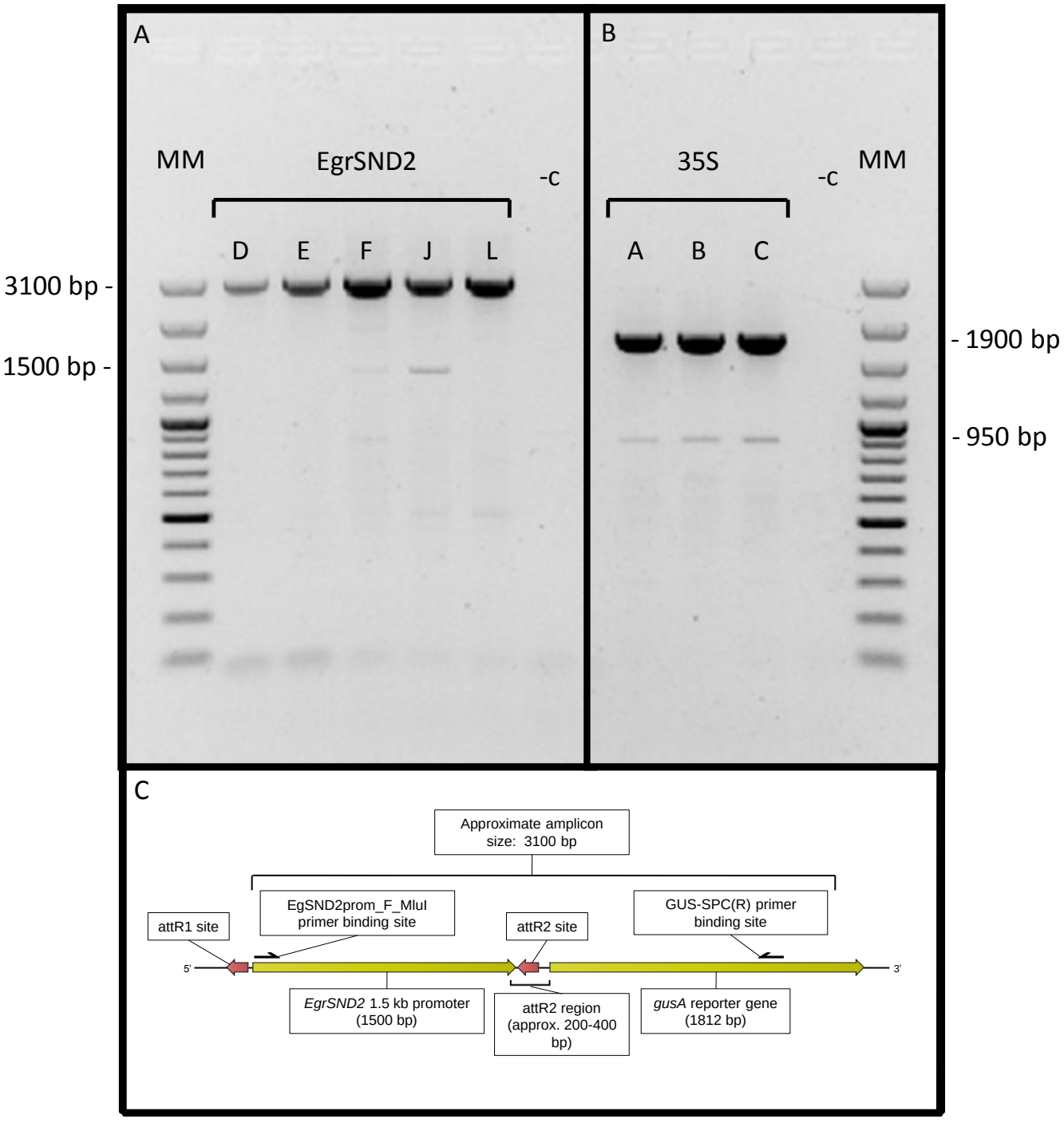


Figure 1: Agarose gel electrophoresis showing PCR products of both the promoter and *GUS* gene in all transgenic lines. (A) PCR using *EgrSND2* promoter_F and *GUS*_R primers. MM indicates 100 bp molecular weight standard; EgrSND2-D, -E, -F, -J and –L indicates transgenic lines carrying the *EgrSND2 1.5 kb promoter::GUS* constructs. (B) PCR using *35SCaMV* promoter_F and *GUS*_R primers. 35S-A, -B, and -C indicates the transgenic lines carrying the *35S CaMV promoter::GUS* constructs. Wild-type template control is indicated by "-c". (C) Schematic representation of the amplicon size expected from the *EgrSND2 1.5 kb promoter:: GUS* plant lines using the EgSND2_F_MluI and GUS-SPC(R) promoter binding sites.

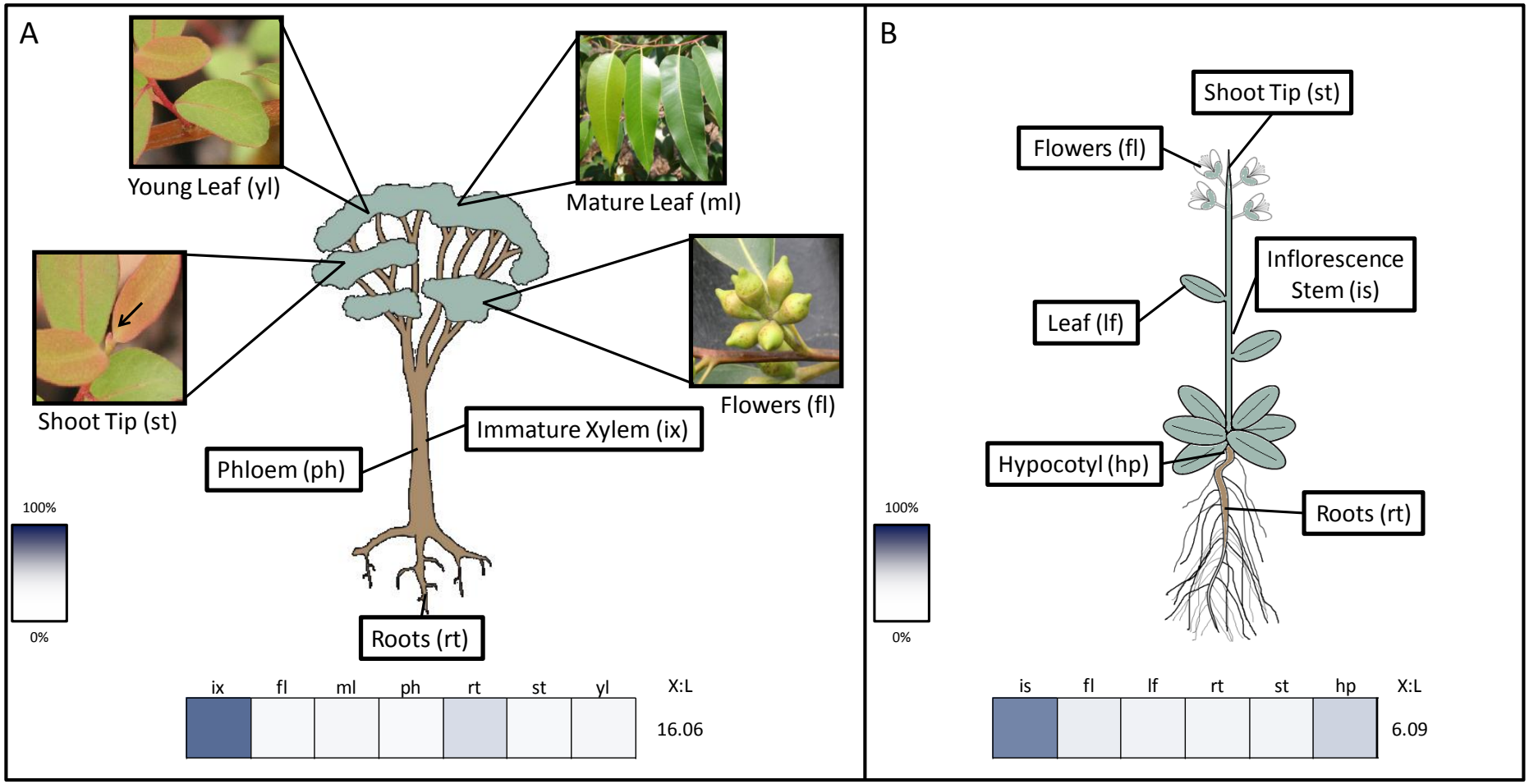


Figure 2: Tissue-specific expression of EgrSND2 and AtSND2 in *E. grandis* and *Arabidopsis*, respectively. (A) Average transcript expression levels from three *E. grandis* trees of *EgrSND2* (Hefer et al., in preparation) represented as a percentage of the total. The immature xylem (ix), flowers (fl), mature leaf (mf), phloem (ph), roots (rt), shoot tips (st) and young leaf (yl) are represented. X:L indicates the xylem to leaf ratio of expression, calculated by dividing the percentage of transcripts from ix by the average of the percentage of transcripts from both ml and yl. (B) Transcript expression levels of AT4G28500 (*AtSND2*) in various *A. thaliana* tissues, adapted from data obtained from the Genevestigator database (Zimmermann et al. 2004) and represented as a percentage of the total. The inflorescence stem (is), flowers (fl), leaf (lf), roots (rt), shoot tip (st) and hypocotyl (hp) are represented. X:L indicates the xylem to leaf ratio of expression, calculated by dividing the average percentage expression from both is and hp by the percentage expression from lf.

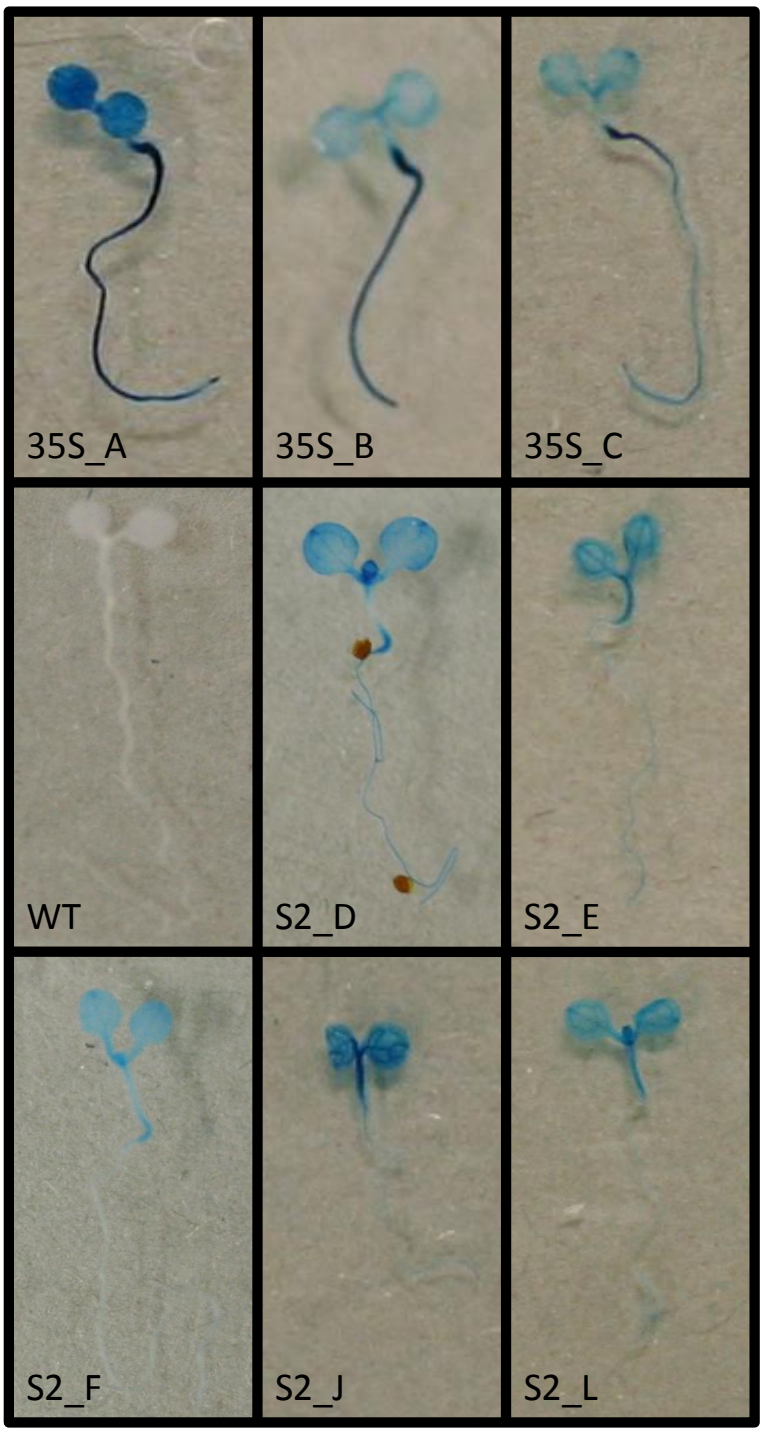


Figure 3: One-week-old *GUS*-stained T2 plants. S2_D, S2_E, S2_F, S2_J and S2_L are the representative *EgrSND2 1.5kb promoter::GUS T2* plant lines, selected to show the consensus phenotype. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate *35S CaMV* promoter positive controls. For each plant line *n*=3. The replicates, as well as non-selected *EgrSND2 1.5 kb promoter::GUS* lines can be seen in Supplementary Figure 5.

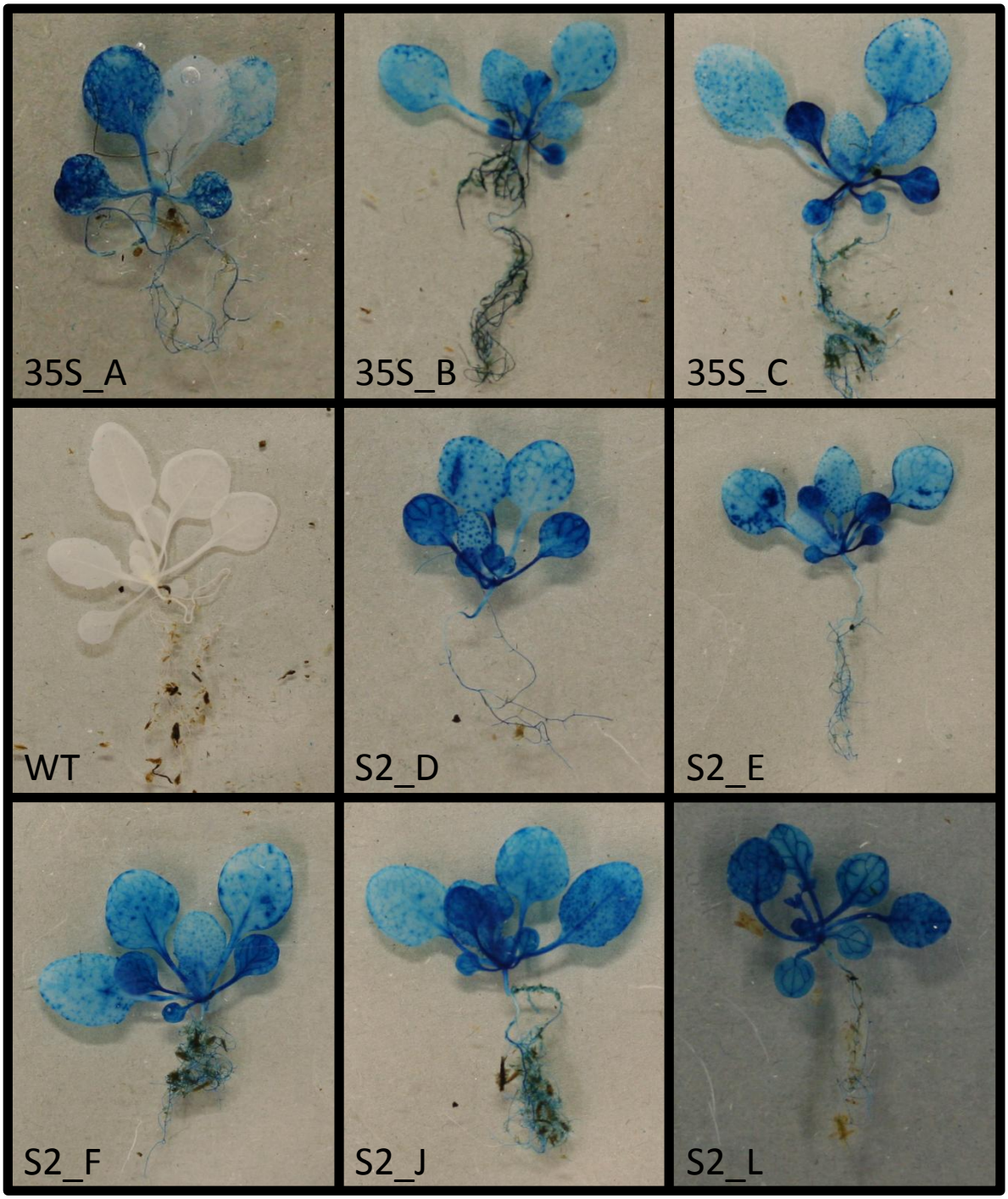


Figure 4: Three-week-old *GUS*-stained T2 plants. S2_D, S2_E, S2_F, S2_J and S2_L are representative of *EgrSND2 1.5kb promoter::GUS* T2 plant lines, selected to show the consensus phenotype. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate 35S CaMV promoter positive controls. For each plant line *n*=3. Additional replicates, as well as non-selected *EgrSND2 1.5 kb promoter::GUS* lines can be seen in Supplementary Figure 6.

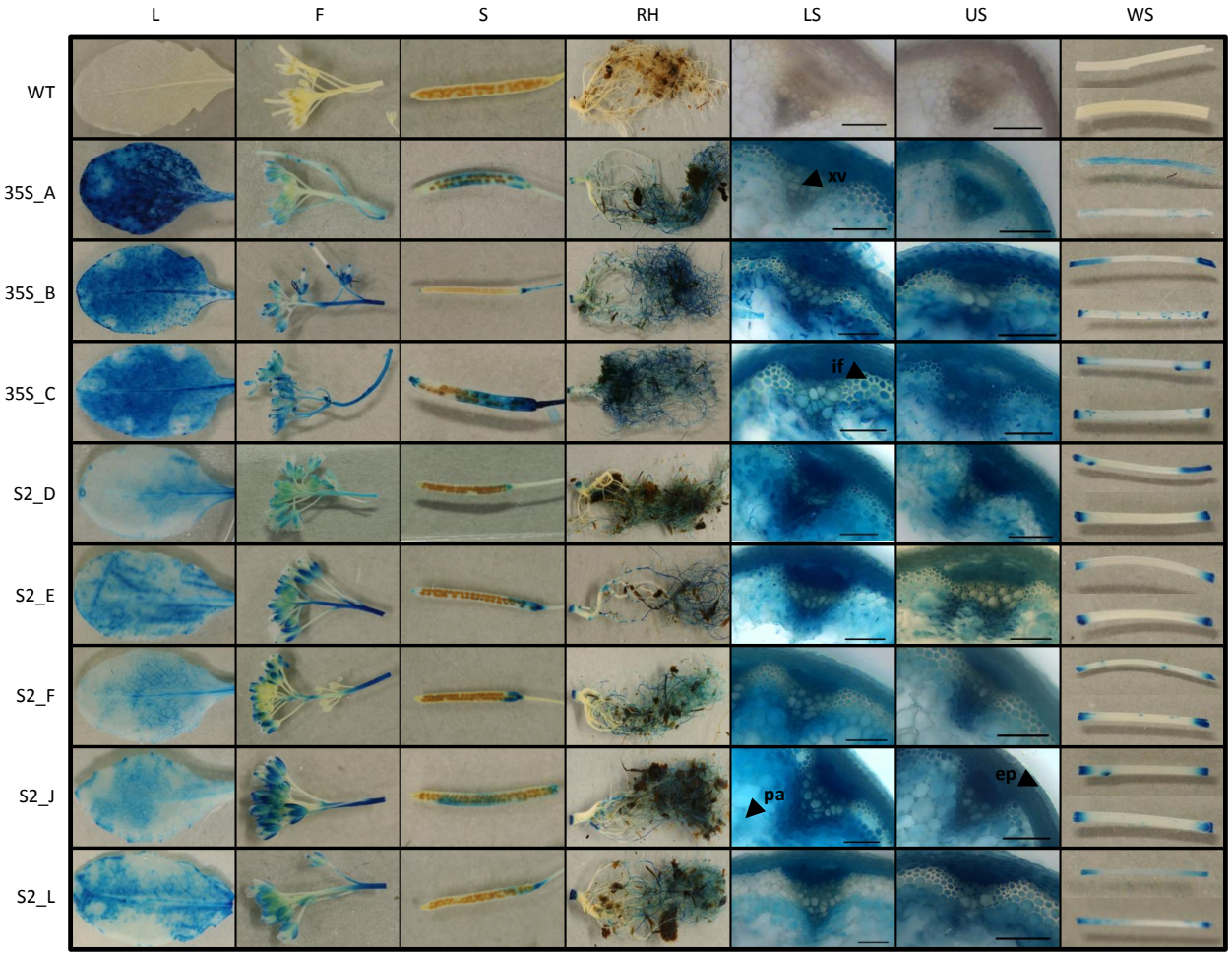


Figure 5: Six-week-old T2 plant *GUS*-stained tissues. S2_D, S2_E, S2_F, S2_J and S2_L are representative *EgrSND2 1.5kb promoter::GUS T2* plant lines, selected to show the consensus phenotype. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate 35S CaMV promoter positive controls. The xylem vessels (xv), interfascicular fibres (if), parenchyma (pa) and epidermal cell layer (ep) are represented in the above stem images with arrows. For each plant line *n*=3. The contamination seen in the root samples is leftover soils from the Jiffies™ in which the plants were grown. Additional replicates, as well as non-selected *EgrSND2 1.5 kb promoter::GUS* lines can be seen in Supplementary Figure 7-12.

