In silico and functional characterization of the *cellulose synthase 8* gene promoter of *Eucalyptus* trees

by

NICOLE MARIE CREUX

Submitted in partial fulfilment of the requirements for the degree

Philosophiae Doctor

In the Faculty of Natural and Agricultural Sciences

Department of Genetics

University of Pretoria

Pretoria

July 2013

Supervisor: Prof. Alexander A. Myburg

Co-supervisor: Dr Christine Maritz-Olivier

Declaration

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

_____

Nicole Marie Creux

_____

Date

TABLE OF CONTENTS

Preface

Cellulose is a highly abundant biopolymer used for many different industrial applications in the paper, textiles, plastics and food industries. Cellulose is deposited into the plant cell wall by a large protein complex embedded in the cell membrane. Cellulose synthase (CesA) protein complexes are comprised of several different CesA subunits each producing a single cellulose chain, which ultimately combine with other chains to form cellulose microfibrils. The *CesA* gene family has been characterised in several different plant species and the number of genes in this family vary from 10 members in *Arabidopsis* to 18 in *Populus* mostly reflecting genome duplication and subsequent gene loss. Expression analysis of the different *CesA* genes revealed the presence of two main expression groups, one associated with primary cell wall formation and another associated with secondary cell wall (SCW) deposition. However, the *CesA* gene expression can also vary within these expression groups depending on the tissue type, developmental stage and the environment of the cell during cellulose deposition. While there is a wealth of information on the *CesA* genes or proteins and how they function, there is comparatively little information on the regulatory regions of these genes and how they modulate these complex expression patterns.

Recently, the main components of a transcriptional network regulating SCW formation in *Arabidopsis* has been elucidated and a number of transcription factors have been identified that regulate SCW biosynthetic genes, including those involved in cellulose biosynthesis. Several SCW-related transcription factors have been shown to activate the *CesA* promoters in *Arabidopsis* and *Populus,* but only *At*BES1 and *At*MYB46 have been shown to directly interact with the promoters of *CesA* genes. The *At*MYB46 binding site (M46RE/SMRE) has been experimentally verified in the *At*CesA promoters and no other studies have investigated which transcription factors interact with the specific regions of a

CesA promoter. While there has been great progress in the reconstruction the *Arabidopsis* SCW regulatory network, almost nothing is known about the SCW regulatory network of *Eucalyptus* which is currently one of the major cellulose producing crops world-wide. The **aim** of this PhD study was to elucidate the regulatory modules of the SCW-related *EgrCesA8* promoter and identify transcription factors which interact with these modules. To achieve this aim the promoters of the six *CesA* genes assessed in our previous study (Creux et al. 2008) were isolated from 13 different *Eucalyptus* species and highly conserved regions which may be functionally conserved at the genus level were identified for each *CesA* gene. The conserved regions of the *CesA8* promoter were functionally tested by promoter deletion studies where reporter gene expression was monitored in *Arabidopsis*, *Eucalyptus* and *Populus* plants. Selected regions of the *EgrCesA8* promoter were then used as bait fragments in a yeast-1-hybrid (Y1H) screen with a panel of 14 SCW-related transcription factors cloned from *Eucalyptus.* Several of the *Eucalyptus* SCW-related transcription factors were identified as potential regulators of *EgrCesA8* gene expression.

Chapter 1 of the dissertation provides a brief review of the literature on the *CesA* genes and proteins and what is currently known about *CesA* gene regulation. In the review, the issue of multiple published names for the *Eucalyptus CesA* genes is highlighted and a new naming convention, similar to *Populus*, is proposed and used consistently throughout the dissertation. An activation and interaction network is also constructed from the literature which shows all of the SCW-associated transcription factors that activate or directly interact with the *CesA* promoters in *Arabidopsis* and *Populus*.

Only a handful of studies have investigated the regulatory modules of *CesA* promoters. In a previous study (Creux et al. 2008) we isolated six *CesA* genes from *Eucalyptus* and then performed a comparative *cis*-element analysis to identify sequences that were over-represented in the *CesA* promoters of *Arabidopsis*, *Populus* and *Eucalyptus*.

Chapter 2 reports on a detailed analysis of DNA sequence variation in six *CesA* promoters from 13 different *Eucalyptus* species. The chapter describes the estimation of species-level nucleotide diversity and mapping of previously published *cis*-elements to the promoters. This enabled us to identify conserved promoter regions which coincided with clusters of *cis*-elements possibly representing functionally constrained regions of the promoters.

Chapter 3 discusses the results of a series of promoter deletions performed on the *EgrCesA8* promoter to identify potential regulatory modules. Several *EgrCesA8* promoter regions were identified which affected reporter gene expression in *Arabidopsis* and *Eucalyptus*. Part of this research required the optimization of the induced somatic sector analysis (ISSA) for testing of promoter::reporter gene constructs in the woody tissues of *Eucalyptus* and *Populus* and this published work appears as an appendix to this dissertation (**Appendix A)**.