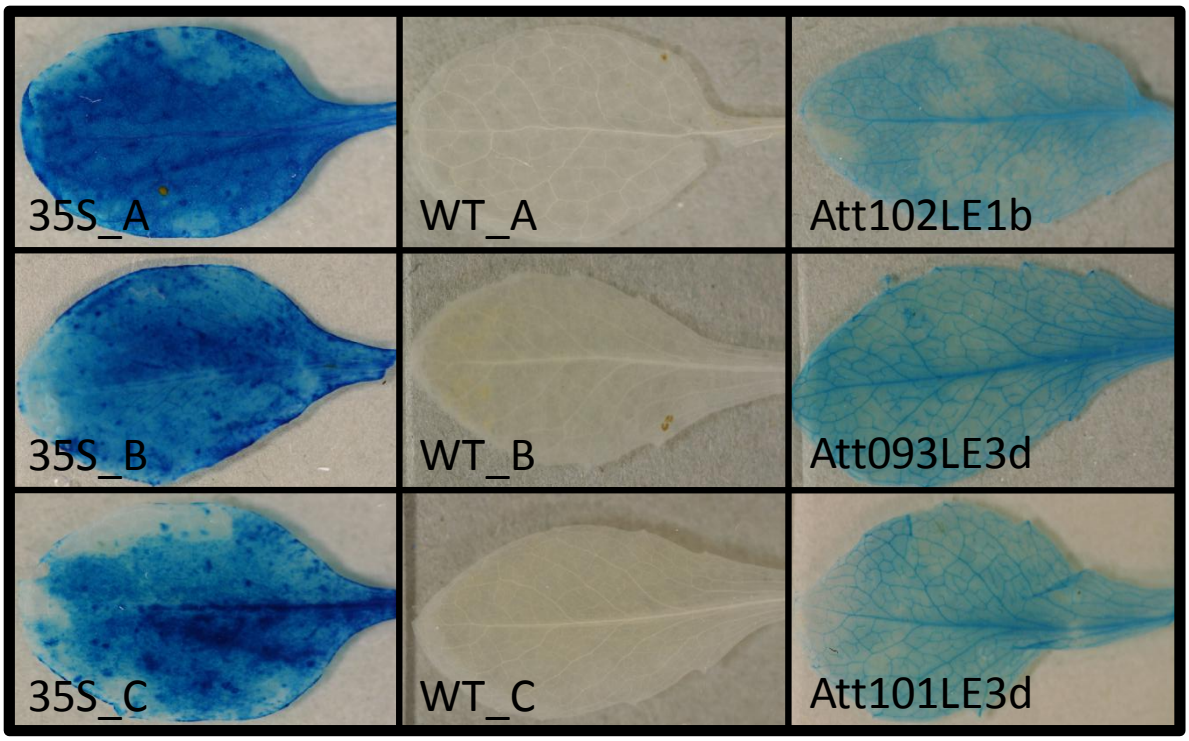


Figure 6: GUS-stained six-week-old plant leaves showing vascular specific expression. Shown with permission of Creux et al., (unpublished data). Plants were transformed with 2 kb *EgrCesA8 promoter::GUS* constructs. 35S_A, 35S_B and 35S_C indicate *35S CaMV* positive controls. WT_A, WT_B and WT_C indicate wild-type negative controls. Att102LE1b, Att093LE3d and Att101LE3d are *2 kb EgrCesA8 promoter::GUS* plant lines showing a vascular-specific expression pattern.

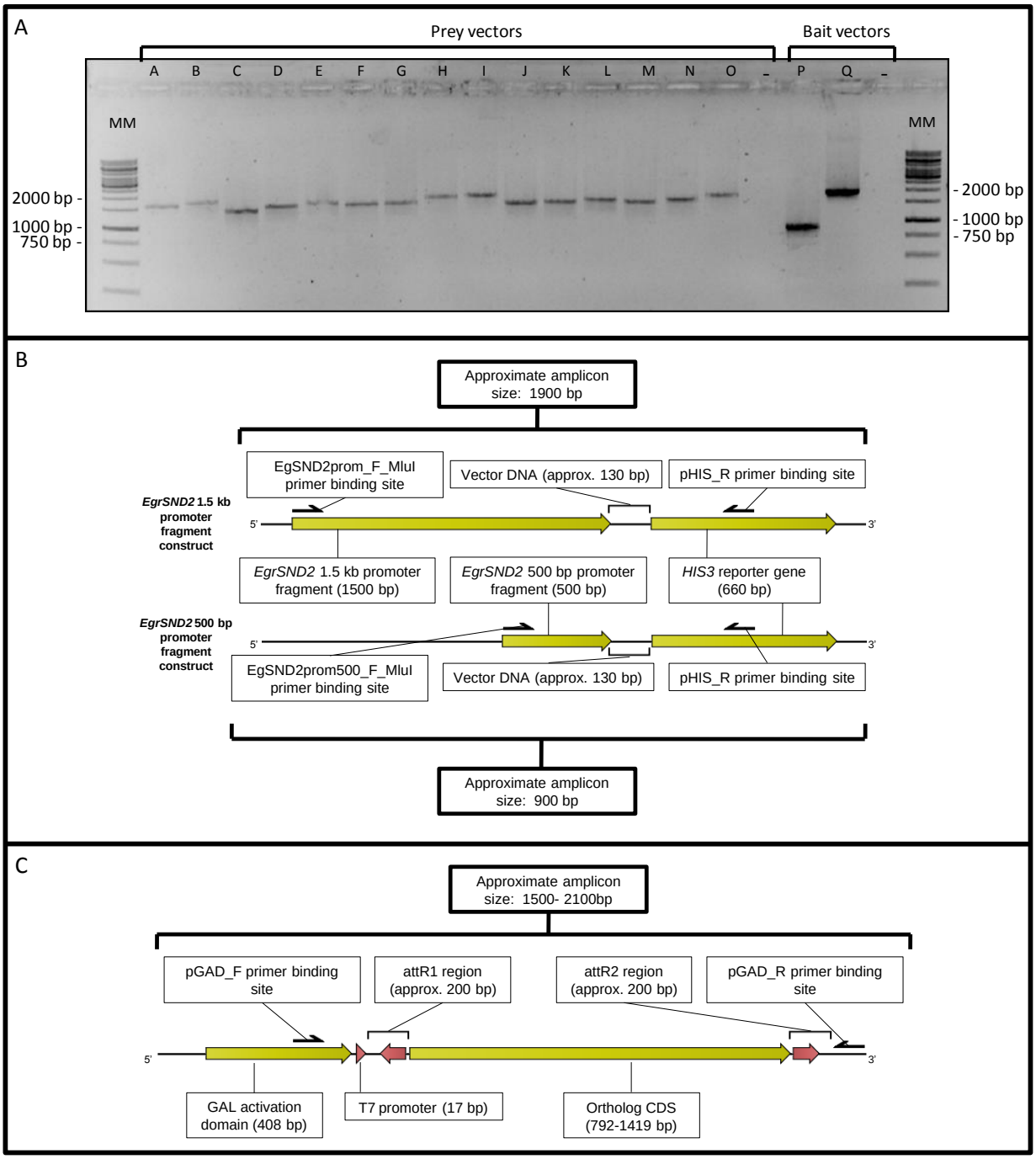


Figure 7: Bait and prey vector amplicons. (A) Agarose gel showing PCR products of the 500 bp and 1.5 kb *EgrSND2* promoter fragments from the pHIS2.1 bait vectors and the PCR amplification of the *Eucalyptus grandis* TF coding sequences from the pDEST-GADT7 prey vector using vector-specific forward and reverse primers (pGAD_F and pGAD_R respectively) (Supplementary Table 1). MM indicates Generuler™ 1 kb ladder. For the prey vectors, A = EgrKNAT7, B = EgrMYB46, C = EgrMYB52/54-A, D = EgrMYB52/54-B, E = EgrMYB83, F = EgrMYB85, G = EgrMYB103, H = EgrNST1, I = EgrSND1, J = EgrSND2, K = EgrSND3, L = EgrVND6, M = EgrVND7, N = EgrZF1 and O = EgrZF2. For the bait vectors, P = 500 bp promoter fragment bait and Q = 1.5 kb promoter fragment bait. Negative control using sterile water as template is indicated by "-". **(B)** Schematic representation of the amplicon size expected from PCR amplification of the 500 bp and 1.5 kb *EgrSND2* promoter fragments from the pHIS2.1 vector using a promoter-specific forward primer and the pHIS_R reverse primer. **(C)** Schematic representation of the amplicon size expected from PCR amplification of the *E. grandis* ortholog CDS sequences from the pDEST-GADT7 vector using the pGAD_F primer and pGAD_R reverse primer.

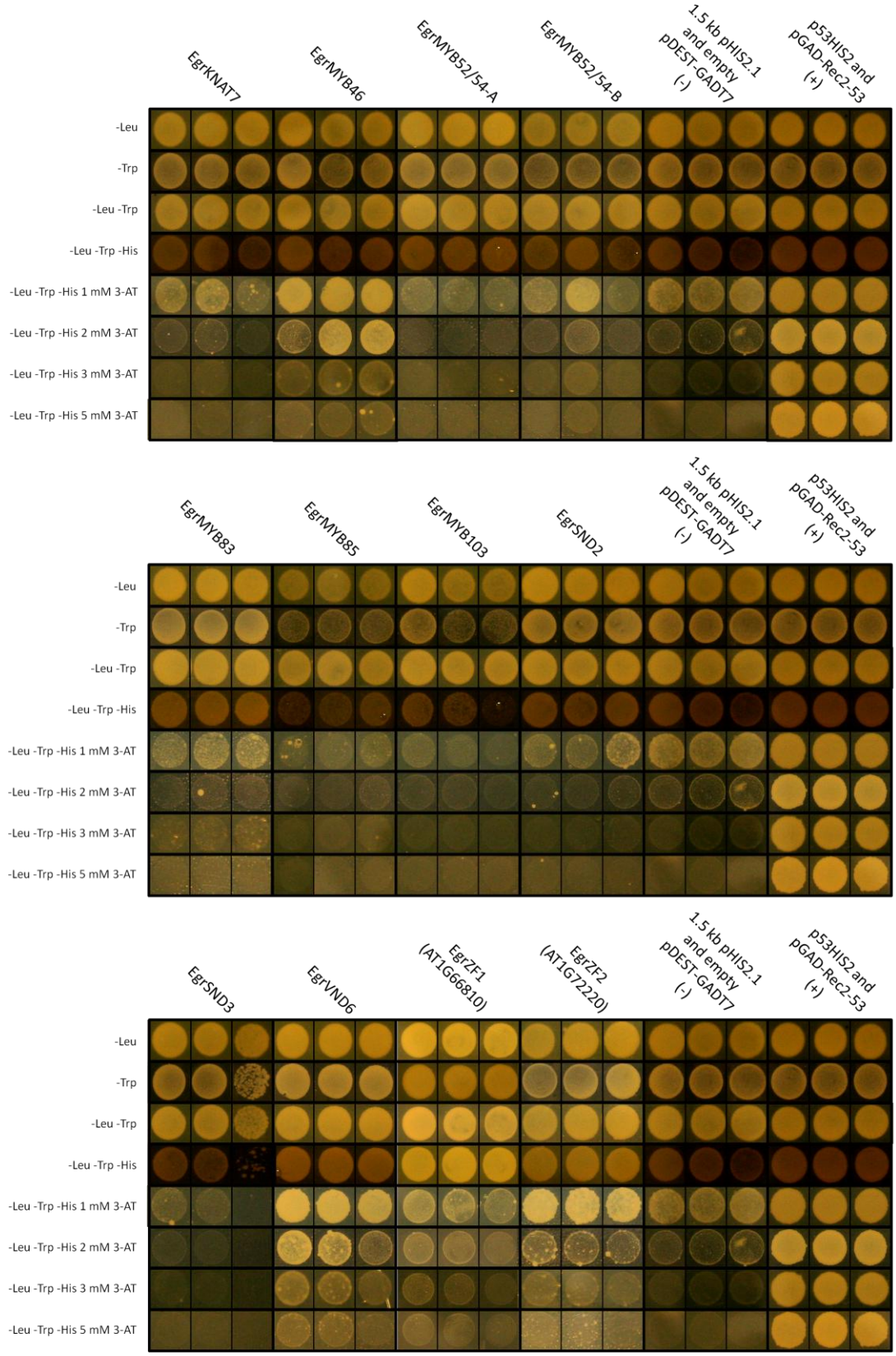


Figure 8: Yeast one-hybrid analysis of the 1.5 kb *EgrSND2* promoter fragment. The selective dropout media on which the yeast colonies were screened are listed to the left of each figure. Proteins which are being screened for interaction with the promoter fragment are listed on the top of each figure The negative control (-) is an empty pDEST-GADT7 plasmid screened against the 1.5 kb fragment. The positive control (+) is that recommended in the Matchmaker™ One-Hybrid Library Construction & Screening kit (Clontech), pGAD-Rec2-53 prey plasmid screened against p53 HIS2 bait plasmid. Three replicates are shown for each interaction.

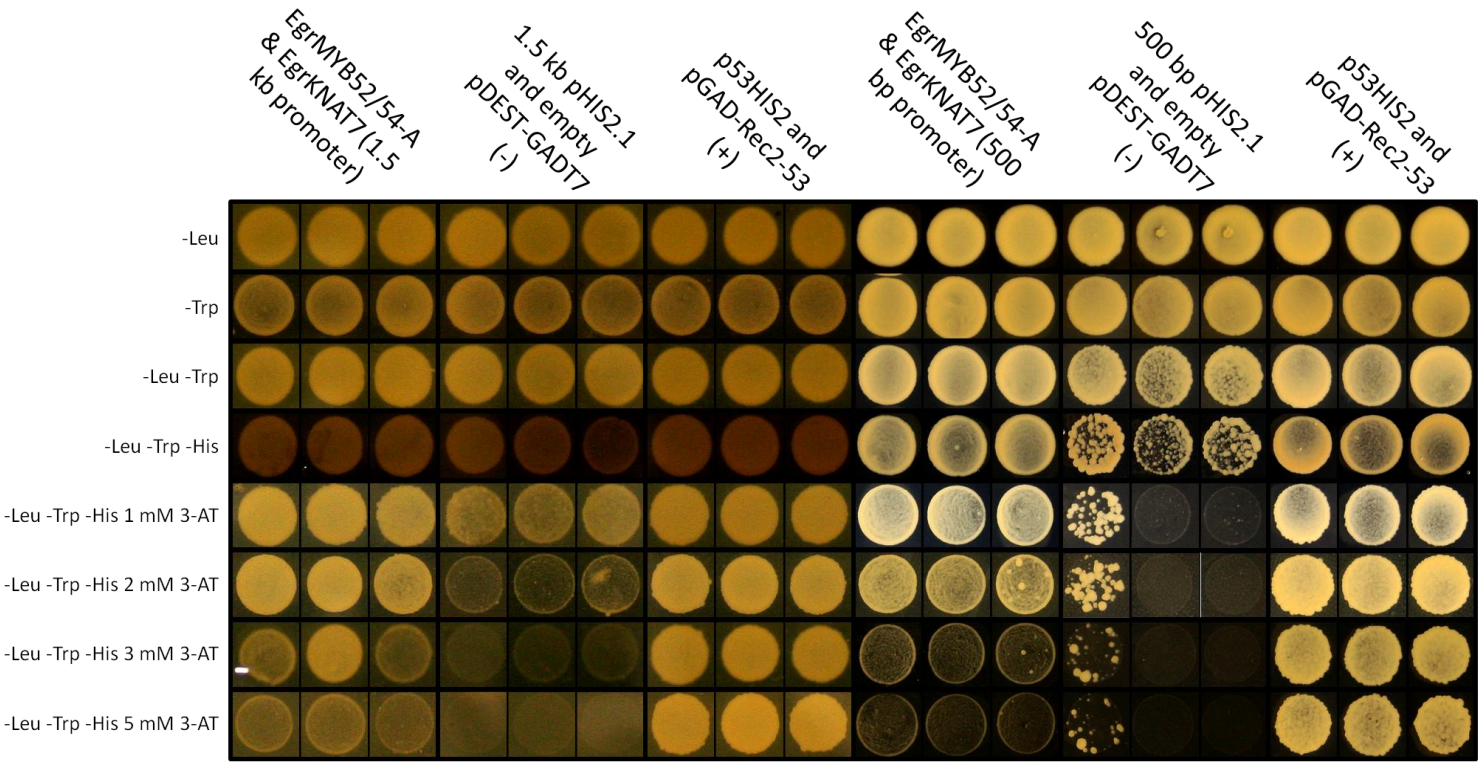


Figure 9: Yeast one-hybrid analysis of the 1.5 kb and 500 bp *EgrSND2* promoter fragments cotransformed with EgrMYB52/54-A and EgrKNAT7 prey constructs. The selective dropout media on which the yeast colonies were screened are listed to the left of each figure. The proteins which are being screened for interaction with the promoter fragment are listed on the top of each figure The negative control (-) is an empty pDEST-GADT7 plasmid screened against the 1.5 kb fragment. The positive control (+) is that recommended in the Matchmaker™ One-Hybrid Library Construction & Screening kit (Clontech), pGAD-Rec2-53 prey plasmid screened against p53HIS2 bait plasmid. Three replicates are shown for each interaction.

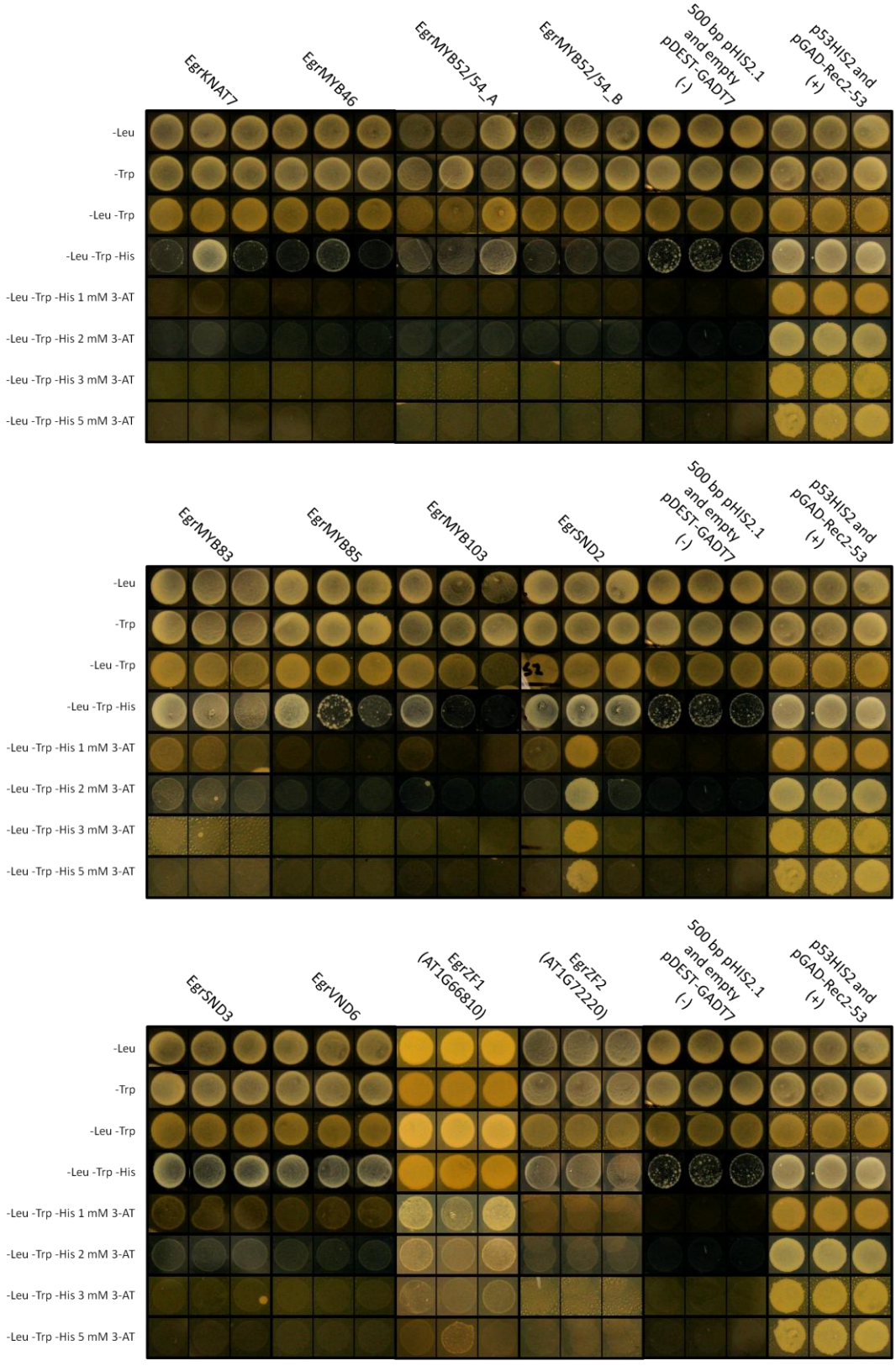


Figure 10: Yeast one-hybrid analysis of the 500 bp *EgrSND2* promoter fragment. The selective dropout media on which the yeast colonies were screened are listed to the left of each figure. The proteins which are being screened for interaction with the promoter fragment are listed on the top of each figure The negative control (-) is an empty pDEST-GADT7 plasmid screened against the 500 bp fragment. The positive control (+) is that recommended in the Matchmaker™ One-Hybrid Library Construction & Screening kit (Clontech), pGAD-Rec2-53 prey plasmid screened against p53HIS2 bait plasmid. Three replicates are shown for each interaction.

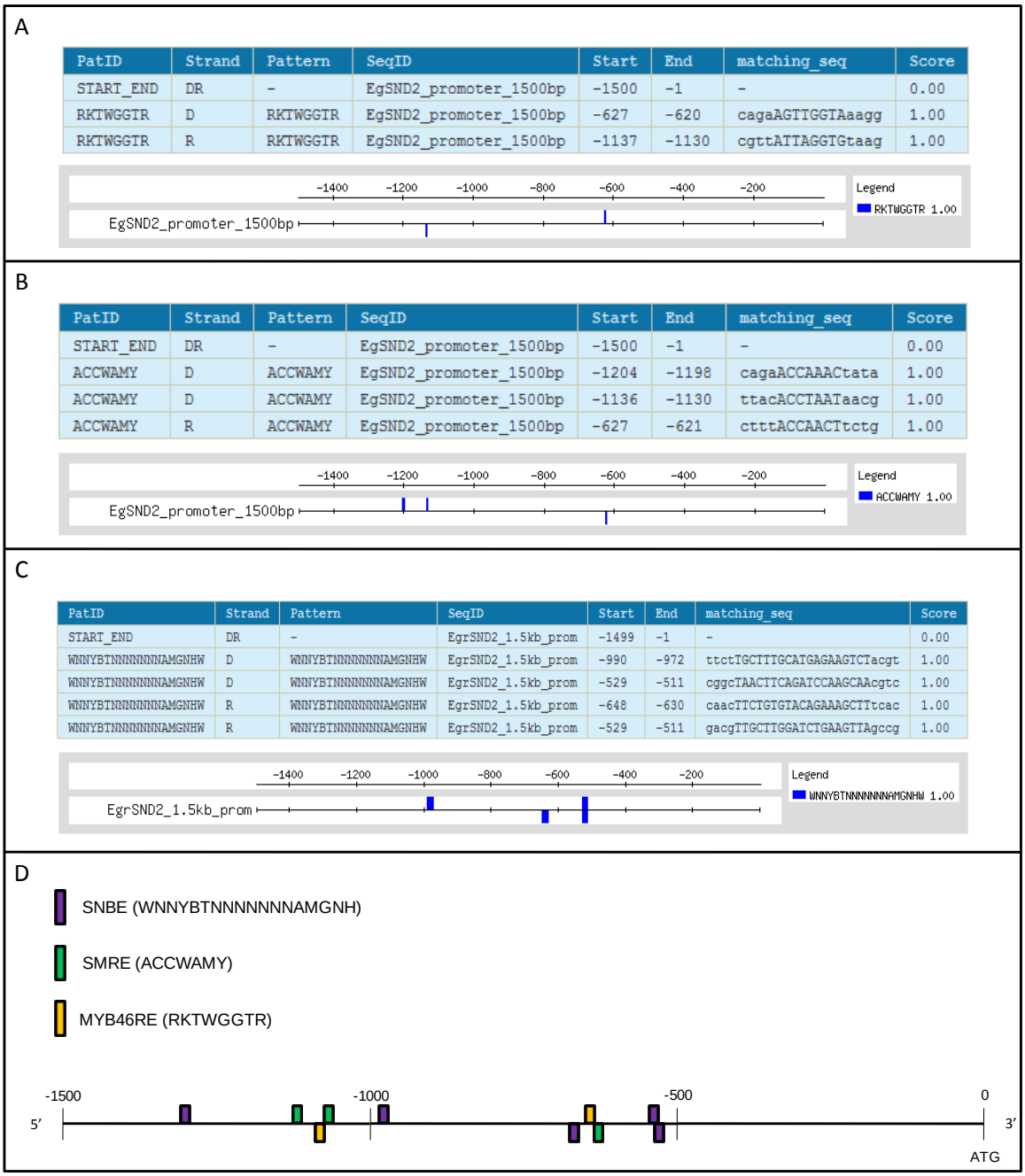


Figure 11: Tables and *cis*-element maps depicting the position and number of putative *cis*-elements in the *EgrSND2* 1.5 kb promoter fragment using the string pattern match function of the RSA tool suite (Thomas-Chollier et al. 2008). In the table, "PatID" indicates the consensus sequence of the element being searched for. "Strand" indicates orientation of the motif, being either direct (forward) or indirect (reverse) orientation. "SeqID" indicates the name of the sequence which is being searched for the motif. "Start" and "End" indicate the first and last base positions of the putative motifs, respectively. "matching_seq" indicates the exact sequence of the identified putative motif. "Score" indicates a significance score assigned by the program, with 0 and 1 the minimum and maximum, respectively. (A) Location of putative M46RE elements (Kim et al. 2012). (B) Location of putative SMRE elements (Zhong and Ye 2012). (C) Location of putative SNBE elements (Zhong et al. 2010b; McCarthy et al. 2011). (D) Schematic diagram of the 1.5 kb *EgrSND2* promoter fragment depicting a summary of the locations of putative *cis*-elements identified using the string pattern match function of the RSA tool suite (Thomas-Chollier et al. 2008). The coloured blocks indicate cis-elements. Blocks above the line indicate elements in direct orientation and blocks below the line indicate elements in indirect orientation. The key is located in the top left corner, with purple blocks representing SNBE elements, green blocks representing SMRE elements and yellow blocks indicating MYB46RE elements. Base positions upstream of the ATG are represented in intervals of 500 bp above the schematic. Position "0" is the location of the start codon.

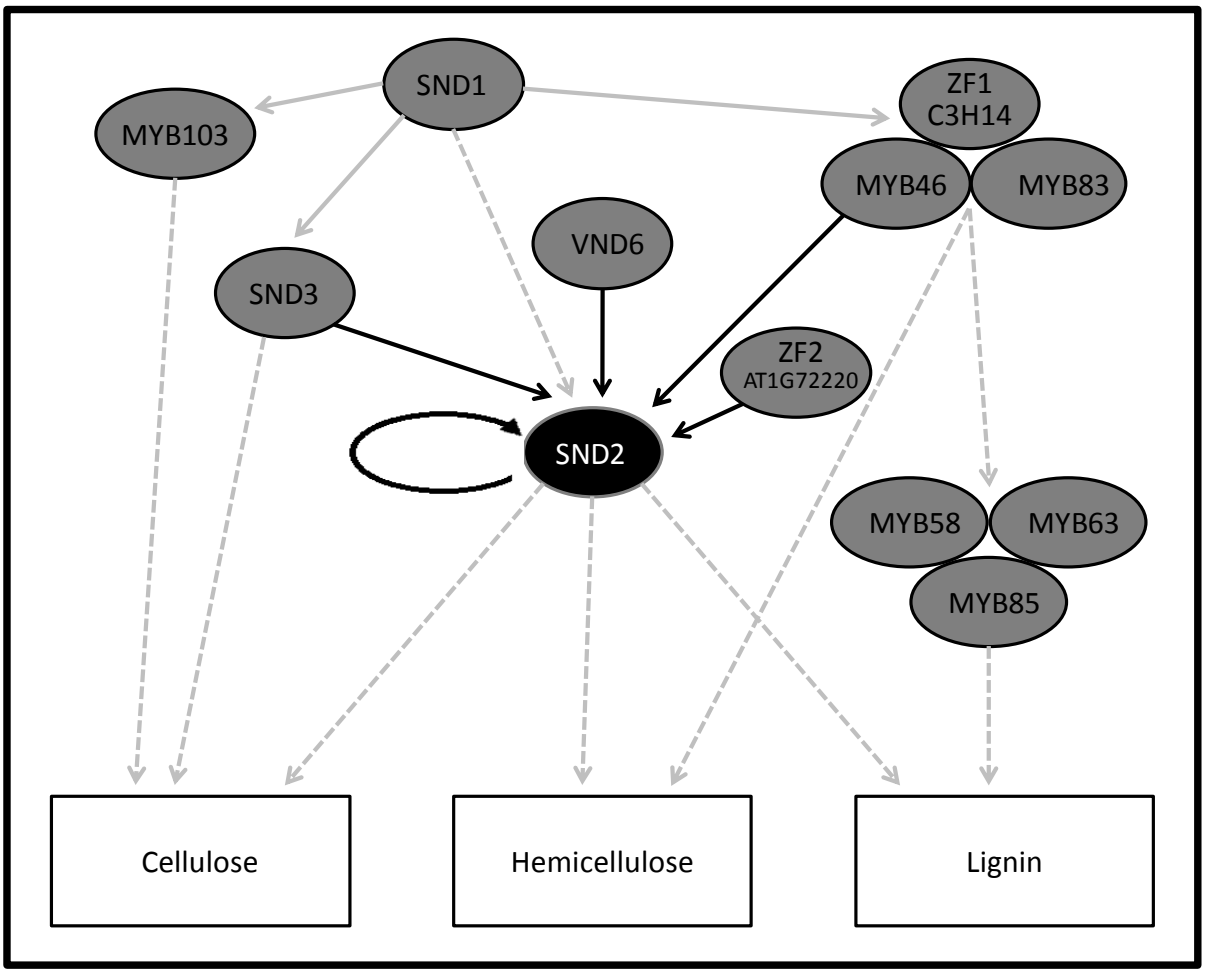


Figure 12: Adapted summary (Hussey et al. 2011) of the transcriptional control of SCW biosynthesis, modified to represent confirmed interactions that occur in both *Arabidopsis thaliana* and *Eucalyptus grandis*. Black lines indicate direct interactions identified in this study using *E. grandis* orthologs of the TFs. Grey lines indicate data from *A. thaliana*. Solid lines indicate direct interactions. Dotted lines indicate interactions without direct binding. The looped line around SND2 indicates a possible self-regulatory mechanism.

2.10 References

Alon, U. (2007) Network motifs: theory and experimental approaches. *Nature Reviews Genetics* 8: 450-461.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215: 403-410.

Badescu, G.O. and Napier, R.M. (2006) Receptors for auxin: will it all end in TIRs? *Trends in Plant Science* 11: 217-223.

Bergmann, D.C., Lukowitz, W. and Somerville, C.R. (2004) Stomatal development and pattern controlled by a MAPKK kinase. *Science* 304: 1494-1497.

Boerjan, W., Ralph, J. and Baucher, M. (2003) Lignin biosynthesis. *Annual Review of Plant Biology* 54: 519-546.

Cassan-Wang, H., Goué, N., Saidi, M.N., Legay, S., Sivadon, P., Goffner, D., et al. (2013) Identification of novel transcription factors regulating secondary cell wall formation in *Arabidopsis*. *Frontiers in plant science* 4: 189.

Chen, C.-C., Fu, S.-F., Lee, Y.-I., Lin, C.-Y., Lin, W.-C. and Huang, H.-J. (2012) Transcriptome analysis of age-related gain of callus-forming capacity in *Arabidopsis* hypocotyls. *Plant and cell physiology* 53: 1457-1469.

Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal* 16: 735-743.

Curtis, M.D. and Grossniklaus, U. (2003) A gateway cloning vector set for high-throughput functional analysis of genes *in planta*. *Plant Physiology* 133: 462-469.

De Micco, V., Ruel, K., Joseleau, J.-P., Grima-Pettenati, J. and Aronne, G. (2012) Xylem anatomy and cell wall ultrastructure of *Nicotiana Tabacum* after lignin genetic modification through transcriptional activator EgMYB2. *IAWA Journal-International Association of Wood Anatomists* 33: 269.

Demura, T. and Fukuda, H. (2007) Transcriptional regulation in wood formation. *Trends in Plant Science* 12: 64-70.

Du, J. and Groover, A. (2010) Transcriptional regulation of secondary growth and wood formation. *Journal of Integrated Plant Biology* 52: 17-27.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792-1797.

Goicoechea, M., Lacombe, E., Legay, S., Mihaljevic, S., Rech, P., Jauneau, A., et al. (2005) EgMYB2, a new transcriptional activator from *Eucalyptus* xylem, regulates secondary cell wall formation and lignin biosynthesis. *The Plant Journal* 43: 553-567.

Gorshkova, T., Brutch, N., Chabbert, B., Deyholos, M., Hayashi, T., Lev-Yadun, S., et al. (2012) Plant fiber formation: state of the art, recent and expected progress, and open questions. *Critical Reviews in Plant Sciences* 31: 201-228.

Grattapaglia, D. and Kirst, M. (2008) *Eucalyptus* applied genomics: from gene sequences to breeding tools. *New Phytologist* 179: 911-929.

Grattapaglia, D., Plomion, C., Kirst, M. and Sederoff, R.R. (2009) Genomics of growth traits in forest trees. *Current Opinion in Plant Biology* 12: 148-156.

Hussey, S., Mizrachi, E., Spokevicius, A., Bossinger, G., Berger, D. and Myburg, A. (2011) *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: 173.

Jefferson, R.A., Kavanagh, T.A. and Bevan, M.W. (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO Journal* 6: 3901-3907.

Katoh, K., Misawa, K., Kuma, K.i. and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30: 3059-3066.

Kim, W.C., Ko, J.H. and Han, K.H. (2012) Identification of a *cis*-acting regulatory motif recognized by MYB46, a master transcriptional regulator of secondary wall biosynthesis. *Plant Molecular Biology* 78: 489-501.

Ko, J.H., Kim, W.C. and Han, K.H. (2009) Ectopic expression of MYB46 identifies transcriptional regulatory genes involved in secondary wall biosynthesis in *Arabidopsis*. *The Plant Journal* 60: 649-665.

Ko, J.H., Yang, S.H., Park, A.H., Lerouxel, O. and Han, K.H. (2007) ANAC012, a member of the plant-specific NAC transcription factor family, negatively regulates xylary fiber development in *Arabidopsis thaliana*. *The Plant Journal* 50: 1035-1048.

Kubo, M., Udagawa, M., Nishikubo, N., Horiguchi, G., Yamaguchi, M., Ito, J., et al. (2005) Transcription switches for protoxylem and metaxylem vessel formation. *Genes & Development* 19: 1855-1860.

Le, B.H., Cheng, C., Bui, A.Q., Wagmaister, J.A., Henry, K.F., Pelletier, J., et al. (2010) Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. *Proceedings of the National Academy of Sciences* 107: 8063-8070.

Legay, S., Lacombe, E., Goicoechea, M., Briere, C., Séguin, A., Mackay, J., et al. (2007) Molecular characterization of *EgMYB1*, a putative transcriptional repressor of the lignin biosynthetic pathway. *Plant Science* 173: 542-549.

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

Li, E., Bhargava, A., Qiang, W., Friedmann, M.C., Forneris, N., Savidge, R.A., et al. (2012) The Class II KNOX gene *KNAT7* negatively regulates secondary wall formation in *Arabidopsis* and is functionally conserved in *Populus*. *New Phytologist* 194: 102-115.

Lucas, W.J., Groover, A., Lichtenberger, R., Furuta, K., Yadav, S.R., Helariutta, Y., et al. (2013) The plant vascular system: evolution, development and functions. *Journal of Integrative Plant Biology* 4: 294-388.

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 and Cell Physiology* 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 Signaling and Behaviour* 6: 1282-1285.

Mellerowicz, E.J., Baucher, M., Sundberg, B. and Boerjan, W. (2001) Unravelling cell wall formation in the woody dicot stem. *Plant Molecular Biology* 47: 239-274.

Mellerowicz, E.J. and Sundberg, B. (2008) Wood cell walls: biosynthesis, developmental dynamics and their implications for wood properties. *Current Opinion Plant Biology* 11: 293-300.

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.

Nagpal, P., Ellis, C.M., Weber, H., Ploense, S.E., Barkawi, L.S., Guilfoyle, T.J., et al. (2005) Auxin response factors ARF6 and ARF8 promote jasmonic acid production and flower maturation. *Development* 132: 4107-4118.

Nakano, Y., Nishikubo, N., Goué, N., Ohtani, M., Yamaguchi, M., Katayama, Y., et al. (2010) MYB transcription factors orchestrating the developmental program of xylem vessels in *Arabidopsis* roots. *Plant Biotechnology* 27: 267-272.

Ohashi-Ito, K., Oda, Y. and Fukuda, H. (2010) Arabidopsis VASCULAR-RELATED NAC-DOMAIN6 directly regulates the genes that govern programmed cell death and secondary wall formation during xylem differentiation. *Plant Cell* 22: 3461-3473.

Ohlrogge, J., Allen, D., Berguson, B., Dellapenna, D., Shachar-Hill, Y. and Stymne, S. (2009) Energy. Driving on biomass. *Science* 324: 1019-1020.

Öhman, D., Demedts, B., Kumar, M., Gerber, L., Gorzsás, A., Goeminne, G., et al. (2013) MYB103 is required for FERULATE-5-HYDROXYLASE expression and syringyl lignin biosynthesis in Arabidopsis stems. *The Plant Journal* 73: 63-76.

Patzlaff, A., McInnis, S., Courtenay, A., Surman, C., Newman, L.J., Smith, C., et al. (2003) Characterisation of a pine MYB that regulates lignification. *The Plant Journal* 36: 743-754.

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

Poethig, R.S. (1997) Leaf morphogenesis in flowering plants. *The Plant Cell* 9: 1077.

Ramirez, V., Agorio, A., Coego, A., Garcia-Andrade, J., Hernandez, M.J., Balaguer, B., et al. (2011) MYB46 modulates disease susceptibility to *Botrytis cinerea* in *Arabidopsis*. *Plant Physiology* 155: 1920-1935.

Ringli, C. (2010) Monitoring the outside: cell wall-sensing mechanisms. *Plant Physiology* 153: 1445-1452.

Schoof, H., Lenhard, M., Haecker, A., Mayer, K.F.X., Jürgens, G. and Laux, T. (2000) The stem cell population of Arabidopsis shoot meristems is maintained by a regulatory loop between the *CLAVATA* and *WUSCHEL* genes. *Cell* 100: 635-644.

Schuetz, M., Smith, R. and Ellis, B. (2013) Xylem tissue specification, patterning, and differentiation mechanisms. *Journal of experimental botany* 64: 11-31.

Shangguan, X.X., Xu, B., Yu, Z.X., Wang, L.J. and Chen, X.Y. (2008) Promoter of a cotton fibre MYB gene functional in trichomes of *Arabidopsis* and glandular trichomes of tobacco. *Journal of Experimental Botany* 59: 3533-3542.

Shen, H., Yin, Y., Chen, F., Xu, Y. and Dixon, R. (2009) A bioinformatic analysis of NAC genes for plant cell wall development in relation to lignocellulosic bioenergy production. *Bioenergy Research* 2: 217-232.

Spitzer-Rimon, B., Farhi, M., Albo, B., Cna'ani, A., Ben Zvi, M.M., Masci, T., et al. (2012) The R2R3-MYB–Like regulatory factor EOBI, acting downstream of EOBII, regulates scent production by activating ODO1 and structural scent-related genes in *Petunia*. *The Plant Cell Online* 24: 5089-5105.

Thomas-Chollier, M., Sand, O., Turatsinze, J.V., Janky, R., Defrance, M., Vervisch, E., et al. (2008) RSAT: regulatory sequence analysis tools. *Nucleic Acids Research* 36: W119-127.