There is little or no information on the SCW-regulatory network in *Eucalyptus* or how these transcription factors may regulate *CesA* gene expression. **Chapter 4** reports the results of a Y1H screen of the *EgrCesA8* regulatory modules (baits) against a panel of 14 SCW-related *Eucalyptus* transcription factors (prey). This section also identifies several binding sites which were conserved in the *CesA8* promoter of 13 *Eucalyptus* species and may facilitate

some of the observed protein-DNA interactions. A model of *EgrCesA8* gene regulation is also provided in this chapter and is based on the combined results of this dissertation.

Concluding Remarks are provided in **Chapter 5**. In this section the results of the dissertation are put into perspective and future projects or prospects stemming from this research are discussed. Conclusions are drawn on the significance of the findings to the field and possible implication for biotechnological applications.

The findings presented in this PhD dissertation represent the outcomes of a study undertaken from June 2007 to June 2013 in the Department of Genetics, University of Pretoria, under the supervision of Prof. A.A. Myburg and Dr C. Martiz-Olivier. Chapter 2 and Appendix A have been published in peer reviewed journals and Chapter 3 and 4 were also prepared in the format of independent manuscripts to maintain the style of the thesis. A certain degree of redundancy may therefore exist between the introductory sections of these chapters and Chapter 1. To submit the results for publication, it is our intention to combine the results of Chapter 3 and 4.

List of publications from the thesis work

Refereed Publications

Creux NM, Bossinger G, Myburg AA, Spokevicius AV (2013) Induced somatic sector analysis of cellulose synthase (*CesA*) promoter regions in woody stem tissues. Planta 237: 799-812

Creux NM, De Castro MH, Ranik M, Maleka MF, Myburgh AA (2013) Diversity and *cis*-element architecture of the promoter regions of cellulose synthase genes in *Eucalyptus*. Tree Genetics and Genomes 9: 989-1004

Refereed conference proceedings

Creux NM, De Castro MH, Ranik M, Spokevicius AV, Bossinger G, Maritz-Olivier C, Myburg AA 2011. *In silico* and functional characterization of the promoter of a *Eucalyptus* secondary cell wall associated cellulose synthase gene (*EgCesA1*). BMC Proceedings 5(Suppl7):P107.

Posters and conference presentations

Creux NM, Ranik M, Myburg AA. (2007) Characterization of the promoter regions of six cellulose synthase genes in *Eucalyptus grandis*. American Society of Plant Biologists (ASPB), Hilton hotel, Chicago, Ill, 7 - 11 July. (Poster presentation 36046)

De Castro MH, **Creux NM**, Ranik M, Mathabatha FM, Myburg AA. (2009) Molecular evolution of the promoter regions of cellulose synthase genes in *Eucalyptus*. IUFRO tree Biotechnology Conference, Whistler, Canada, 28 June - 2 July. (Poster presentation)

Creux NM, De Castro MH, Ranik M, Myburg AA (2010) Functional characterization of the diversity and architecture in *Eucalyptus CesA* promoters. South African Genetics Society (SAGS) Conference, Bloemfontein, 8 – 11 April. (Oral presentation - Awarded best PhD oral presentation)

Creux NM, De Castro MH, Ranik M, Myburg AA. (2010) Diversity, architecture and function of *cis*-regulatory sequences in the promoters of cellulose synthase genes in *Eucalyptus* plantation trees. American Society of Plant Biologists (ASPB), Montreal, Canada, 31 July - 4 August. (Poster presentation 08020)

Creux NM, De Castro MH, Ranik M, Spokevicius AV, Bossinger G, Maritz-Olivier C, Myburg AA (2011) *In silico* and functional characterization of the promoter of a *Eucalyptus* secondary cell wall associated cellulose synthase gene (*EgCesA1*). IUFRO Tree Biotechnology Conference, Arraial d'Ajuda, Bahia, Brazil, 26 June - 2 July (Poster presentation – Runner-up poster award)

Creux NM, Ranik M, De Castro MH, Spokevicius AV, Bossinger G, Maritz-Olivier C, Myburg AA (2012) Understanding the transcriptional regulation of the *Eucalyptus Cellulose synthase1* gene. 38th Conference of South African Association of Botanists (SAAB), Pretoria, 15 - 18 January. (Oral presentation)

Creux NM, Martiz-Olivier C, Spokevicius AV, Bossinger G, Myburg AA (2013) DNA-protein interactions modulating *Eucalyptus grandis cellulose synthase 8* gene expression. IUFRO Tree Biotechnology Conference, Asheville, NC, USA, 26th May – 1st June. (Poster presentation)

Dedication

I would like to dedicate this work to the memory of my uncle, Prof Renier van Rooyen, for being one of the great academic inspirations in my life. He was a Professor in Medicine at the University of Pretoria where he taught passionately for many years and held various medical degrees including MBChB, MMed (Int) and MD (Endo). As a hobby he studied many subjects part-time and obtained several other degrees including BSc, MSc, Bluris, LLM and additional qualifications in Psychology, Philosophy, Statistical methods, Geology and Biochemistry. His insatiable thirst for knowledge and interest in a broad array of topics has always inspired me and guided me along my academic path.