Timell, T. (1986) Chemical properties of compression wood. *Compression wood in gymnosperms* 1: 289-408.

Van Aken, O., Zhang, B., Law, S., Narsai, R. and Whelan, J. (2013) AtWRKY40 and AtWRKY63 modulate the expression of stress-responsive nuclear genes encoding mitochondrial and chloroplast proteins. *Plant Physiology* 162: 254-271.

Van Beveren, K.S., Spokevicius, A.V., Tibbits, J., Wang, Q. and Bossinger, G. (2006) Transformation of cambial tissue *in vivo* provides an efficient means for induced somatic sector analysis and gene testing in stems of woody plant species. *Functional plant biology* 33: 629-638.

Wang, D., Guo, Y., Wu, C., Yang, G., Li, Y. and Zheng, C. (2008) Genome-wide analysis of CCCH zinc finger family in *Arabidopsis* and rice. *BMC Genomics* 9: 44.

Werker, E. (2000) Trichome diversity and development. *Advances in Botanical Research* 31: 1-35.

Winzell, A., Aspeborg, H., Wang, Y. and Ezcurra, I. (2010) Conserved CA-rich motifs in gene promoters of Pt×tMYB021-responsive secondary cell wall carbohydrate-active enzymes in *Populus*. *Biochemical and Biophysical Research Communications* 394: 848-853.

Yamaguchi, M., Goue, N., Igarashi, H., Ohtani, M., Nakano, Y., Mortimer, J.C., et al. (2010) VASCULAR-RELATED NAC-DOMAIN6 and VASCULAR-RELATED NAC-DOMAIN7 effectively induce transdifferentiation into xylem vessel elements under control of an induction system. *Plant Physiology* 153: 906-914.

Yu, H. and Gerstein, M. (2006) Genomic analysis of the hierarchical structure of regulatory networks. *Proceedings of the National Academy of Sciences* 103: 14724-14731.

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 and Cell Physiology* 52: 1856-1871.

Zhong, R., Lee, C. and Ye, Z.H. (2010a) Evolutionary conservation of the transcriptional network regulating secondary cell wall biosynthesis. *Trends in Plant Science* 15: 625-632.

Zhong, R., Lee, C. and Ye, Z.H. (2010b) 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., Richardson, E.A. and Ye, Z.H. (2007) The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in *Arabidopsis*. *Plant Cell* 19: 2776-2792.

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 and Cell Physiology* 53: 368-380.

Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L. and Gruissem, W. (2004) GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiology* 136: 2621-2632.

CHAPTER 3

CONCLUDING REMARKS

Hardwood crops such as *Eucalyptus* are an important renewable resource

Eucalyptus tree species are an important source of wood fibres and biopolymers that are used in many different areas such as the paper, textile, food and pharmaceutical industries. Xylem fibres from *Eucalyptus* are particularly valuable because of their high cellulose content and eucalypt plantations are consequently a prime source of chemical cellulose world-wide. This is especially significant considering the potential of cellulosic biomass as a possible source of fermentable sugar for biofuel production. An additional point that makes *Eucalyptus* an attractive species for commercial use is its innate hardiness, which allows it to be grown in many different locations throughout the world including marginal lands not competing with food crops. However, survivability and viability of a particular *Eucalyptus* species or hybrid may come at the cost of commercially desirable traits such as wood quality and chemical composition. Genetic modification and directed breeding programmes may be able to solve this problem by targeting and modifying specific traits in the tree.

The transcriptional network of secondary cell wall biosynthesis has not been fully characterized

Much work has been done on the transcriptional network of secondary cell wall biosynthesis in the model plant *Arabidopsis thaliana*. Even though the network is thought to be conserved in woody and herbaceous plants (Zhong et al. 2010), it is reasonable to expect that large woody perennials such as *Eucalyptus* may differ in some aspects of transcriptional regulation. The focus of studies so far has been mainly on discovery and identification of the major players in the network. A number of studies have identified TFs important for secondary cell wall deposition (Zhong et al. 2008; Ko et al. 2009). This is a good starting point; however, except for a few notable exceptions, such as AtSND1, AtVND6, AtVND7, AtMYB46 and AtMYB83, not much characterisation of these TFs has been performed. There

are a number of TFs in the network (e.g. AtMYB54 and AtSND3) which apart from the occasional expression study (Zhong et al. 2008; Nakano et al. 2010), have not been the focus of functional research. A transcriptional network is dynamic and requires much fine-tuning. Therefore, while it is important to know about the core proteins involved in this network and how they interact to regulate expression, if we are to properly utilise and manipulate this network, we will need to know more about the uncharacterised TFs which may be responsible for fine control of regulation. Also, TFs may differ slightly in their expression, localisation and interactions between xylem cell types, so a technique such as single cell transcriptomics may help to resolve the overlaps in TF expression and function between regulatory networks and determine cell-specific regulatory programmes.

Another aspect of this network which has been neglected is the various layers of regulation that are occurring. We tend to take a simplistic view of these networks, thinking of them in terms of a collection of single interactions between proteins and promoters that operate in a cascade-like fashion to regulate secondary cell wall biosynthesis. In truth, this process is much more complex. Firstly, the interaction between protein and promoter is more complicated than a single TF binding a single motif. A single protein may recognize more than one motif, or may need to bind in a complex with other factors to perform its function (Lelli et al. 2012). The specificity and affinity of proteins may also change depending on their binding partners. Secondly, there has been very little research performed on which proteins interact with each other in this network. While we do know that MYB and NAC proteins can bind as homo- or heterodimers (Olsen et al. 2005; Dubos et al. 2010), we do not know the extent to which this occurs in the network, and we have no information on how many of these may form part of larger protein complexes, or what their role would be therein. Thirdly, we are not certain how strongly epigenetic factors influence expression in this network. Phenomena such as DNA methylation, histone modification and chromosome position in the nucleus may affect chromatin accessibility and activity (Lelli et al. 2012). So far, great strides have been made in this field with the ENCODE project (Dunham et al. 2012), greatly

increasing our understanding of transcriptional regulation in humans, but this has yet to be applied extensively in plants. Fourthly, non-coding RNAs (ncRNAs) need to be considered. These molecules can perform many regulatory functions, from silencing genes performed by microRNAs (miRNAs; Zhang et al. 2006) to directing methylation in gene sequences by long non-coding RNAs (lncRNAs; Wang and Chang 2011). As such ncRNAs are an important piece of the network. Lastly, we need to consider the question of whether transcript abundance is an accurate representation of protein abundance. For the purpose of most studies, this is assumed to be the case. However, it is more than likely that some transcripts that are produced are degraded or transported elsewhere, or some transcripts may be translated multiple times, meaning that that actual protein abundance would differ from that of the transcript. This is an issue for all proteins and not just TFs, but it may have a large effect on TF-related studies. Therefore there are many great opportunities for research to be performed that addresses these issues.

In this study, protein interactions with the promoter of a putative ortholog of *AtSND2* in *E. grandis* (*EgrSND2*) were identified. The findings of this study are the first reports of direct interactions between the TFs tested and the *EgrSND2* promoter. This work has provided insight into the functions and interactions of *EgrSND2*. This not only provides some of the groundwork for characterising the SCW regulatory network in *E. grandis*, but also serves as evidence for possible interactions in the *Arabidopsis* SCW regulatory network.

Conservation of the transcriptional regulatory network of secondary cell wall biosynthesis across plant species

As mentioned previously, the transcriptional regulatory network of secondary cell wall biosynthesis has been extensively studied in *Arabidopsis thaliana*. However, if we are to apply this research in a way that is meaningful to industry, these findings will have to be applied to commercial hardwood crops. This should be possible, as studies have been performed which show that this network is conserved across plant species (Zhong et al.

2010). Of particular note is EgMYB2, a TF from *E. gunnii* (Goicoechea et al. 2005), which is able to activate SCW biosynthesis in tobacco xylem cells when overexpressed (De Micco et al. 2012), and has been shown to be an ortholog of AtMYB46 (Zhong et al. 2010), providing some evidence of conservation of the regulatory network in *Eucalyptus* species. While there is conservation, we are not sure of the extent to which the network is conserved between plant species. Tree species produce much larger quantities of xylem than *Arabidopsis,* so we would expect differences in both the scale and programming of xylogenesis in these plants, which should be reflected in the regulatory machinery underlying these processes.

Genome-wide duplications have occurred in the genome of *Eucalyptus grandis* (Myburg et al. 2011). This may be reflected in the gene families encoding transcription factors (e.g. NACs, MYBs and zinc finger proteins) that are involved in secondary cell wall biosynthesis. This is most likely to result in the generation of paralogs of the genes involved, which may allow for subfunctionalization of retained duplicates. If even a small number of these changes have occurred, the manipulation of the network in trees based on what is known in *Arabidopsis* may become very difficult. Therefore more studies focussing on the conservation of biological function of the network should be performed. This will be valuable as once we know the extent of conservation, we will have an indication of how far the research performed in *Arabidopsis* can be applied to other species.

If we were able to study the SCW regulatory network in a tree species such as *Eucalyptus*, it could be a valuable model. There are difficulties in doing this. First, it is not feasible to use conventional *Arabidopsis* methods of transformation, such as floral dipping, on tree species. Plant callus transformation is possible, but is slow and labour intensive. Secondly, tree species such as *Eucalyptus* take much longer to grow and require more space than *Arabidopsis,* meaning that any transgenic tree line would take years and much greenhouse space to analyse. There are ways to work around this problem. For example, one may express a tree protein in *Arabidopsis.* If the network is highly conserved, any effects we see should translate to tree species. We may also use specialised techniques such as Induced

Somatic Sector Analysis (ISSA; Creux et al. 2013a), which allows us to directly transform the vascular cambium of trees, though this is limited to expression analysis in vascular tissues, as mainly the vascular cambium of mature trees are transformed in this analysis. Currently there is no feasible way to quickly analyse and characterise TFs from the transcriptional regulatory network of SCW biosynthesis in trees so our best course of action may be to determine the extent of functional conservation of the network, so that studies of tree proteins performed in *Arabidopsis* may be adequate to infer functions in tree species such as *Eucalyptus.*

New TF-promoter interactions in tree species

Determining the expression pattern of the *E. grandis SND2* ortholog promoter fragment is difficult. Apart from techniques such as ISSA (Van Beveren et al. 2006; Creux et al. 2011), there is no viable way to analyse expression in a native genetic background. As of yet, there is no standard *Eucalyptus* transformation protocol and even if routine transformation were possible, it would take many years for the trees to reach maturity. Transformation in *Populus* is possible, but apart from the same problem of growing time and despite *Populus* being a woody species, it is still relatively distantly related to *Eucalyptus* (Wang et al. 2009). RNA-seq analyses are ideal for determining expression patterns of TFs in trees. However, they can be very expensive, analysis can be difficult if a genome sequence is not available and results would need to be confirmed with RT-qPCR. A viable alternative is to test the promoter fragment in a model species, such as *Arabidopsis,* which is what was done in this study. However, the results of this study were different to those obtained previously (Zhong et al. 2008), showing expression in many different tissues, rather than a fibre specific expression pattern. This may be due to the use of the promoter fragment in a heterologous species, though it is equally likely that the difference in constructs used between the different studies caused the difference in expression patterns.

Generally, plant promoter studies are performed by selecting a 500 bp-2kb fragment upstream of the gene (Assunção et al. 2010; Ohashi-Ito et al. 2010; Creux et al. 2013b; Downs et al. 2013). In the case of this study, we used a 1.5 kb fragment upstream of the start codon. In the study performed by Zhong et al. (2008), 3 kb upstream of the start codon, the entire CDS of the protein and 2 kb downstream of the termination codon were used, with the reporter gene inserted in-frame just before the termination codon. These differences may well account for the different expression patterns seen in this study, due to the presence or absence of regulatory elements in the sequence. In order to eliminate problems such as these, studies should be performed in *Eucalyptus*, similar to those done in *Arabidopsis* (Molina and Grotewold 2005) and the ENCODE project (Dunham et al. 2012) in which promoter regions are characterised on a genome-wide scale, using a range of sequence and functional data. This would be relatively easy to achieve, since many of the resources required for the endeavour are readily available (e.g. the *Eucalyptus grandis* genome sequence) or could be easily obtained (e.g. ChIP-seq and DNase I hypersensitivity data). This may allow us to better estimate the size of the promoter for the class of *Eucalyptus* gene we are working on, and could serve as a guide for selecting appropriate promoter fragments for functional studies.

The value of the research performed in the study is twofold. Not only is it the first to identify some of the direct regulators of SND2, but it is the first to do it with newly isolated *Eucalyptus* transcription factors. This means that it not only elucidates new connections in the transcriptional network, but is also important groundwork for elucidating the network in *Eucalyptus* species. It is important to note that these interactions are shown *in vivo,* using Y1-H analyses. To expand upon this work, the first step would be to confirm the protein-DNA interactions seen in this study using Electrophoretic Mobility Shift Assays (EMSAs). If this study were to be repeated, it may be worthwhile to use a different analysis, such as a transactivation analysis using the bait promoter sequence and a GUS or GFP reporter gene in protoplast cells derived from a non-woody tissue such as mature *E. grandis* leaves. It

would be beneficial to use a non-woody tissue as the secondary cell wall biosynthetic program should not be as active in these cells, meaning there is less chance that it will interfere with the analysis. However, even tissues such as leaves would still have some small amount of secondary cell wall biosynthesis occurring, so it would still be wise to confirm the findings using an *in vitro* technique such as EMSA.

In this study, EgrMYB46, EgrMYB83, EgrVND6, EgrSND2, EgrSND3 and EgrZF1 were able to interact with the *EgrSND2* 1.5 kb promoter fragment. Interaction of both EgrMYB46 and EgrMYB83, the two intermediate master regulators (McCarthy et al. 2009) with the *EgrSND2* promoter region indicates that EgrSND2 may be involved in SCW biosynthesis in vessel cells as well as fibre cells. The ortholog of the potential mid-level master regulator AtC3H14 (Ko et al. 2009), designated EgrZF1, also showed interaction with the promoter fragment, as did EgrVND6, the ortholog of the master switch of SCW biosynthesis in metaxylem vessels (Yamaguchi et al. 2010), further supporting this hypothesis. The interaction of EgrSND3, the ortholog of AtSND3 which was associated with cellulose biosynthesis in *Arabidopsis* (Zhong et al. 2008), with the promoter fragment suggests a role of EgrSND2 in cellulose biosynthesis. Seeing as in *Arabidopsis* AtSND3 is a direct target of AtSND1, but AtSND2 is not (Zhong et al. 2008), EgrSND3 may provide a path through which EgrSND2 is regulated by the master regulator EgrSND1 in xylem fibres, though EgrMYB46, EgrMYB83 and EgrZF1 are possible alternative routes. Lastly, EgrSND2 shows interaction with its own promoter, indicating possible autoregulation through a feedback loop. This may affect the temporal expression of EgrSND2, as autoregulation may increase or decrease response time to a stimulus (Alon 2007). The findings of this study are an important step to understanding SCW biosynthesis in commercially important tree species such as *E. grandis*, as well as other plants like *Arabidopsis*. This may eventually allow for the breeding and modification of trees for more desirable wood properties for industry and will increase understanding of SCW biosynthesis as a whole.

3.5 References

Alon, U. (2007) Network motifs: theory and experimental approaches. *Nature Reviews Genetics* 8: 450-461.

Assunção, A.G., Herrero, E., Lin, Y.-F., Huettel, B., Talukdar, S., Smaczniak, C., et al. (2010) *Arabidopsis thaliana* transcription factors bZIP19 and bZIP23 regulate the adaptation to zinc deficiency. *Proceedings of the National Academy of Sciences* 107: 10296-10301.

Creux, N., Myburg, A., Bossinger, G. and Spokevicius, A. (2011) Analysis of *cellulose synthase* (*CesA*) promoter function in trees using Induced Somatic Sector Analysis (ISSA). *BMC Proceedings* p. O41. BioMed Central Ltd.

Creux, N.M., Bossinger, G., Myburg, A.A. and Spokevicius, A.V. (2013a) Induced somatic sector analysis of *cellulose synthase* (*CesA*) promoter regions in woody stem tissues. *Planta* 3: 799-812.

Creux, N.M., De Castro, M.H., Ranik, M., Maleka, M.F. and Myburg, A.A. (2013b) Diversity and *cis*-element architecture of the promoter regions of *cellulose synthase* genes in *Eucalyptus*. *Tree Genetics & Genomes* 9: 989-1004.

De Micco, V., Ruel, K., Joseleau, J.-P., Grima-Pettenati, J. and Aronne, G. (2012) Xylem anatomy and cell wall ultrastructure of *Nicotiana Tabacum* after lignin genetic modification through transcriptional activator EGMYB2. *IAWA Journal-International Association of Wood Anatomists* 33: 269-286.

Downs, G.S., Bi, Y.-M., Colasanti, J., Wu, W., Chen, X., Zhu, T., et al. (2013) A developmental transcriptional network for *Zea mays* defines coexpression modules. *Plant Physiology*.

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.

Dunham, I., Birney, E., Lajoie, B.R., Sanyal, A., Dong, X., Greven, M., et al. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57-74.

Goicoechea, M., Lacombe, E., Legay, S., Mihaljevic, S., Rech, P., Jauneau, A., et al. (2005) EgMYB2, a new transcriptional activator from *Eucalyptus* xylem, regulates secondary cell wall formation and lignin biosynthesis. *The Plant Journal* 43: 553-567.

Ko, J.H., Kim, W.C. and Han, K.H. (2009) Ectopic expression of MYB46 identifies transcriptional regulatory genes involved in secondary wall biosynthesis in *Arabidopsis*. The *Plant Journal* 60: 649-665.

Lelli, K.M., Slattery, M. and Mann, R.S. (2012) Disentangling the many layers of eukaryotic transcriptional regulation. *Annual Review of Genetics* 46: 43-68.

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 and cell physiology* 50: 1950-1964.

Molina, C. and Grotewold, E. (2005) Genome wide analysis of *Arabidopsis* core promoters. *BMC Genomics* 6: 25.

Myburg, A., Grattapaglia, D., Tuskan, G., Jenkins, J., Schmutz, J., Mizrachi, E., et al. (2011) The *Eucalyptus grandis* Genome Project: Genome and transcriptome resources for comparative analysis of woody plant biology. *BMC Proceedings* 5: I20.

Nakano, Y., Nishikubo, N., Goué, N., Ohtani, M., Yamaguchi, M., Katayama, Y., et al. (2010) MYB transcription factors orchestrating the developmental program of xylem vessels in *Arabidopsis* roots. *Plant Biotechnology* 27: 267-272.

Ohashi-Ito, K., Oda, Y. and Fukuda, H. (2010) Arabidopsis VASCULAR-RELATED NAC-DOMAIN6 directly regulates the genes that govern programmed cell death and secondary wall formation during xylem differentiation. *Plant Cell* 22: 3461-3473.

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.

Van Beveren, K.S., Spokevicius, A.V., Tibbits, J., Wang, Q. and Bossinger, G. (2006) Transformation of cambial tissue *in vivo* provides an efficient means for induced somatic sector analysis and gene testing in stems of woody plant species. *Functional plant biology* 33: 629-638.

Wang, H., Moore, M.J., Soltis, P.S., Bell, C.D., Brockington, S.F., Alexandre, R., et al. (2009) Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proceedings of the National Academy of Sciences* 106: 3853-3858.

Wang, Kevin C. and Chang, Howard Y. (2011) Molecular Mechanisms of Long Noncoding RNAs. *Molecular Cell* 43: 904-914.

Yamaguchi, M., Goue, N., Igarashi, H., Ohtani, M., Nakano, Y., Mortimer, J.C., et al. (2010) VASCULAR-RELATED NAC-DOMAIN6 and VASCULAR-RELATED NAC-DOMAIN7 effectively induce transdifferentiation into xylem vessel elements under control of an induction system. *Plant Physiology* 153: 906-914.

Zhang, B., Pan, X., Cobb, G.P. and Anderson, T.A. (2006) Plant microRNA: a small regulatory molecule with big impact. *Developmental biology* 289: 3-16.

Zhong, R., Lee, C. and Ye, Z.H. (2010) Evolutionary conservation of the transcriptional network regulating secondary cell wall biosynthesis. *Trends in Plant Science* 15: 625-632.

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.