Acknowledgements

I gratefully acknowledge the following people, organizations and institutes for assisting me in the completion of this study:

- To Prof. Zander Myburg, his continued support, guidance and patience throughout my postgraduate career are greatly appreciated. The strong leadership he provided while still allowing me the freedom to explore the field both at home and aboard has enhanced both this study and my understanding of the subject. I am also thankful for the exciting and dynamic research atmosphere he has created within the group and the time he has taken to edit my manuscripts and chapters during this degree.
- To Dr Christine Maritz-Olivier, the insight, advice and support she provided was crucial to the completion of this study. I greatly appreciated the time she took to don a lab coat and personally teach me some tricks of the trade. I am also very thankful for the shoulder to cry on during the difficult stages of this study and now I know why her students refer to her as their “lab mom”.
- To Mr Martin Ranik for his advice and contributions to Chapter 2 and its publication. I am also thankful to him for being a supportive mentor and friend through all of my postgraduate studies. His encouragement and help has kept me going through it all.
- To Ms Minique (Ouma) de Castro, she has been a fierce friend and colleague for many years. Her attention to detail is unparalleled and was vital to the analysis and preparation of the Chapter 2 manuscript. I have enjoyed all our discussion, especially when they were accompanied by a “wyntjie”.
- To Dr Antanas V. Sopkevicius for mentoring and teaching me all there is to know about ISSA. His contribution, guidance and advice were crucial to publishing the Appendix A manuscript. I will always be grateful to him and his wife (Rachaele) for

graciously opening their home to me when I was visiting the University of Melbourne; they made me feel like part of the family.

- To Prof Gerd Bossinger and his research group, they made me feel welcome at the University of Melbourne and provided a fantastic work environment. I enjoyed the months I spent in his laboratory with him and the team. His support, insight and advice on the Appendix A manuscript were invaluable and greatly appreciated.
- To my past mentorship students, Andrew De Santos and Matt Laubscher, for tackling every task I set them with enthusiasm. They worked tirelessly to screen all of the *Arabidopsis* plants used in this study and for this I am very grateful.
- To my past and present Honours students, Desre Pinard and Matt Laubscher for reminding me how exciting science is and for letting me try out my mentoring skills on them.
- To all the past and present members of the Forest Molecular Genetics (FMG) programme and the *Eucalyptus* Pine Pathogen Interactions (EPPI) for all their support and friendship during this study. A special thank you to Ms Elna Cowley (Laboratory Manager) and Ms Adri Veale (*Arabidopsis* Facility Manager) for all their efforts in keeping the labs running smoothly. I would also like to thank Mr Eshchar Mizrachi and Mr Steven Hussey for reviewing Chapter 1 and 4 respectively, their insight was greatly appreciated. My thanks also goes to Mr Jonathan Botha for cloning and sequencing the transcription factors for Y1H prey panel as this added great value to the study.
- To my parents for their unwavering support both emotionally and financially without which this study would not have been completed. I am also grateful for all the love and friendship they have given me. I am thankful to my sister (Michelle) and her

husband (Andre) for always opening their home to me and providing emotional support when I needed a break.

- To my special friends (Bronwyn, Jana, Vinet, Esh, Jo, Ritesh, Marja and Marius) for all the love and support they have given me. The kind words of encouragement and hearty laughter helped me maintain the momentum to complete this study. I am also thankful to Dr Martin Kemler for reminding me that science is not only about genes and proteins but encouraged me to think more broadly and read more widely.
- To Prof Paulette Bloomer and the Genetics Department of the University of Pretoria for providing a sound, nurturing academic environment and exposure to all aspects of science.
- To Prof Michael J. Wingfield and the Forestry and Agricultural Biotechnology Institute (FABI) for the opportunities I have received to explore different aspects of academia and the excellent facilities which I used during the study.
- To Sappi and Mondi for providing plant material, strong support and funding contributions during this study.
- To the National Research Foundation of South Africa (NRF), for funding supplied by the Scarce Skills Scholarship.
- To The Human Resources and Technology for Industry Programme (THRIP), for financial support of the research.

THESIS SUMMARY

In silico and functional characterization of the *cellulose synthase 8* gene promoter of *Eucalyptus* trees

Nicky M. Creux

Supervised by ***Prof A.A. Myburg*** *and* ***Dr C. Maritz-Olivier***

Submitted in partial fulfilment of the requirements for the degree ***Philosophiae Doctor***

Department of Genetics

University of Pretoria

Cellulose is a highly abundant biopolymer which forms the basis of several industrial applications including paper and textile products. Cellulose is deposited into the plant cell wall by a large membrane bound protein complex which is comprised of different cellulose synthase (CesA) subunits. Plants maintain several different *CesA* genes which have specific expression patterns depending on the cell wall type, tissue type, developmental stage and environment of the cell. While *CesA* genes and proteins have been the focus of many studies, the upstream regulatory regions which govern their complex expression patterns have remained largely unexplored. The aim of this study was to use the previously identified *CesA* promoter regulatory modules and putative *cis*-elements to identify conserved *cis*-element clusters in the *Eucalyptus CesA* and the transcription factors which interact with the regulatory regions of the *EgrCesA8* promoter.

The promoters of six cellulose synthase genes (*CesA1*, *CesA3*, *CesA6*, *CesA4*, *CesA7* and *CesA8*) were isolated from 13 *Eucalyptus* species of different sections in the *Symphyomyrtus* subgenus of *Eucalyptus*. Species-level nucleotide diversity was calculated for the promoters of each gene. The promoters each contained a highly conserved region at the transcriptional start site (TSS), possibly marking the core promoter. The *Eucalyptus* promoters appeared to be TATA-less and *cis*-elements which resembled alternate core plant promoter elements were found clustered close to the TSS. Other localised regions of low species-level nucleotide diversity were identified upstream of the TSS in each promoter set and could indicate the location of *cis*-regulatory modules (CRMs).

The conserved promoter regions and *cis*-element maps of the SCW-associated *EgrCesA8* promoter were used to direct promoter truncation for reporter gene analysis in *Arabidopsis*, *Eucalyptus* and *Populus*. Comparative analysis of the *cis*-element maps and GUS expression data revealed that two main conserved regions of the *CesA8* promoter harboured clusters of *cis*-elements and modulated GUS expression. The $CT_{(11)}$-microsatellite in the conserved TSS-associated *cis*-element cluster produced strong non-specific GUS expression in *Eucalyptus* and *Arabidopsis* when appended to the 5'UTR which suggests a role in the *EgrCesA8* core promoter. Further upstream in the promoter a second conserved promoter region coincided with a cluster of SCW-associated *cis*-elements and caused a loss of expression in leaf vasculature, suggesting a role for this CRM in modulating tissue-specific expression of *EgrCesA8*.

The conserved *EgrCesA8* promoter regions which coincided with *cis*-element clusters and GUS regulatory modules were used as baits in a yeast-1-hybrid screen against the a panel of 14 *Eucalyptus* SCW transcription factors. *Egr*MYB31 (*At*MYB46) and *Egr*ZincFinger-A were found to interact with the *EgrCesA8* 5'UTR. The interaction of these transcription factors with the 5'UTR were blocked by the presence of the $CT_{(11)}$-microsatellite and could

explain the loss of tissue-specific expression. *Egr*NAC170 (*At*SND2) directly interacted with the CRM containing a cluster of SCW-related *cis*-elements. A dual Y1H assay revealed that *Egr*KNAT7 in the presence of *Egr*MYB80 (*At*MYB52) or *Egr*MYB87 (*At*MYB54) could also interact with the *ErgCesA8* CRM. Together the results of the dissertation indicate that *EgrCesA8* regulation is modulated by different protein-DNA and protein-protein interactions acting at highly conserved regions of the promoter.

List of Tables

List of Figures

Appendix and Supplementary Data

All supplementary data are listed as separate electronic files on the accompanying disc secured on the back cover of the dissertation. **Appendix A** can be located at the end of the dissertation (pg 215).

Appendix

Appendix A: Creux NM, Bossinger G, Myburg AA, Spokevicius AV (2013) Induced somatic sector analysis of *cellulose synthase* (*CesA*) promoter regions in woody stem tissues. Planta 237:799-812

Supplementary Files

Chapter 2

Supplementary file 2.1: *Eucalyptus* species used in this study were predominantly from the commercially important subgenus *Symphyomyrtus* and one out group from the subgenus *Eucalyptus*.

Supplementary file 2.2: Primer sequences for the amplification of promoter regions of six *Eucalyptus* cellulose synthase genes, as well as the first exon and intron of the *CesA8* gene.

Supplementary file 2.3: *Arabidopsis thaliana* cellulose synthase (*AtCesA*, TAIR9 – www.arabidopsis.org) genes and their corresponding *Populus trichocarpa* (*PtiCesA*, Kumar et al. 2009, http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html) and *Eucalyptus grandis* (*EgCesA*, Yin et al. 2009, Ranik and Myburg 2006) orthologs included in this study.

Supplementary file 2.4: Alignments of the promoter sequences of 13 *Eucalyptus* species for each of the six orthologous groups of *cellulose synthase* genes indicating the locations of mapped *cis*-element occurrences.

Supplementary file 2.5: Graphical representation of two *Eucalyptus CesA1* promoter haplotypes observed in this study and nucleotide diversity within the region.

Supplementary file 2.6: Excel spreadsheet of the raw data and statistical analysis for the evaluation of *cis*-element evolution

Supplementary file 2.7: Comparison of consensus motif sequences at the genus (*Eucalyptus*, *Populus* and *Arabidopsis*) and species (*Eucalyptus*) level.