SUMMARY

Summary: Qualitative GUS and targeted yeast one-hybrid analyses of the Eucalyptus grandis SND2 promoter region

Jonathan Botha

Supervised by ***Prof A.A. Myburg, Dr C. Maritz-Olivier*** *and* ***Mr E. Mizrachi***

Submitted in partial fulfilment of the requirements for the degree Magister Scientiae

Department of Genetics

University of Pretoria

Woody tissues of tree species are known to have thick SCWs that are rich in three main biopolymers: namely lignin, hemicellulose and cellulose. Cellulose is of particular interest to industry, as it is a source of chemical cellulose and may be a viable alternative to fossil fuels. SCW biosynthesis is a complex process regulated by a transcriptional cascade. SND1 is believed to be the master regulator for SCW biosynthesis in xylem fibres and cells which have thickened SCWs for structural support as overexpression of SND1 results in the upregulation of a number of proteins important for SCW biosynthesis. Three of these, namely SND2, SND3 and MYB103 were able to activate expression of the *AtCesA8* promoter, indicating a role in cellulose biosynthesis. Of these three proteins, only SND2 was not a direct target of SND1. The aims of this study were firstly to determine in which tissues of *Arabidopsis* the 1.5 kb *Eucalyptus grandis SND2* promoter fragment was active using qualitative *GUS* analysis and, secondly, to determine which *E. grandis* orthologs of a subset of proteins important for SCW biosynthesis were able to bind *in vivo* directly to the 1.5 kb promoter fragment and a 500 bp promoter truncation using a yeast one-hybrid (Y1-H) analysis.

Wild-type plants were transfected with *Agrobacterium* containing the β-Glucuronidase reporter gene under the control of the 1.5 kb *EgrSND2* promoter fragment. The resulting transgenic plant lines were submitted to qualitative GUS analysis at one, three and six

weeks. An identical analysis was performed concurrently on the *EgrSND2* 1.5 kb promoter fragment for comparison. It was found that at no stage of development was activity restricted to the vasculature, though there was strong expression in the vasculature. Expression appeared to be localised mainly in young, fast growing tissues. This is not in line with previous observations, but may be due to inherent differences between the construct used in this study and those used previously. Also the 1.5 kb *EgrSND2* promoter fragment showed an expression pattern similar to that of the 1.5 kb *EgrSND3* promoter fragment.

The EgrSND2 1.5 kb promoter fragment and a 500 bp truncation of the promoter fragment were subjected to Y1-H screening against the *E. grandis* orthologs of a number of transcription factors (TFs) thought to be important for SCW biosynthesis. A number of orthologs such as EgrMYB46, EgrMYB83, EgrSND2, EgrSND3, EgrVND6, EgrZF1 and EgrZF2 were able to bind to the 1.5 kb and 500 bp promoter fragments. Results also confirm that *EgrSND2* plays a role in SCW biosynthesis. Orthologs of the master regulators for xylem vessel development were also able to bind, indicating that *EgrSND2* may be involved in SCW biosynthesis in many cell types. Also, *EgrSND2* was able to bind to the *EgrCesA8* promoter, implicating a possible role in cellulose biosynthesis. These analyses provide the first evidence of direct interaction of EgrMYB46, EgrMYB83, EgrSND2, EgrSND3, EgrVND6, EgrZF1 and EgrZF2 with egrSND2 *in vivo*, and, as such, are a valuable framework for further study of the transcriptional network of SCW biosynthesis in both *E. grandis* and *Arabidopsis.*

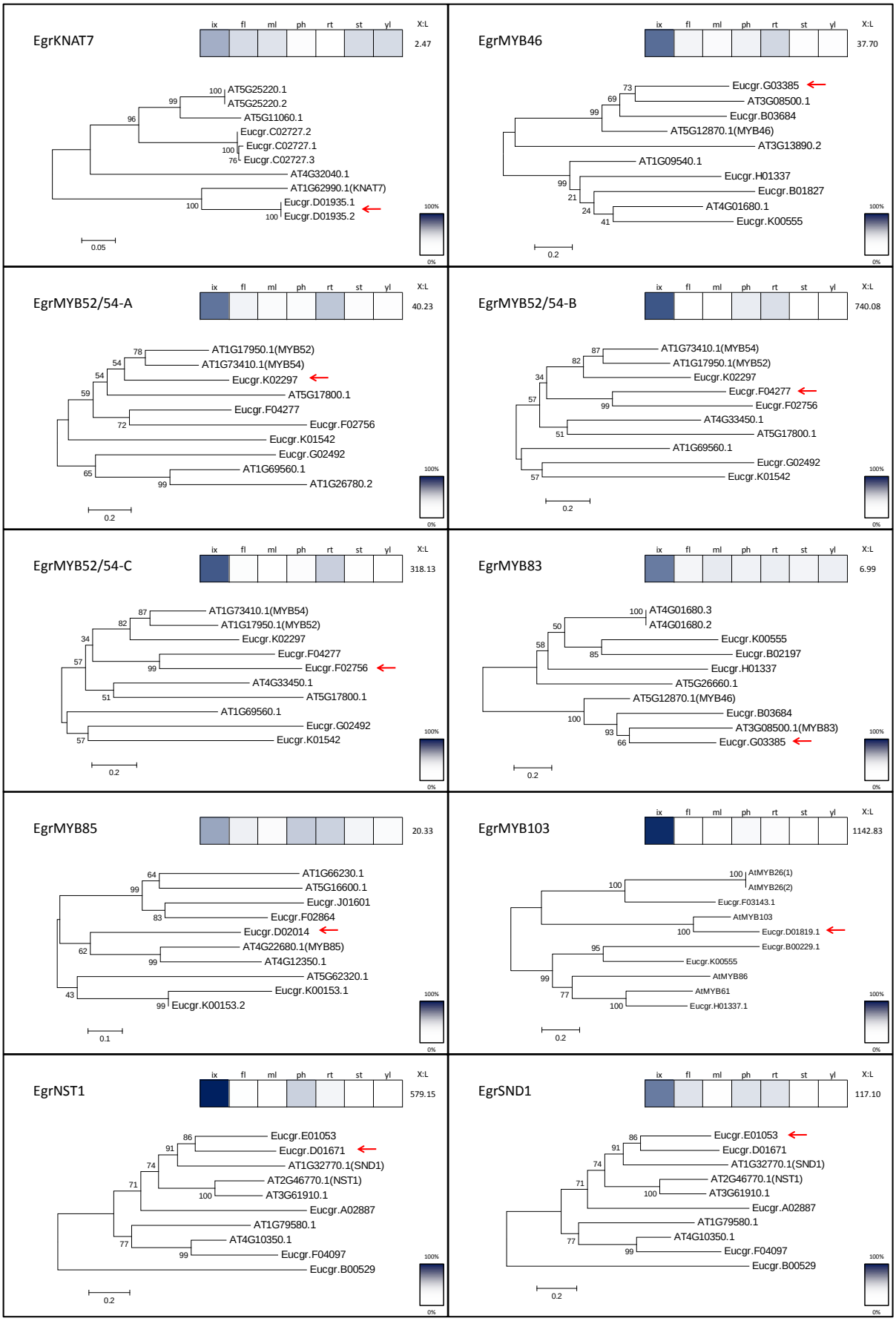
APPENDIX 1

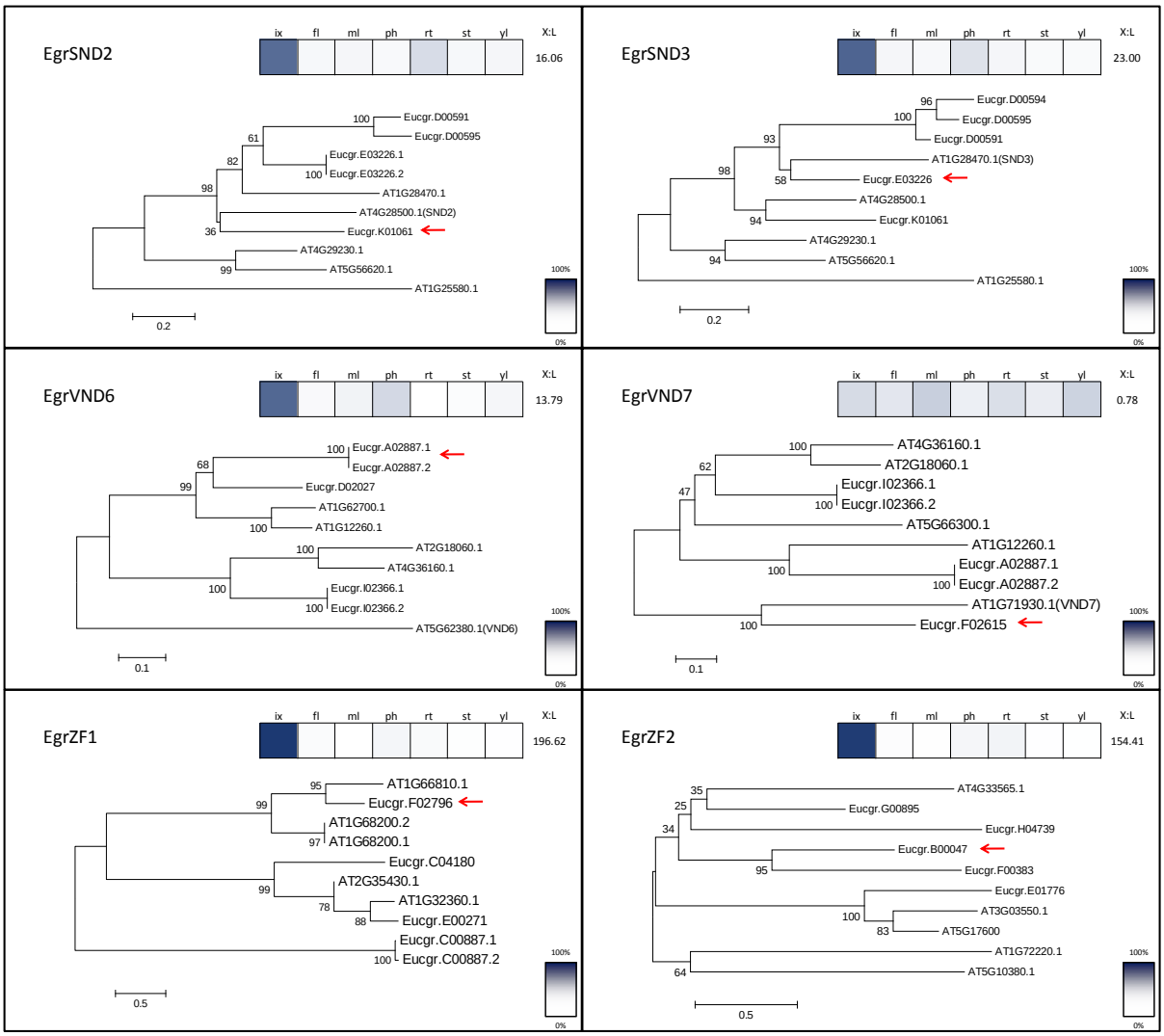
Supplementary Tables and Figures

Supplementary Table 1: **Names and sequences of oligonucleotide primers used for the promoter-GUS PCR and the colony PCR orientation screens**.

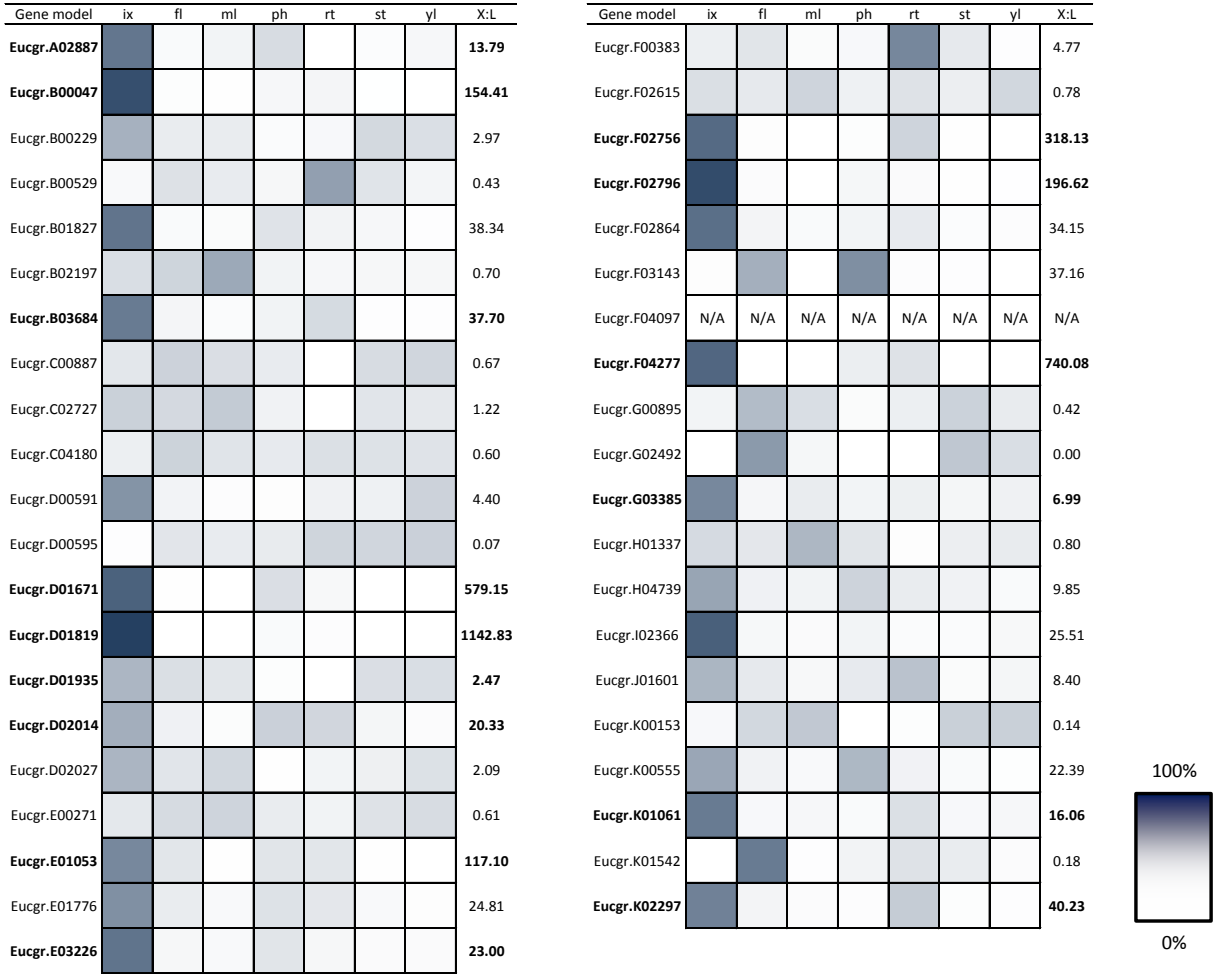
Primer name[a]	Primer sequence (5'-3')[b]	Tm (°C)[c]	Applications[d]
EgSND2prom_F_MluI	GAACGCGTTTATAGGGCACGGGCGAACC	58	Amplification, orientation screening and sequencing of the 500 bp and 1.5 kb *EgrSND2* promoter fragment
EgSND2prom500_F_MluI	ACGCGTGCGGTTCTGTCTCTGCAGATCATA	58	
EgSND2prom_R_SpeI	TCACTAGTGATTTCCTGCTCTTCGCTTTTT	58	
EgSND3prom_F_MluI	GAACGCGTAGGATGCCCTCAAACTCAGA	58	Amplification, orientation screening and sequencing of the 1.5 kb *EgrSND3* promoter fragment
EgSND3prom500_F_MluI	ACGCGTGCGGCCTCCTTTCGCAATAAAATCAA	58	
EgSND3prom_R_SpeI	TCACTAGTGATCTCTTTCTCTTCGCTTCGAGTCT	58	
2x35S-F	GGTCAACATGGTGGAGCACGACACA	56	Amplification, orientation screening and sequencing of the *35S CaMV* promoter
2x35S-R	AACTAGTTCTAGAGTCGAGGTCCTC	56	
GUS-SPC(F)	CATGTCGCGCAAGACTGTAA	56	Amplification, orientation screening and sequencing of the *β-Glucuronidase (GUS)* gene.
GUS-SPC(R)	TCCGGTTCGTTGGCAATACT	56	
M13F-pUC	GTTTTCCCAGTCACGAC	58	Standard primers used for amplification, orientation screening and sequencing of Gateway™ vectors
M13R-pUC	CAGGAAACAGCTATGAC	58	
pGAD_F	AGTAGCAACGGTCCGAACCT	60	Amplification, orientation screening and sequencing of prey (pDEST-GADT7) vectors
pGAD_R	GATGGTGCACGATGCACAGT	60	
pHIS_R	GGTGTGATGGTCGTCTATGT	58	Amplification, orientation screening and sequencing of bait (pHIS2.1) vectors

[a]The name of the primers used
[b]The sequence of the primer in a 5' to 3' configuration
[c]The annealing temperature used in PCR amplification
[d]Possible applications in which the primer is used

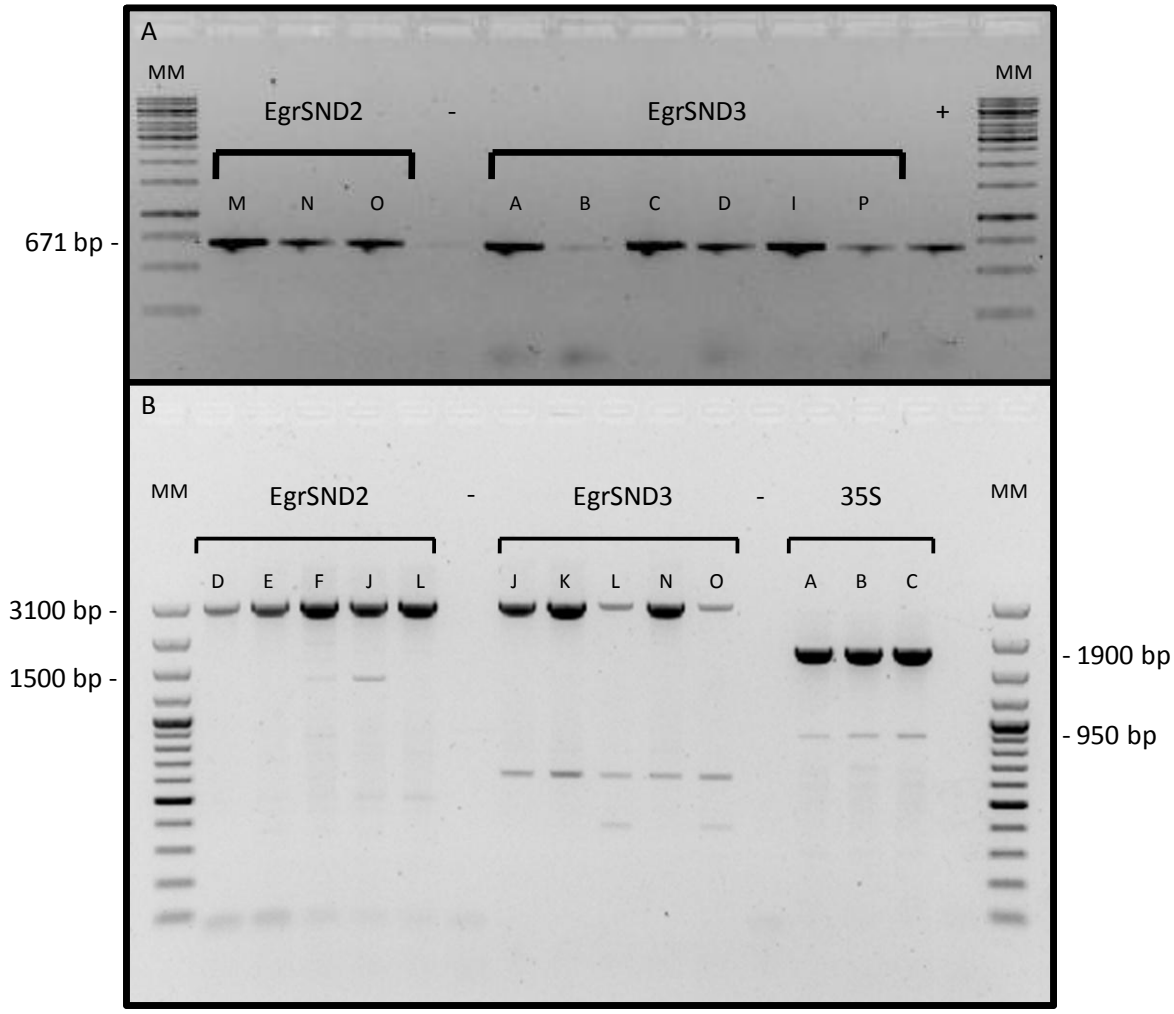




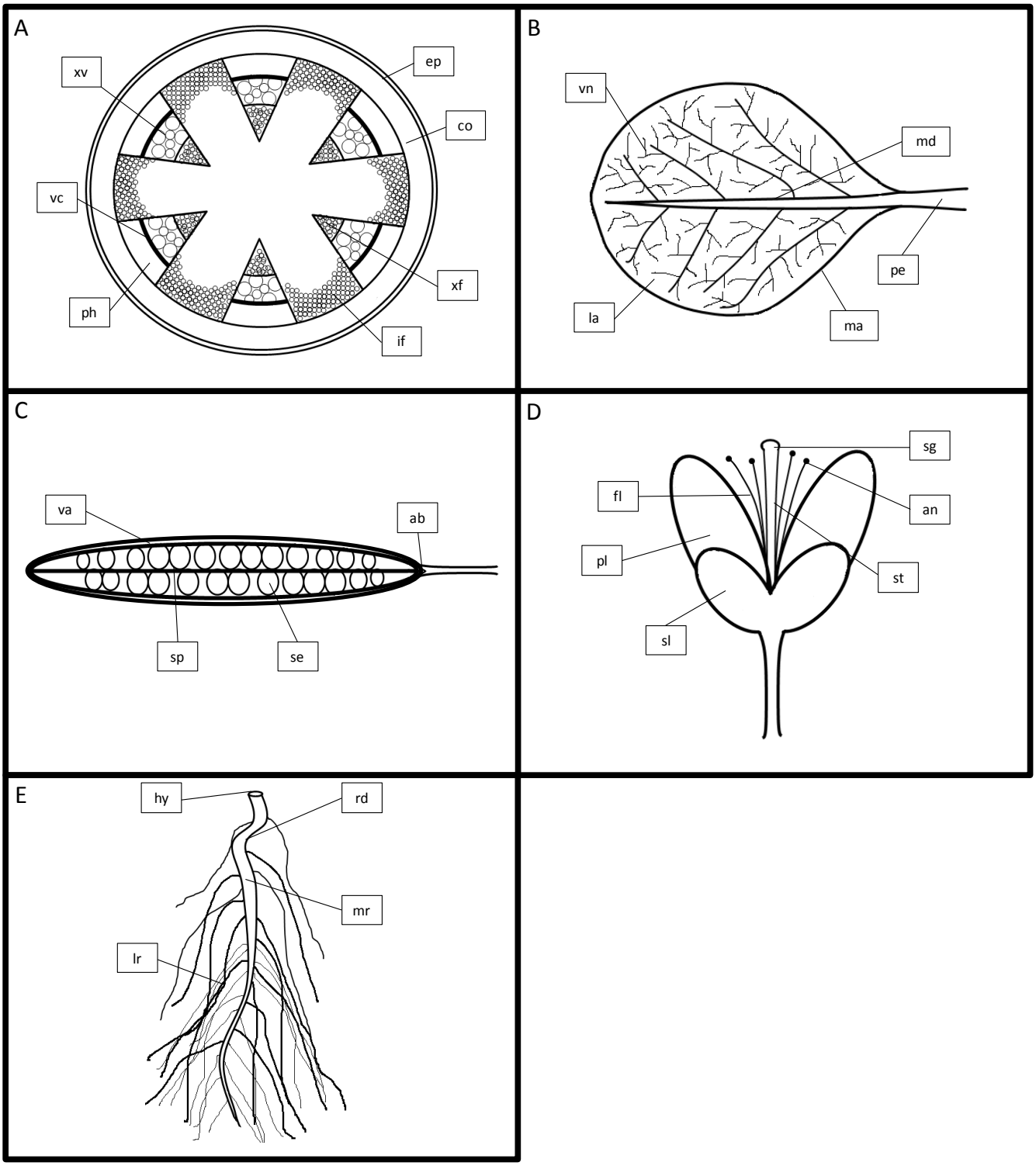
Supplementary Figure 1: Neighbour-joining trees used to identify the putative *Eucalyptus grandis* orthologs of the selected *Arabidopsis* TFs. Branch lengths reflect the number of amino acid substitutions that have taken place. Bootstrap values are based on 1000 bootstrap replicates. Proteins used in the tree were obtained from TAIR 10 and the v1.1 *E. grandis* genome build available on www.phytozome.net. The name of the putative *E. grandis* ortholog (Table 3) is displayed in the top left corner of each block. The red arrow indicates the ortholog selected for the yeast one-hybrid analysis. The relative transcript levels of the selected ortholog in various tissues calculated from the average expression from three *E. grandis* trees (Hefer et al., in preparation) is displayed in the top right. The immature xylem (ix), flowers (fl), mature leaf (mf), phloem (ph), roots (rt), shoot tips (st) and young leaf (yl) are represented. X:L indicates the xylem to leaf ratio of expression, calculated by dividing the percentage of transcripts from ix by the average of the percentage of transcripts from both ml and yl.



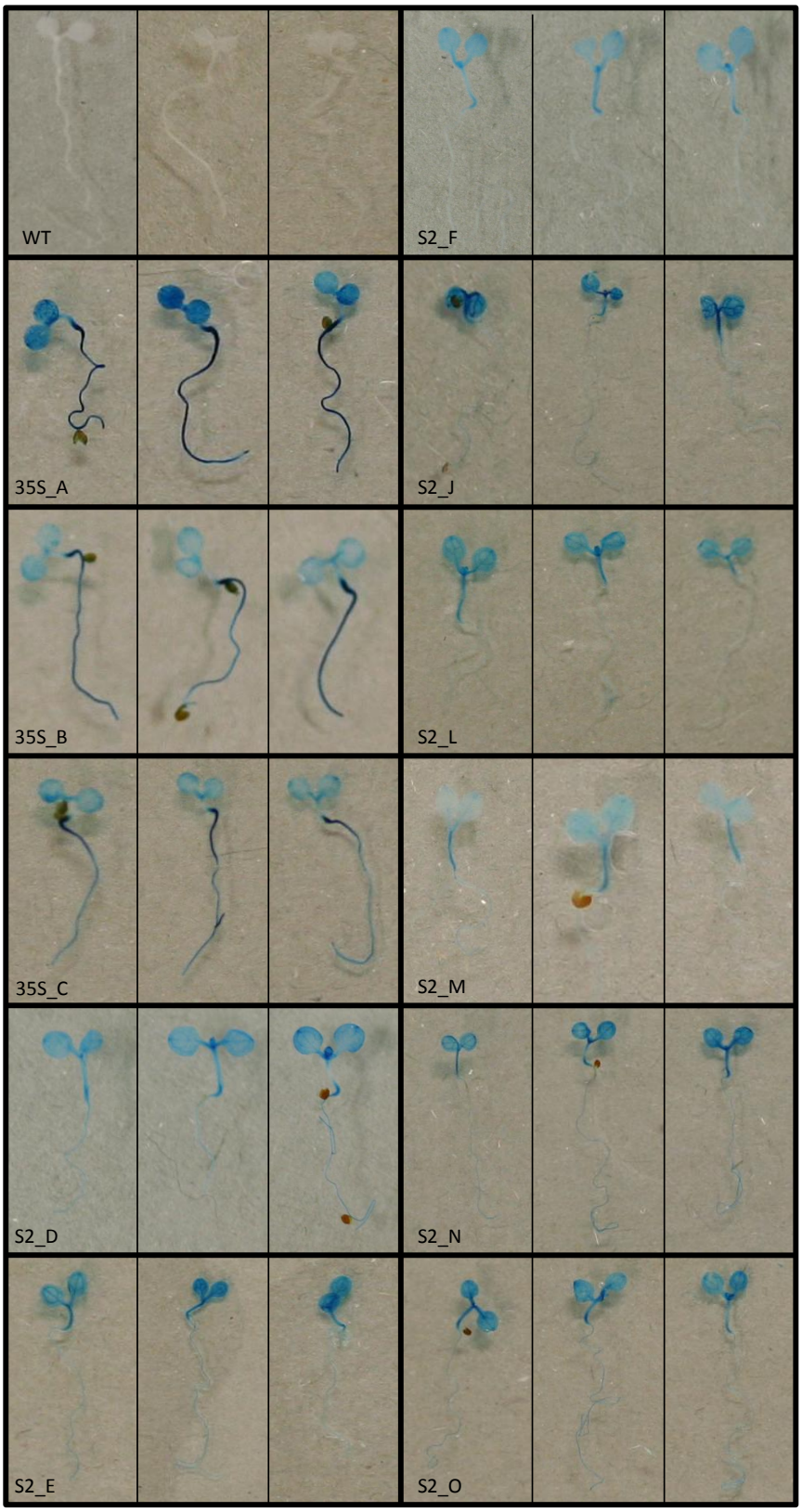
Supplementary Figure 2: The relative transcript levels of the *E. grandis* genemodels used in the phylogenetic analysis in Supplementary Figure 1 in various tissues calculated from the RNA-seq data from three *E. grandis* trees (Hefer et al., in preparation). The immature xylem (ix), flowers (fl), mature leaf (mf), phloem (ph), roots (rt), shoot tips (st) and young leaf (yl) are represented. X:L indicates the xylem to leaf ratio of expression, calculated by dividing the percentage of transcripts from ix by the average of the percentage of transcripts from both ml and yl. The genemodels in bold are those that were selected as putative orthologs and whose expression is shown in the right corner of the phylogenies.



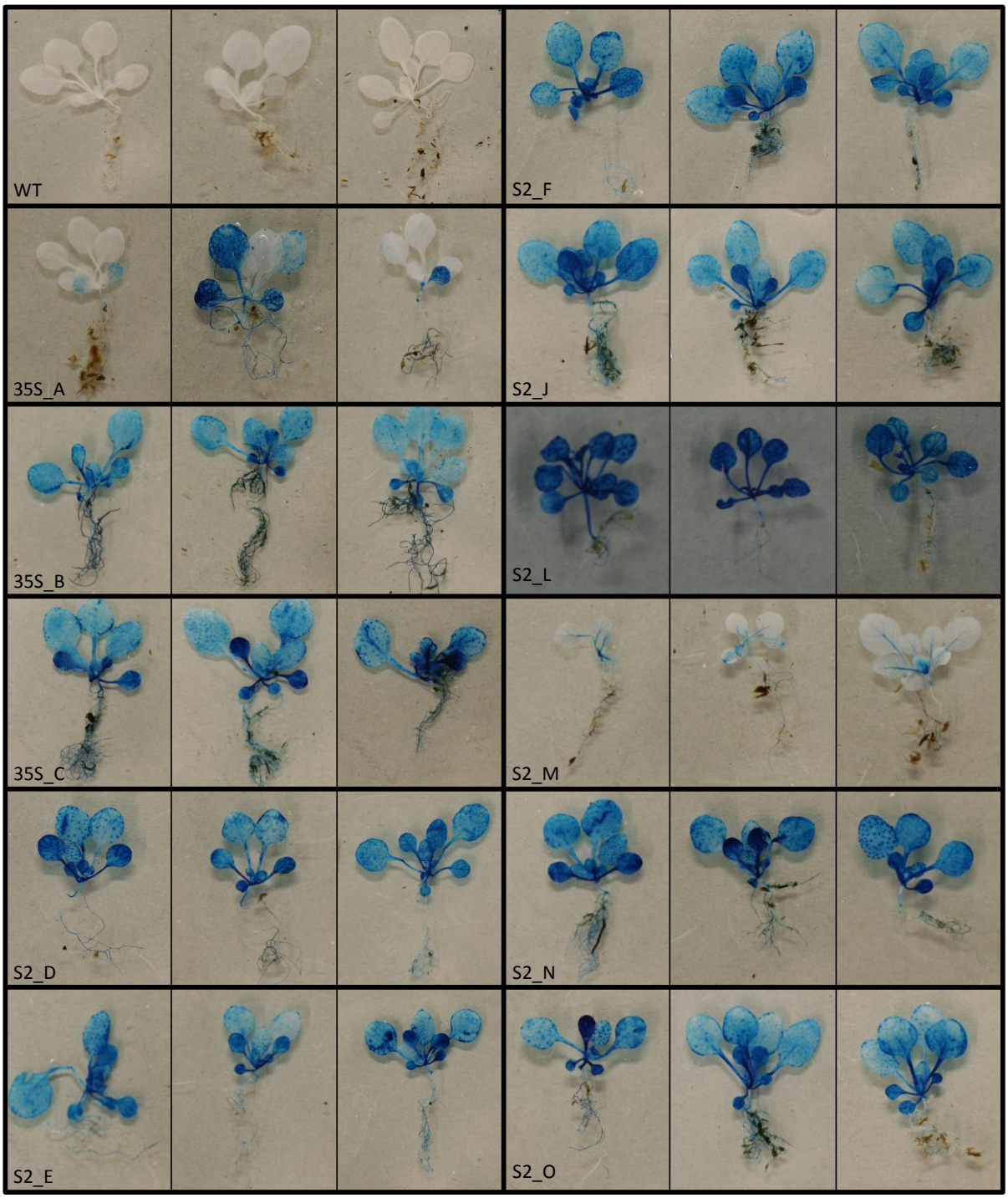
Supplementary Figure 3: Agarose gels showing the presence of the reporter gene in the transformed T2 plant lines. A: *GUS* reporter genes amplified from *Arabidopsis* genomic DNA using *GUS*-specific forward and reverse primers (GUS-SPC(F) and GUS-SPC(R) respectively, Supplementary Table 1). The labels above the line brackets indicate the promoter fragment which is present in the transformed T2 plant lines being tested. The letters indicated the designation of the plant line for each promoter construct. MM indicates the Fermentas 1 kb molecular weight standard. "+" indicates a positive control plant line containing the *35SCaMV* promoter upstream of the *GUS* reporter gene. "-" indicates the wild-type control. B: Promoter and *GUS* reporter genes amplified from *Arabidopsis* genomic DNA using gene-specific forward primers (EgSND2prom_F_MluI, EgSND3prom_F_MluI and 2x35S-F for the *EgrSND2, EgrSND3* and *35SCaMV* promoter constructs respectively) and *GUS* reporter gene specific reverse primers (GUS-SPC(R) Supplementary Table 1).



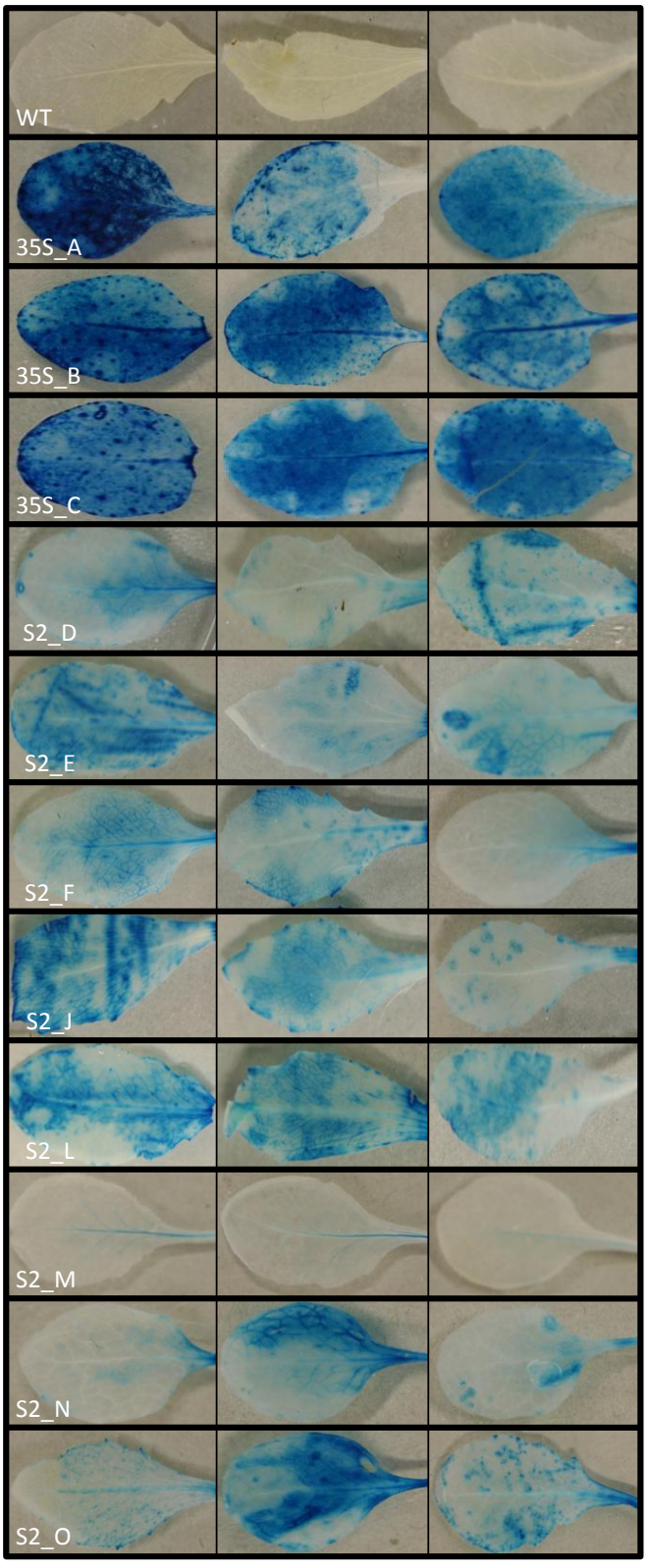
Supplementary Figure 4: Schematic representations of *Arabidopsis thaliana* organs subjected to quantitative GUS analysis. A: Inflorescence stem; co = cortex, ep = epidermis, if = interfascicular fibre, ph = phloem, vc = vascular cambium, xf = xylem fibre, xv = xylem vessel. B: Leaf; la = lamina, ma = margin, md = midrib, pe = petiole, vn = vein. C: Silique; ab = abscisson zone, se = seed, sp = septum, va = valve (carpel). D: Flower; an = anther, fl = filament, pl = petal, sg = stigma, sl = sepal, st = style. E: Root; hy = hypocotyl, lr = lateral root, mr = main root, rd = root dermis.



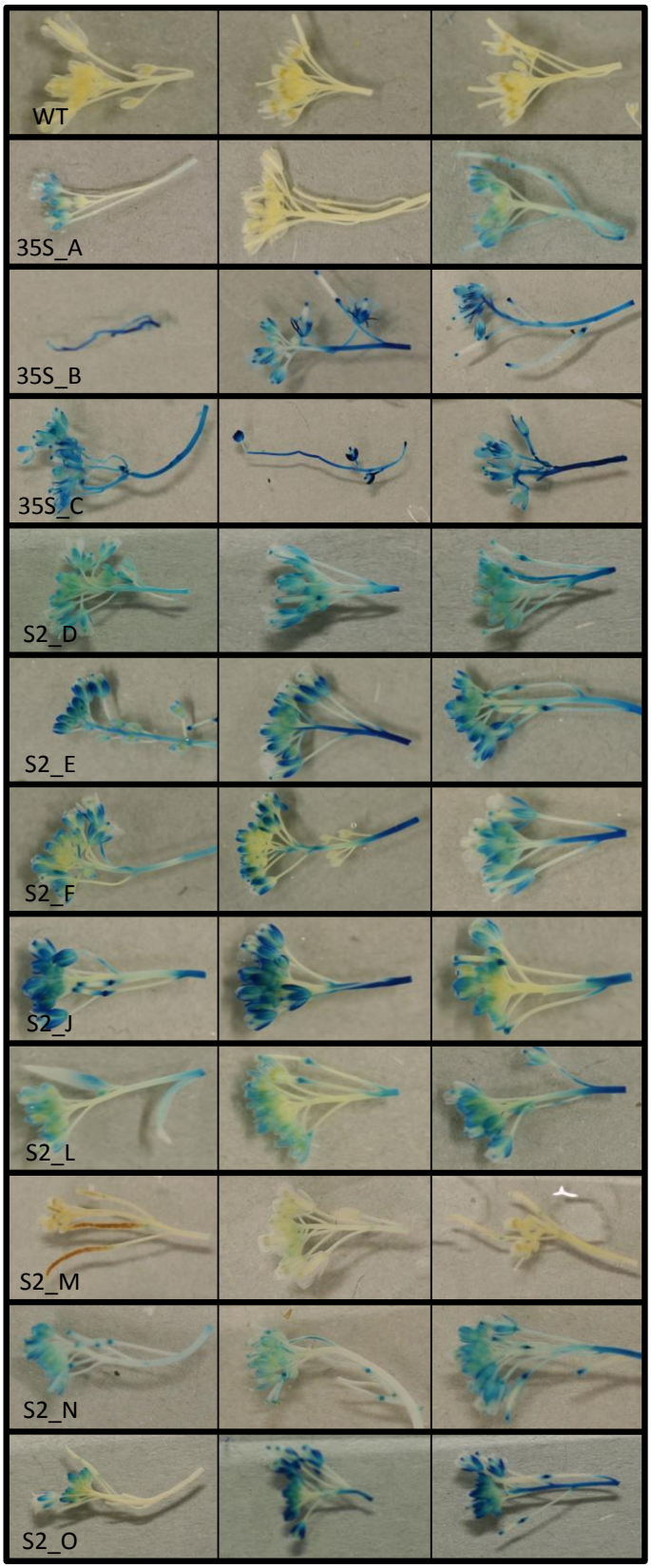
Supplementary Figure 5: One-week-old *EgrSND2 1.5kb promoter::GUS* plant *GUS*-stained whole plants. S2_A, S2_D, S2_E, S2_F, S2_J, S2_L, S2_M, S2_N and S2_O are *EgrSND2 1.5kb promoter::GUS* plant line designations. Three replicates of each are shown. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate *35SCaMV* promoter positive controls.