Chapter 3

Supplementary file 3.1: GUS expression in each line for the eight *EgrCesA8* promoter fragments and the 35S positive control in *Arabidopsis* leaves.

Supplementary file 3.2: GUS expression in each line for the eight *EgrCesA8* promoter fragments and the 35S positive control in *Arabidopsis* roots and hypocotyls.

Chapter 4

Supplementary file 4.1: Yeast cell suspension calculations and raw plate images for each bait promoter fragment and transcription factor combination.

Supplementary file 4.2: Raw images of the plates used in the yeast-1-hybrid screens with varying concentrations of 3-amino-1,2,4-triazole (3AT) to eliminate leaky HIS expression and identify interactions with baits that confer auto-activation.

Supplementary file 4.3: Yeast cell suspension calculations, raw images of the plates used in the dual yeast-1-hybrid screens on selective media and PCR confirmation that both prey vectors are present.

Chapter 1

LITERATURE REVIEW

Transcriptional regulation of *cellulose synthase* genes

1.1 Introduction

Cellulose is a highly abundant biopolymer used by many different industries for applications in pulp, paper, textiles and biofuel production. Cellulose consists of glucose molecules which have been joined together by β-(1-4)-glycosidic linkages. These glucan chains then aggregate to form the cellulose microfibrils that are deposited into the plant cell wall (reviewed in Somerville 2006). Cellulose is produced by the cellulose synthase complex (CSC) which is a large rosette shaped protein complex consisting of multiple cellulose synthase (CesA) proteins embedded in the cell membrane (reviewed in Doblin et al. 2002). In the CSC, CesA composition varies depending on whether a primary or secondary cell wall (SCW) is deposited. Primary cell wall-associated CSCs are predominantly composed of CesA1, CesA3 and CesA6, while the SCW CSCs consist of CesA4, CesA7 and CesA8 (Wang et al. 2008; Atanassov et al. 2009; Song et al. 2010), but the exact composition, ratio and arrangement of CesA proteins in the CSC is not known.

Plant-specific *CesA* genes have been identified in a range of plant species from green algae and moss to higher order species such as *Populus* and *Arabidopsis* (Yin et al. 2009). Expression pattern analysis supported the findings that three *CesA* genes are required during SCW formation and a separate set of *CesA* genes are expressed during primary cell wall deposition (reviewed in Taylor 2008). Phylogenetic analysis of the *CesA* genes revealed that this gene family is highly conserved across different plant species (Vergara and Carpita 2001; Ranik and Myburg 2006; Liang and Joshi 2004; Carroll and Specht 2011). Kumar et al. (2009) suggested that based on these relationships the previously identified *Populus CesA* genes should be renamed to reflect the expression and phylogenetic links to their *Arabidopsis* counterparts. Other expression studies have also suggested that *CesA* genes may respond to different stresses such as mechanical, drought and pathogen attack (Bhandari et al. 2006; Hernández-Blanco et al. 2007; Cal et al. 2013), indicating that *CesA* genes require a number

of regulatory mechanisms to ensure their expression patterns are maintained and modulated during the life cycle of the plant.

The deposition of cellulose, hemicelluloses and lignin during SCW formation is regulated by a network of transcription factors which modulate the precise spatio-temporal expression patterns observed for the cell wall biosynthetic genes (reviewed in Demura and Ye 2010; Zhong et al. 2010a). The SCW regulatory network is mainly comprised of the MYB (MYELOBLASTOSIS) and NAC (NAM/ATAF/CUC) transcription factor families (Stracke et al. 2001; Wilkins et al. 2009; Hu et al. 2010). MYB transcription factors such as *At*MYB46, *At*MYB83 and *At*MYB63 have been shown to interact with the promoters of cellulose, lignin and hemicellulose biosynthetic genes respectively (Goicoechea et al. 2005; McCarthy et al. 2009; Zhou et al. 2009; Zhong et al. 2011; Kim et al. 2013). Similarly, the NAC transcription factors such as the vascular-related NAC domain proteins (*At*VND6 and *At*VND7) and secondary wall-associated NAC domain proteins (*At*SND1) have also been shown to directly interact with the promoters of SCW structural genes (Ohashi-Ito et al. 2010; Zhong et al. 2010c; Yamaguchi et al. 2011). While these studies have gone a long way in elucidating the SCW-regulatory network, there are no studies which focus on a particular set of promoters within the network, such as the cellulose synthase gene family's promoters.

The promoter regions of the *CesA* genes are not as well characterised as their coding counterparts and have been the focus of only a few studies. The *CesA* promoter regions have been isolated from plants species such as *Eucalyptus*, *Populus* and cotton (Creux et al. 2008; Lu et al. 2008; Wu et al. 2009; Ko et al. 2012). These promoter regions modulated tissue-specific reporter gene expression *in planta* and were subjected to deletion studies, where specific regions were identified that could regulate reporter gene expression in a tissue-/response-dependent manner (Lu et al. 2008; Wu et al. 2009). The *CesA* promoters have also been subjected to *in silico* analysis where conserved sequence motifs were identified and

were suggested to play a role in the regulation of *CesA* expression (Creux et al. 2008; Lu et al. 2008). Only two transcription factors (*At*MYB46 and *At*BES1) have been directly linked to their respective *cis*-elements in the *CesA* promoters (Xie et al. 2011; Zhong et al. 2011; Kim et al. 2013).