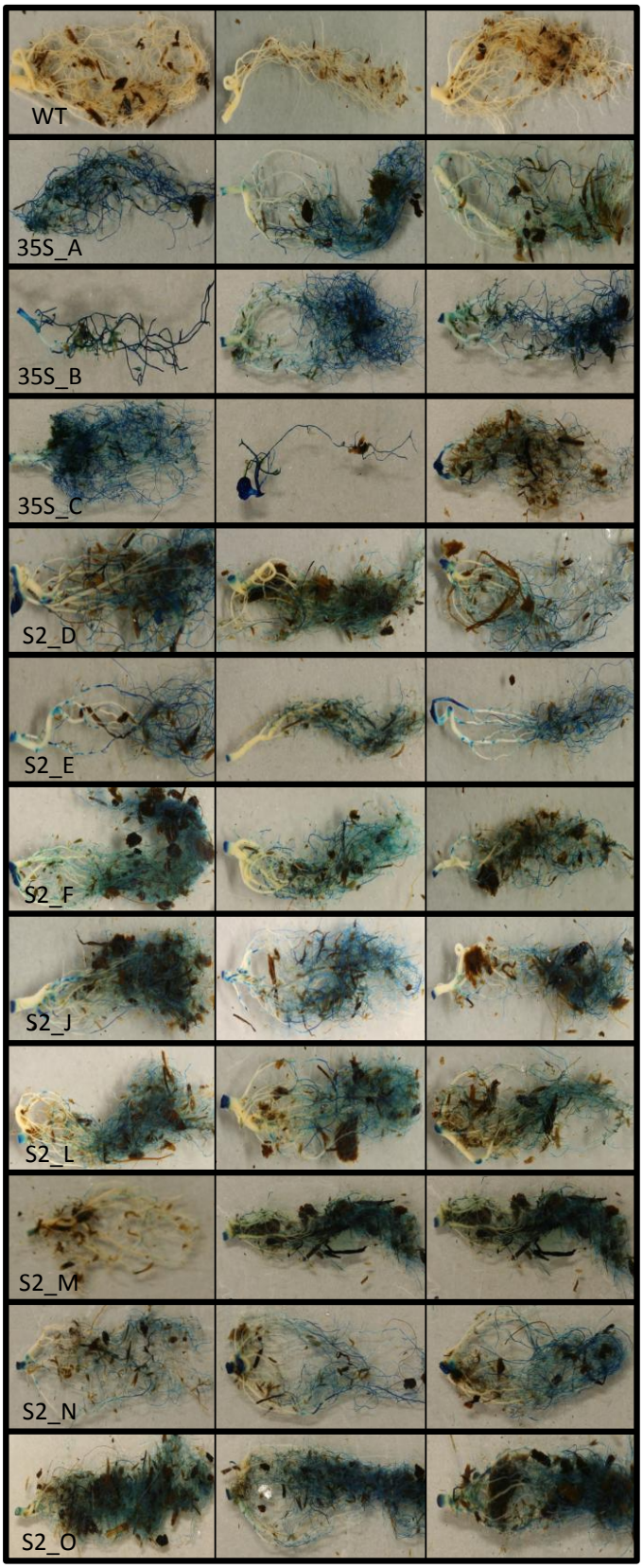
Supplementary Figure 6: Three-week-old *EgrSND2 1.5kb promoter::GUS* plant *GUS*-stained whole plants. S2_A, S2_D, S2_E, S2_F, S2_J, S2_L, S2_M, S2_N and S2_O are *EgrSND2 1.5kb promoter::GUS* plant line designations. Three replicates of each are shown. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate *35SCaMV* promoter positive controls.



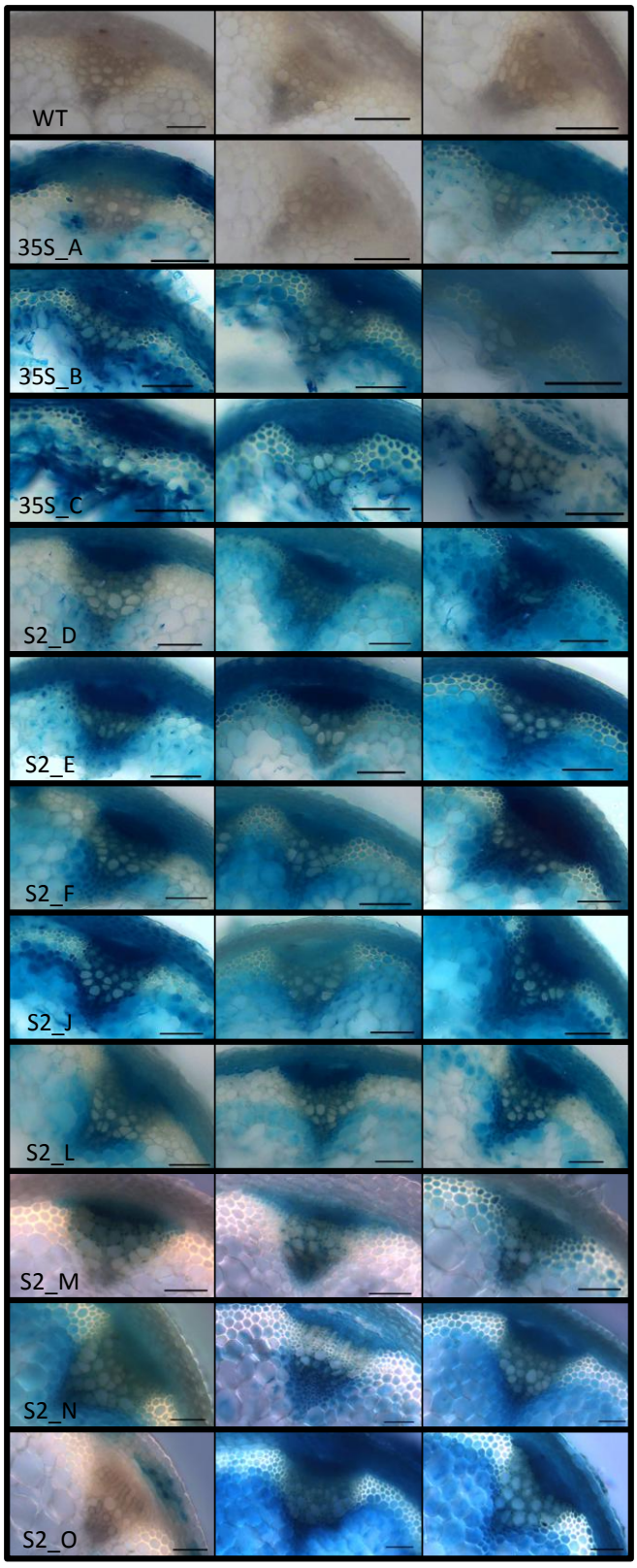
Supplementary Figure 7: Six-week-old *EgrSND2 1.5kb promoter::GUS* plant *GUS*-stained leaves. S2_A, S2_D, S2_E, S2_F, S2_J, S2_L, S2_M, S2_N and S2_O are *EgrSND2 1.5kb promoter::GUS* plant line designations. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate *35SCaMV* promoter positive controls.



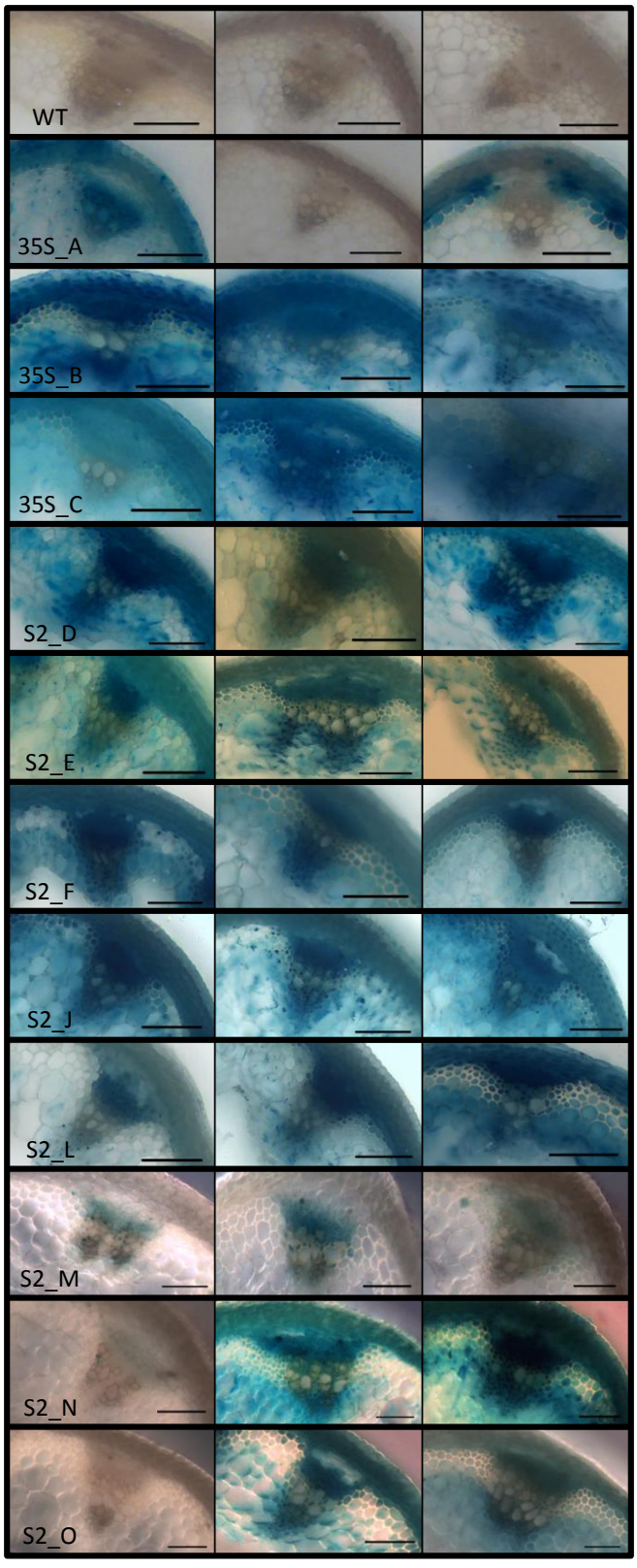
Supplementary Figure 8: Six-week-old *EgrSND2 1.5kb promoter::GUS* plant *GUS*-stained flowers. S2_A, S2_D, S2_E, S2_F, S2_J, S2_L, S2_M, S2_N and S2_O are *EgrSND2 1.5kb promoter::GUS* plant line designations. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate *35SCaMV* promoter positive controls.



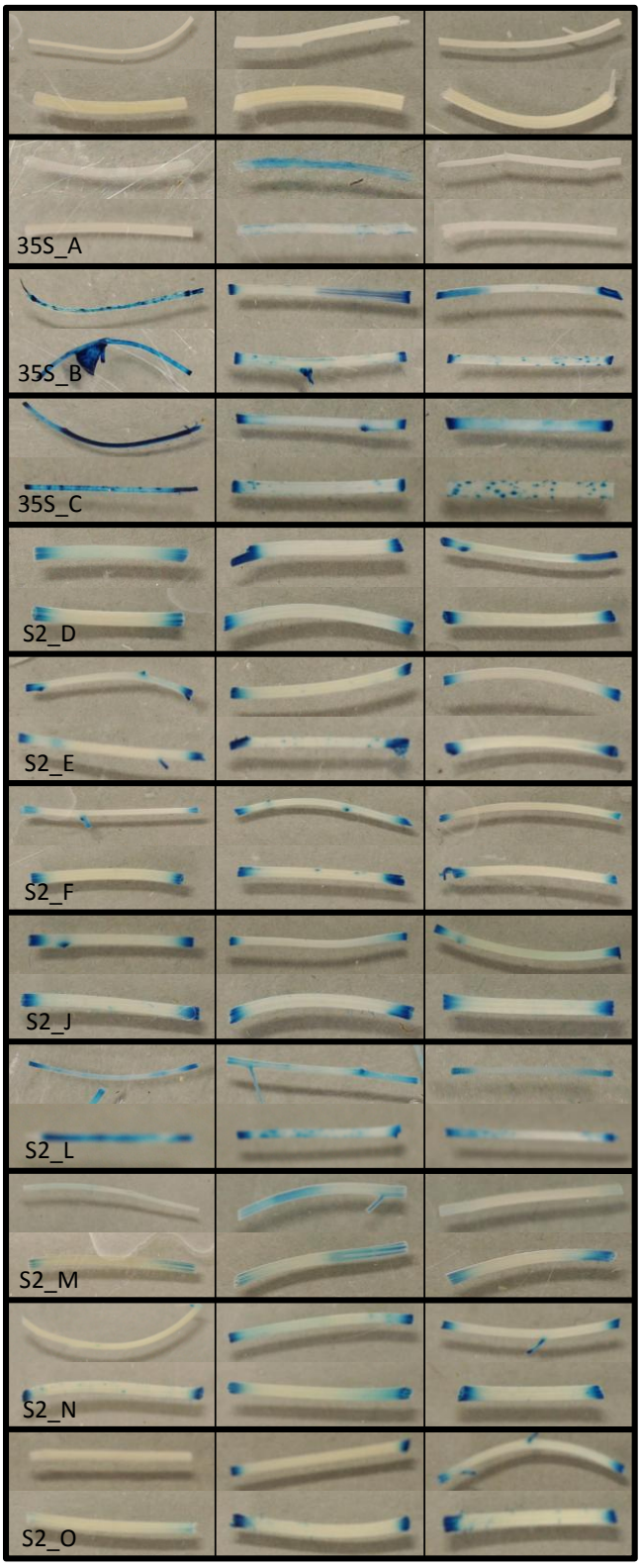
Supplementary Figure 9: Six-week-old *EgrSND2 1.5kb promoter::GUS* plant *GUS*-stained roots and hypocotyls. S2_A, S2_D, S2_E, S2_F, S2_J, S2_L, S2_M, S2_N and S2_O are *EgrSND2 1.5kb promoter::GUS* plant line designations. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate *35SCaMV* promoter positive controls.



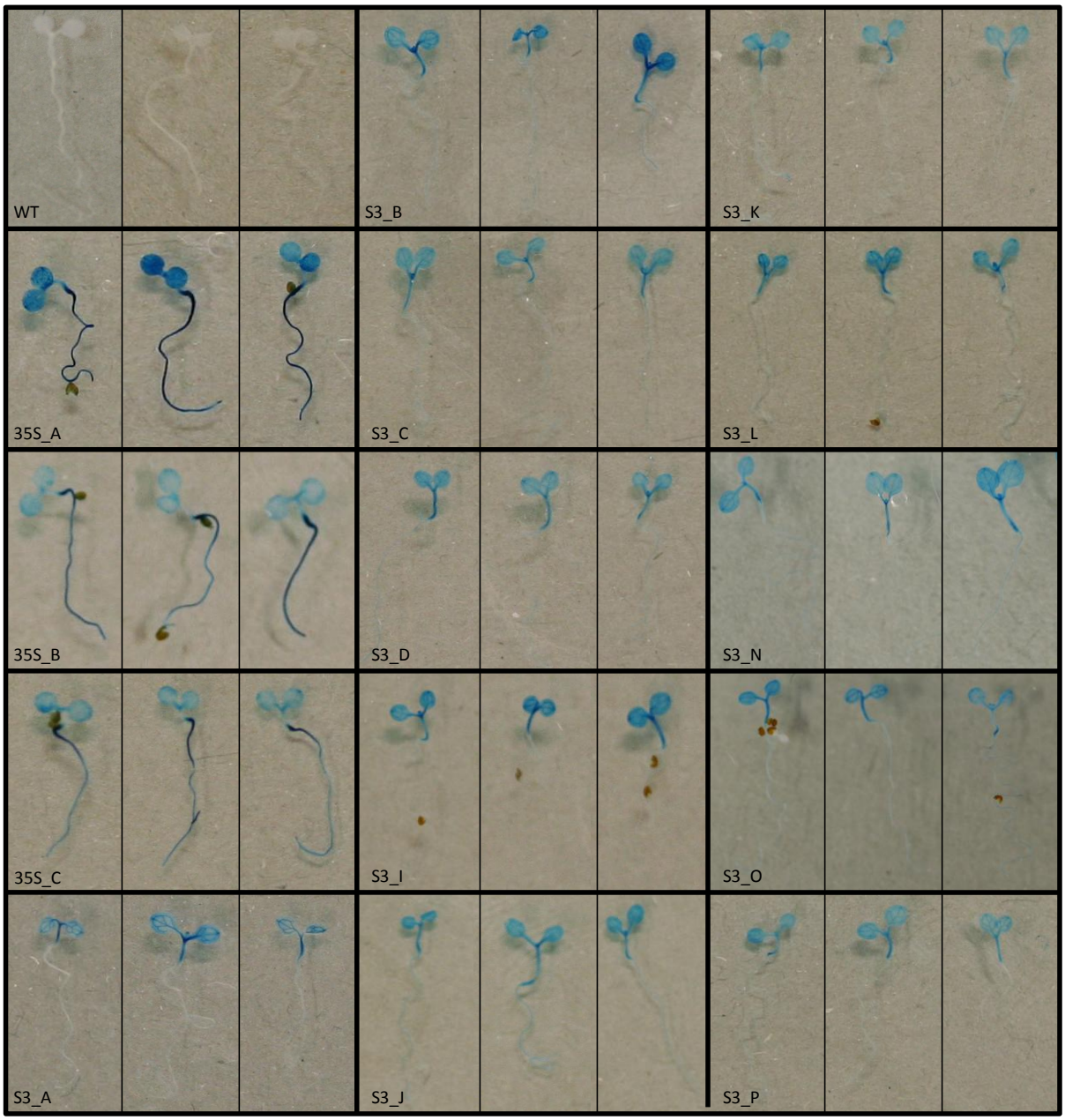
Supplementary Figure 10: Six-week-old *EgrSND2 1.5kb promoter::GUS* plant *GUS*-stained lower inflorescence stem microscope sections. S2_A, S2_D, S2_E, S2_F, S2_J, S2_L, S2_M, S2_N and S2_O are *EgrSND2 1.5kb promoter::GUS* plant line designations. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate *35SCaMV* promoter positive controls. Black scale bars indicate 100 nm.



Supplementary Figure 11: Six-week-old *EgrSND2 1.5kb promoter::GUS* plant *GUS*-stained upper inflorescence stem microscope sections. S2_A, S2_D, S2_E, S2_F, S2_J, S2_L, S2_M, S2_N and S2_O are *EgrSND2 1.5kb promoter::GUS* plant line designations. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate *35SCaMV* promoter positive controls. Black scale bars indicate 100 nm.

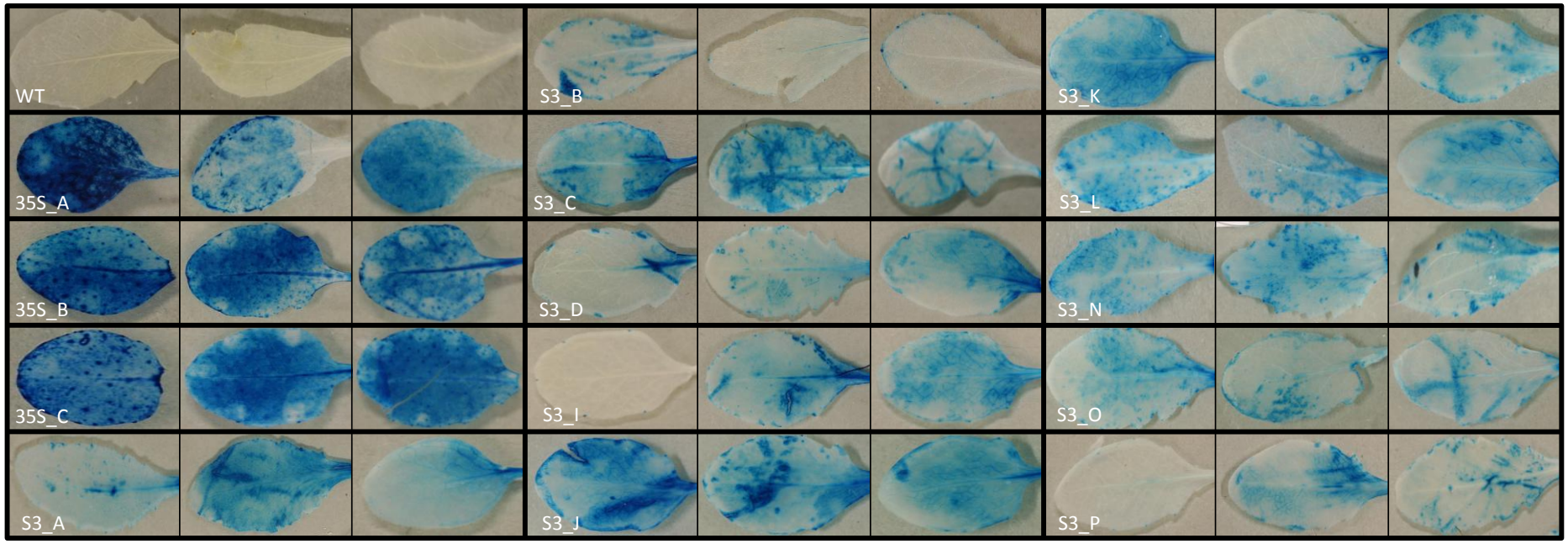


Supplementary Figure 12: Six-week-old *EgrSND2 1.5kb promoter::GUS* plant *GUS*-stained whole inflorescence stem segments. In each picture, an upper stem segment is seen at the top and a lower stem segment is seen at the bottom. S2_A, S2_D, S2_E, S2_F, S2_J, S2_L, S2_M, S2_N and S2_O are *EgrSND2 1.5kb promoter::GUS* plant line designations. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate *35SCaMV* promoter positive controls.