This review will first briefly give an overview of the CesA proteins and how they function but this will not be a detailed review as this has been extensively reviewed in the past (Somerville 2006; Taylor 2008; Carpita 2011; Lei et al. 2012). Next the conservation of these genes across vast evolutionary distances and how the relationships of these genes can be used to assign orthology is discussed. A new naming convention for the *Eucalyptus CesA* genes is also proposed which could alleviate the confusion surrounding genes with multiple names. Finally, the review concludes with a survey of what is currently known about plant *CesA* gene regulation in terms of promoter analysis and interactions with transcription factors of the SCW-regulatory network.

1.2. The cellulose synthase proteins and genes

1.2.1. Cellulose synthase complexes (CSC)

Cellulose is deposited into the plant cell wall by a large membrane bound protein complex, which has a rosette structure with six subunits (Kimura et al. 1999). Each subunit consists of several different CesA proteins depending on the cell wall type (primary or secondary), the tissue type or developmental stage of the cell (reviewed in Taylor 2008). In *Arabidopsis* the SCW associated complexes are comprised of *At*CesA4, *At*CesA7 and *At*CesA8 and the deposition of the SCW is disrupted when a null mutant is produced for any one of these genes suggesting that all three are necessary for cellulose deposition in the SCW (Taylor et al.

2000; Gardiner et al. 2003; Taylor et al. 2003). Similarly, the primary-wall associated CSC is composed of three different CesA proteins, but which three CesA proteins may differ depending on tissue type (Desprez et al. 2007; Persson et al. 2007). *At*CesA1 and *At*CesA3 have been shown to be essential for the primary wall-associated CSCs and a null mutation in either of these genes is lethal to the plant (Persson et al. 2007). The third member of this complex is less constrained although in general it is reported to be *At*CesA6, which is partially redundant with *At*CesA2, *At*CesA5 and *At*CesA9 (Persson et al. 2007). Studies have shown that in specific tissues *At*CesA6 may be replaced, for example in the seed coat *At*CesA5 is involved in cell wall deposition and *At*CesA9 is active in more mature root cells during embryogenesis (Desprez et al. 2007). In *Arabidopsis* cells, primary wall-associated and secondary wall-associated CSCs have distinct roles and only CesA4, 7 and 8 subunits could be isolated from stem tissues depositing SCWs (Atanassov et al. 2009). However, in *Populus* immature xylem, both primary and SCW-associated CSCs were identified in the same cell (Song et al. 2010). The authors suggest that the type of CSC may be dependent on the cellulose crystalinity required and therefore both types of CSCs could occur at the same developmental stage during xylogenesis. The mechanism of switching from primary wall-associated complexes to SCW-associated complexes during plant cell development and the transcriptional regulation involved in this process has yet to be fully elucidated.

Many authors have proposed models for CSC movement within the cell membrane and the deposition of cellulose into the plant cell wall (reviewed in Lei et al. 2012). Previous studies have indicated that the CSCs are associated with the cortical microtubules and move bi-directionally along the microtubules (Paredez et al. 2006), but for a long time the mechanism of association remained elusive. Recently, the cellulose synthase interacting 1 (CSI1) protein was identified as the link between the CSCs and the cortical microtubules. *Arabidopsis* mutants for CSI1/POMPOM2 showed reduced CSC movement along the

microtubule tracks (Bringmann et al. 2012; Li et al. 2012b). The movement of the CSCs along the microtubules appears to be driven by the synthesis and deposition of the cellulose microfibrils into the cell wall, which pushes the complexes along the cell membrane (Robinson and Quader 1981; Paredez et al. 2006; Diotallevi and Mulder 2007). Differences in primary wall-associated and secondary wall-associated CSCs have also been noted with regards to their microtubule associations and velocity during cellulose deposition. Wightman et al. (2009) showed that GFP tagged *At*CesA5 (Primary wall-associated) particles moved at a much lower velocity than the SCW-associated *At*CesA7 YFP tagged particles suggesting there are key differences between the primary- and secondary-wall associated CSC and the mechanisms by which they are regulated.