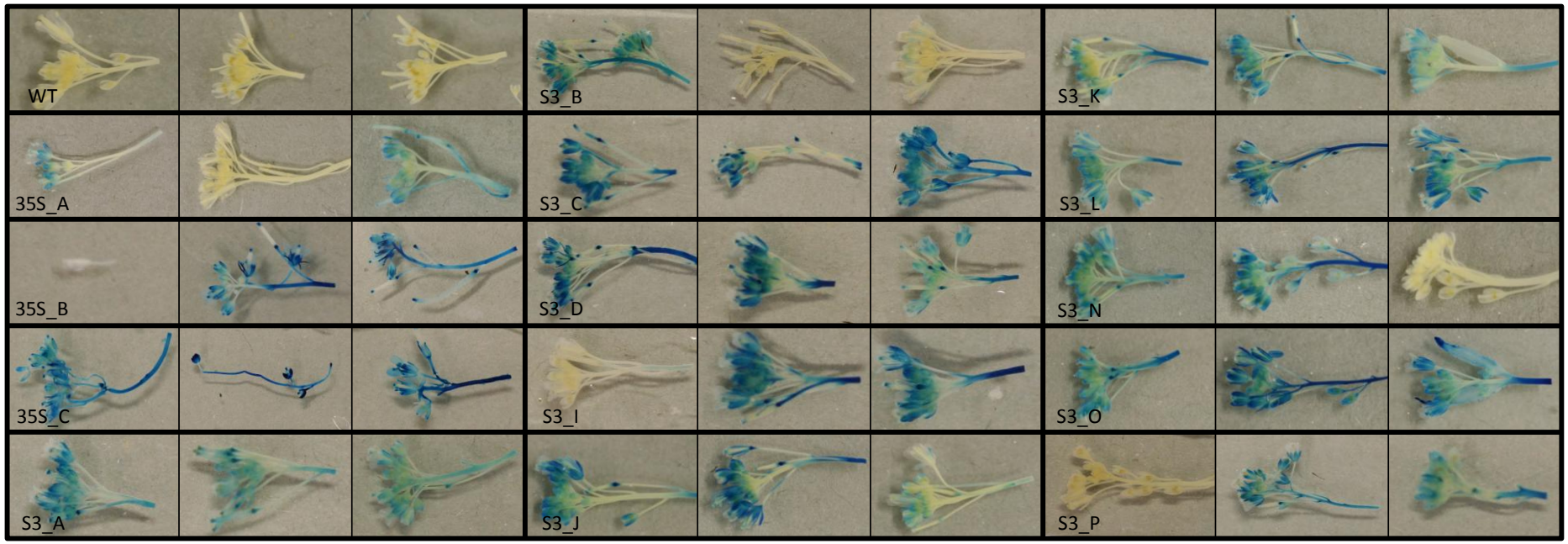


Supplementary Figure 13: One-week-old *EgrSND3 1.5kb promoter::GUS* plant *GUS*-stained whole plants. S3_A, S3_B, S3_C, S3_D, S3_I, S3_J, S3_K, S3_L, S3_N, S3_O and S3_P are *EgrSND3 1.5kb promoter::GUS* plant line designations. Three replicates of each are shown. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate *35SCaMV* promoter positive controls.

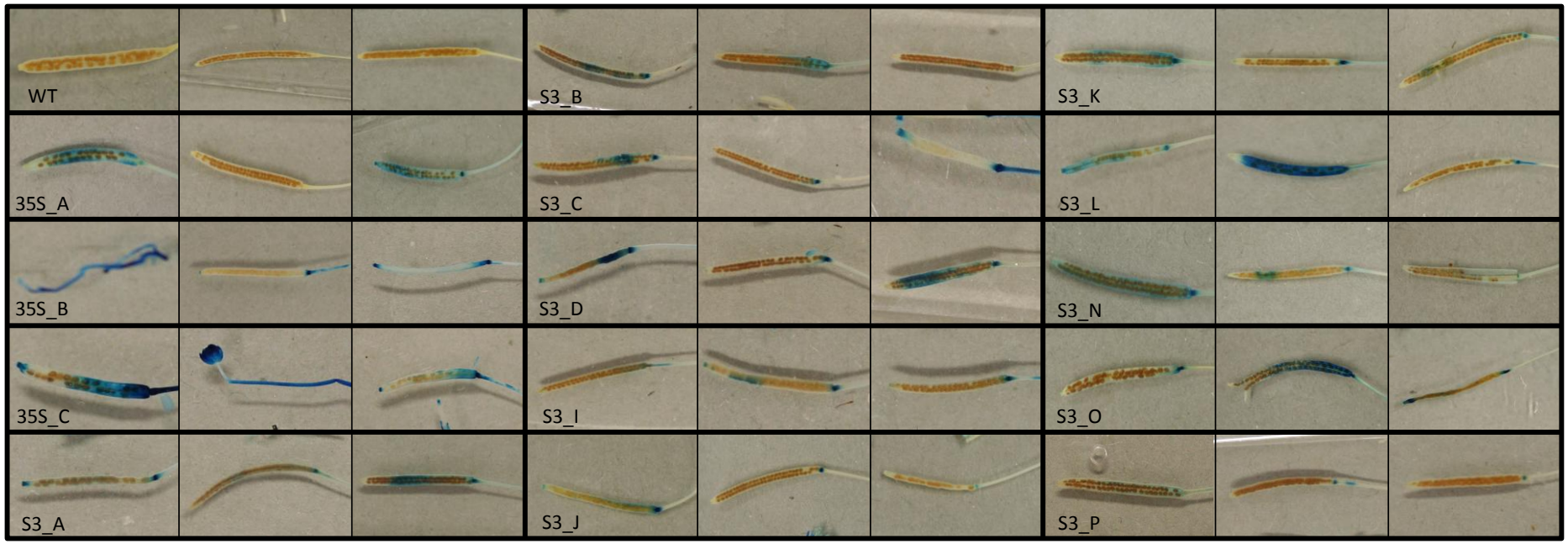
Supplementary Figure 14: Three-week-old *EgrSND3 1.5kb promoter::GUS* plant *GUS*-stained whole plants. S3_A, S3_B, S3_C, S3_D, S3_I, S3_J, S3_K, S3_L, S3_N, S3_O and S3_P are *EgrSND3 1.5kb promoter::GUS* plant line designations. Three replicates of each are shown. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate *35SCaMV* promoter positive controls.



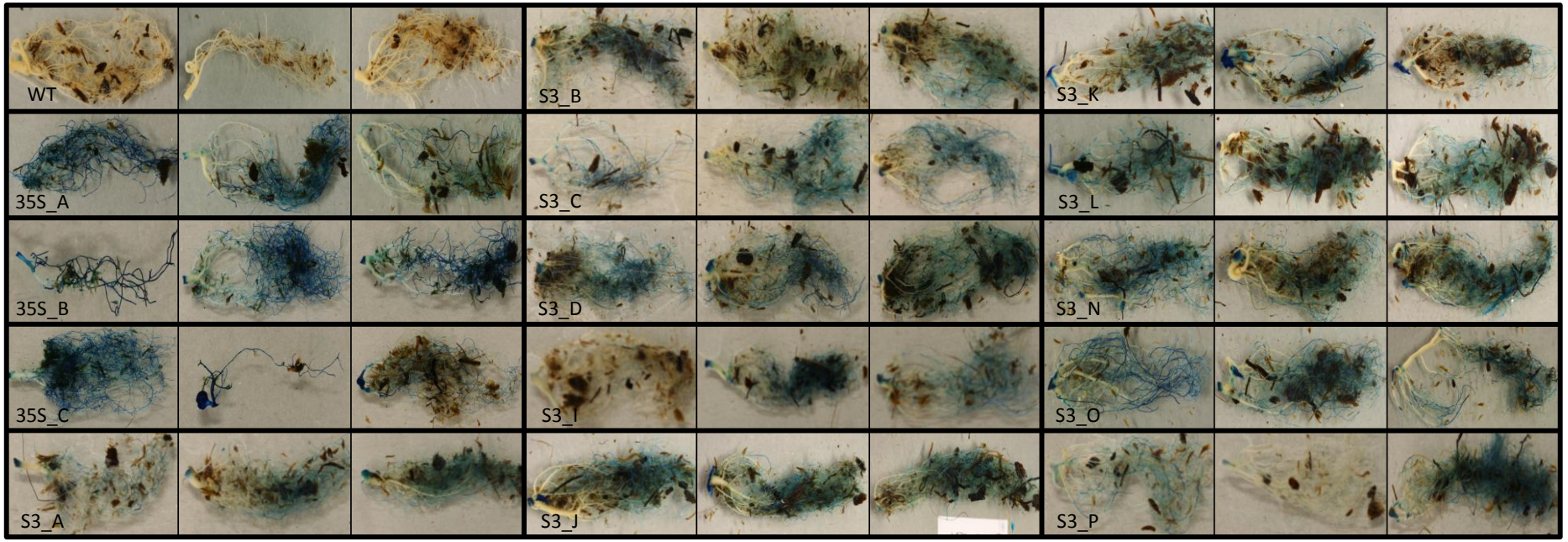
Supplementary Figure 15: Six-week-old plant *GUS*-stained leaves. S3_A, S3_B, S3_C, S3_D, S3_I, S3_J, S3_K, S3_L, S3_N, S3_O and S3_P are *EgrSND3 1.5kb promoter::GUS* plant line designations. Three replicates of each are shown. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate *35SCaMV* promoter positive controls.



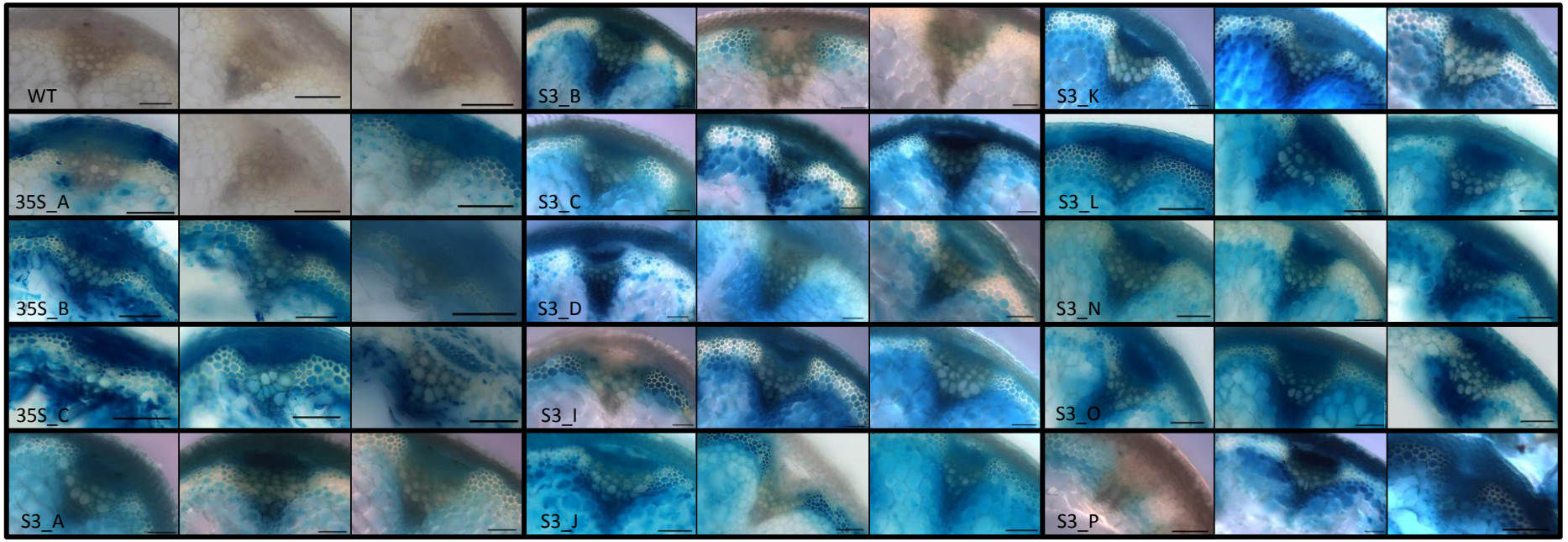
Supplementary Figure 16: Six-week-old plant GUS-stained flowers. S3_A, S3_B, S3_C, S3_D, S3_I, S3_J, S3_K, S3_L, S3_N, S3_O and S3_P are *EgrSND3 1.5kb promoter::GUS* plant line designations. Three replicates of each are shown. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate *35SCaMV* promoter positive controls.



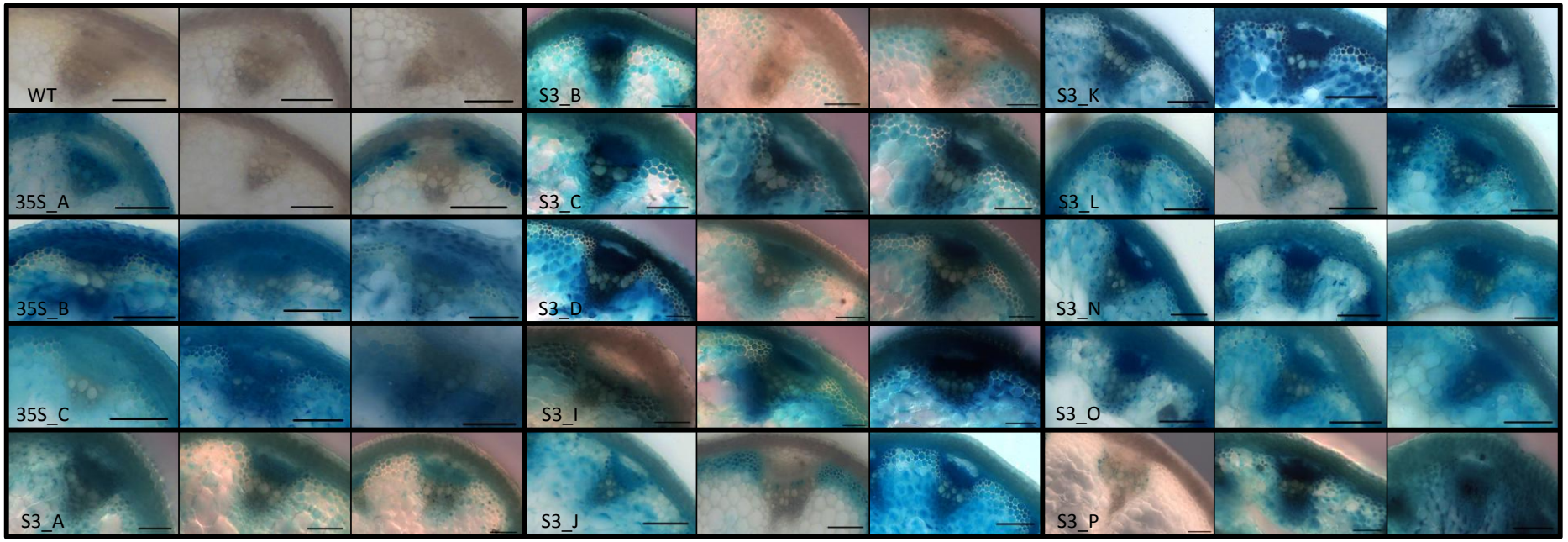
Supplementary Figure 17: Six-week-old plant *GUS*-stained siliques. S3_A, S3_B, S3_C, S3_D, S3_I, S3_J, S3_K, S3_L, S3_N, S3_O and S3_P are *EgrSND3 1.5kb promoter::GUS* plant line designations. Three replicates of each are shown. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate *35SCaMV* promoter positive controls.



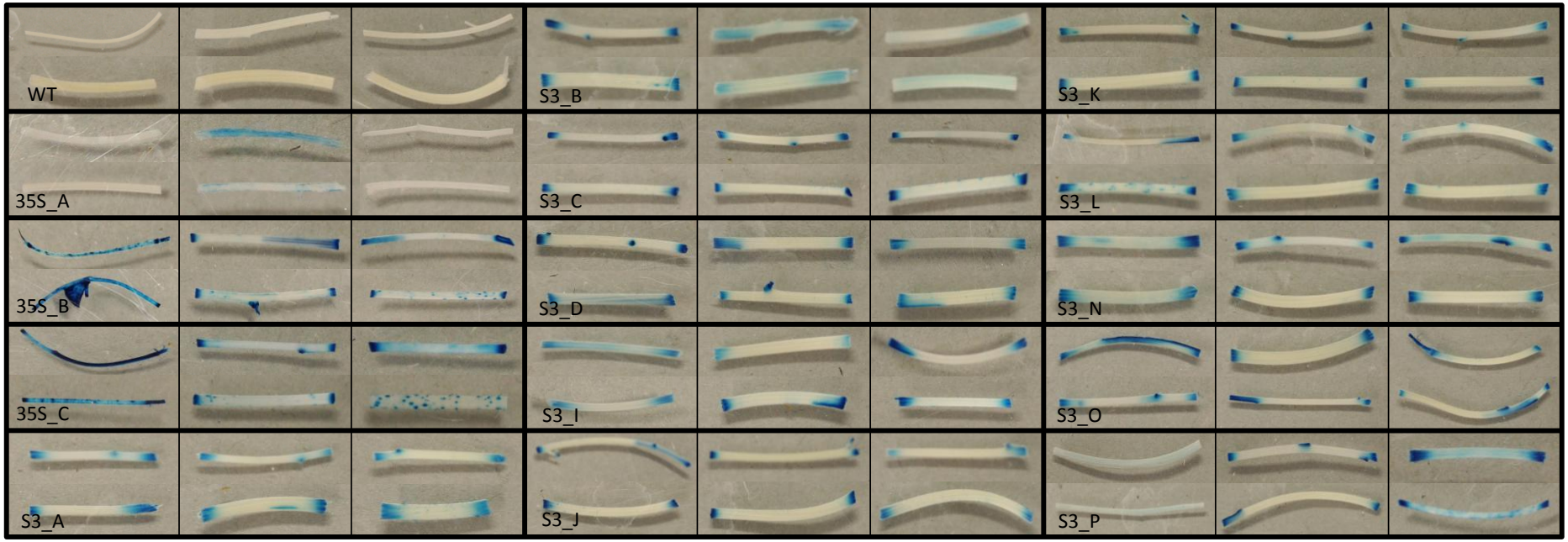
Supplementary Figure 18: Six-week-old plant *GUS*-stained roots and hypocotyls. S3_A, S3_B, S3_C, S3_D, S3_I, S3_J, S3_K, S3_L, S3_N, S3_O and S3_P are *EgrSND3 1.5kb promoter::GUS* plant line designations. Three replicates of each are shown. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate *35SCaMV* promoter positive controls.



Supplementary Figure 19: Six-week-old plant *GUS*-stained lower inflorescence stem microscope sections. S3_A, S3_B, S3_C, S3_D, S3_I, S3_J, S3_K, S3_L, S3_N, S3_O and S3_P are *EgrSND3 1.5kb promoter::GUS* plant line designations. Three replicates of each are shown. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate *35SCaMV* promoter positive controls. Black scale bars indicate 100 nm.



Supplementary Figure 20: Six-week-old plant *GUS*-stained upper inflorescence stem microscope sections. S3_A, S3_B, S3_C, S3_D, S3_I, S3_J, S3_K, S3_L, S3_N, S3_O and S3_P are *EgrSND3 1.5kb promoter::GUS* plant line designations. Three replicates of each are shown. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate *35SCaMV* promoter positive controls. Black scale bars indicate 100 nm.



Supplementary Figure 21: Six-week-old plant *GUS*-stained whole inflorescence stem segments. In each picture, an upper stem segment is seen at the top and a lower stem segment is seen at the bottom. S3_A, S3_B, S3_C, S3_D, S3_I, S3_J, S3_K, S3_L, S3_N, S3_O and S3_P are *EgrSND3 1.5kb promoter::GUS* plant line designations. Three replicates of each are shown. WT indicates the wild-type negative control; 35S_A, 35S_B and 35S_C indicate *35SCaMV* promoter positive controls.