Overall CesA protein structure is conserved, even when proteins from eukaryotes and prokaryotes are compared (Pear et al. 1996; Saxena et al. 2001). CesA proteins have a number of common features (Figure 1.1) which play a role in structure and function of the CSC. A RING finger zinc-binding domain, which contains four CXXC motifs, has been hypothesized to play a role in CesA-CesA interactions in the CSC (Doblin et al. 2002; Kurek et al. 2002). A short plant-specific region after the zinc-binding domain is termed the hyper variable region I (HVRI)/class specific region I (CSRI) and this region shows very low similarity to other *CesA* genes from the same plant species (Vergara and Carpita 2001). The CSRI is followed by the first two of eight transmembrane domains (Figure 1.1), after which the glycosyltransferase catalytic domain is located (Saxena et al. 2001). The glycosyltransferase catalytic domain loops into the cytosol and can be further dissected into different protein regions. The glycosyltransferase catalytic domain contains the conserved D,D,D,QXXRW protein motifs (Figure 1.1) which are believed to play a role in the glycosyltransferase activity and is the signature motif for all glycosyltransferase family 2 (GT2) proteins (Saxena et al. 1995; Campbell et al. 1997). The first two conserved aspartic

acid residues fall into the A domain of the catalytic region. In plants this region has an insertion which has been termed the plant-conserved region (Pear et al. 1996; Delmer 1999; Joshi and Mansfield 2007). After the plant-conserved region, a second hyper variable region (HVRII)/class specific region II (CSRII) was identified (Vergara and Carpita 2001), which also exhibited limited similarity between *CesA* genes from the same species. The B domain follows the CSRII and contains the rest of the highly conserved residues (D and QXXRW) in the catalytic region of the protein (Figure 1.1). Directly after the B domain, which is the end of the glycosyltransferase catalytic region, are six more transmembrane domains and the C-terminal of the protein (reviewed in Joshi and Mansfield 2007).

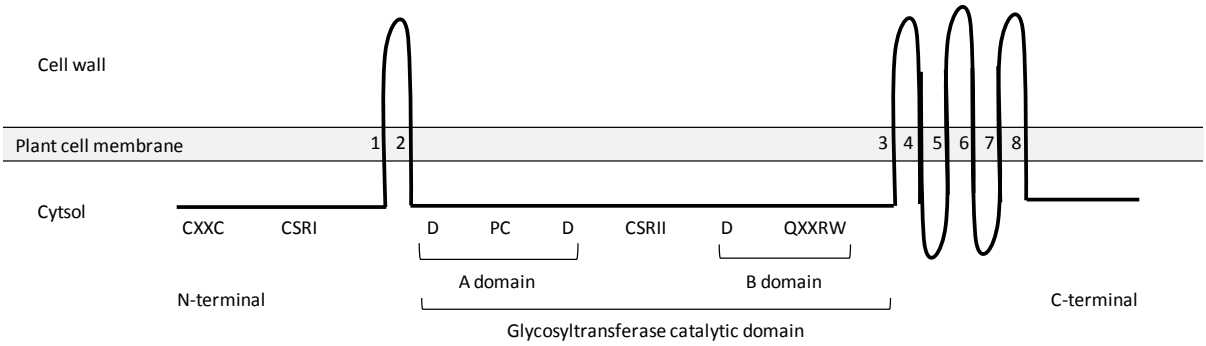


Figure 1.1 Line diagram of the general CesA structure compiled from the literature. The position of the cell wall, membrane (grey shading) and cytosol are indicated on the left. The zinc finger (CXXC), plant conserved region (PC), A domain, B domain, class specific region I and II (CSRI and CSRII) are indicated. The transmembrane domains (1-8) extend through the plant membrane. The glycosyltransferase catalytic domain containing the conserved aspartic acid residues (D) and the conserved QXXRW protein domain is indicated at the bottom of the figure.

The structural and functional models of CesA proteins have mainly been constructed from bacterial CesA studies because plant CesA proteins have proven more difficult to examine. Pear et al. (1996) isolated the first plant CesA and showed that the glycosyltransferase catalytic unit could bind UDP-glucose (the substrate for cellulose synthesis) when over expressed in *E. coli*. The cultures also showed that if part of the A-

domain containing the first conserved aspartic acid residue was deleted, the protein lost its ability to bind UDP-glucose. This indicates the importance of the highly conserved residues in the catalytic domain of both plant and bacterial CesA proteins. When these conserved CesA residues were mutated by site directed mutagenesis in a bacterial *CesA* gene the cellulose synthase activity decreased substantially suggesting these residues are important in both plant and bacterial CesA proteins (Saxena et al. 2001). Recently the first crystal structure of a CesA protein was resolved from *Rhodobacter sphaeroides* (Morgan et al. 2012). The model suggested that the first two conserved aspartate residues most likely co-ordinate the positioning of the UDP-glucose molecules, while the third aspatic acid is possibly the catalytic site of the reaction. The study also suggested that the QXXRW domain plays a role in the cytoplasmic entry to the glucan channel, which is surrounded by the transmembrane domains (Morgan et al. 2012). An independently predicted 3D model of a plant CesA protein (*Gh*CesA1) (orthologous to *At*CesA8) revealed many of the same structural features identified in the resolved structure by Morgan et al. (2012). However, a plant 3D CesA model was also able to address several plant-specific features (Sethaphong et al. 2013). The authors propose that the CSRII and the plant conserved region may play a role in CesA rosette formation. While the 3D plant model and the resolution of the *Rs*CesA structure have begun to provide clues to the function of various CesA protein domains, the plant specific-domains are yet to be convincingly resolved.

1.2.2. Cellulose synthase (CesA) gene conservation and phylogeny

It is hypothesised that plants acquired *CesA* genes through an ancient endosymbiosis event with cyanobacteria (reviewed in Popper et al. 2011). Following this ancient acquisition the genes duplicated and diversified, leading to the cellulose synthase super family which

includes “true” *CesA* genes and *CesA-like* (*Csl*) genes (Yin et al. 2009). Due to the closely related nature of the *CesA* genes they have many features in common. The CSRII region is an interesting conserved feature of the *CesA* genes because, while all plant *CesA* genes possess a CRSII region, it is more similar between orthologs of different species than between paralogs of the same species (Vergara and Carpita 2001). Liang and Joshi (2004) showed that the CSRII region can be used to separate the plant *CesA* genes into six main groups, where three where associated with primary cell wall formation and another three with SCW deposition. Using this information Ranik and Myburg (2006) isolated six *Eucalyptus CesA* genes, three of which were expressed in tissues depositing SCWs and three during primary cell wall deposition.

Over the years *CesA* genes have been isolated from many different plant species including rice, maize, *Arabidopsis*, *Populus* and *Eucalyptus* (Richmond and Somerville 2000; Hazen et al. 2002; Samuga and Joshi 2002; Appenzeller et al. 2004; Ranik and Myburg 2006) and with the increasing number of available genome sequences the number will still increase significantly. Each study which identifies a new set of *CesA* genes or adds to the list of *CesA* genes, uses a phylogenetic analysis to separate the genes and uncover their relationships (Burton et al. 2004; Djerbi et al. 2005; Suzuki et al. 2006; Carroll and Specht 2011). *CesA* phylogenetic studies have all shown that across plant species there are six clear groupings of *CesA* genes and that these are associated with the deposition of cellulose in either the primary or secondary cell wall. Despite the relatedness of the *CesA* genes, each study has applied a different naming system which is often based on the order in which each gene was identified rather than on their phylogenetic relationships. This has led to some confusion, and in some cases the same genes in a species have received multiple names. Examples of this are seen in *Populus* and *Eucalyptus*, where in *Populus PtrCesA3A*, *PtCesA1* and *PtCesA4* all represent the same *Populus trichocarpa CesA* gene (Joshi et al. 2004; Djerbi et al. 2005; Suzuki et al.

2006). Similarly, in *Eucalyptus* the names *EgCesA1* (Ranik and Myburg 2006) and *EgraCesA3* (Lu et al. 2008) represent the same *Eucalyptus grandis CesA* gene. Kumar et al. (2009) later proposed a naming convention which brought the *Populus CesA* names in line with the naming convention first used by the *Arabidopsis* community. In this review six classes of *CesA* genes for six different plant species including two monocots (rice and maize), one herbaceous dicot (*Arabidopsis*), two woody genera (*Poplulus* and *Eucalyptus*) and a basal rosid (*Vitis*) have been grouped by comparing the previously published phylogenetic groupings for these genes (Table 1.1).

With the recent completion of the *E. grandis* genome sequence (http://phytozome.net/) and the already confusing nomenclature of the *E. grandis CesA* genes, it may be prudent to adopt a similar nomenclature to that proposed by Kumar et al. (2009). To this end a last row was included in Table 1.1 where the currently published *E. grandis CesA* sequences (Ranik and Myburg 2006; Lu et al. 2008) have been renamed to reflect their homology to the *Populus* and *Arabidopsis* orthologs. For example, *EgCesA1* should be renamed *EgrCesA8*, to reflect that it is most similar to the *AtCesA8* and *PtiCesA8A* and *PtiCesA8B* genes (Table 1.1; grey shading). The *Populus* genome has undergone whole-genome duplication and so many of the genes, including the *CesA* genes, have been duplicated and some have subsequently been retained (Tuskan et al. 2006). To denote this Kumar et al. (2009) suggested the convention of adding a letter suffix (A-Z) to denote putatively duplicated genes (paralogs) which could not be distinguished from each other. A preliminary investigation of the *E. grandis CesA* genes (data not shown) has indicated there are eleven *CesA* coding genes, and that there are several paralogs which may be indistinguishable from one another. This is also denoted with a lettered suffix for *Eucalyptus* where A indicates that this is the first gene to be reported, but that there are likely other paralogs still to be investigated (Table 1.1; grey shading). While proposing a new naming

convention for the *Eucalyptus CesA* genes based on previous studies, a full analysis of the *Eucalyptus CesA* superfamily will be useful for further research in this important plantation tree species.

Table 1.1 *Cellulose synthase* (*CesA*) orthologs from different plant species always group into six clades representing the three different *CesAs* associated with the primary or secondary cell wall genes.

Plant species	**References**[a]	**Primary cell wall CSC-associated *CesA* genes**[b]			**Secondary cell wall CSC-associated *CesA* genes**[c]		
		CesA1-type[d]	*CesA3-type*[d]	*CesA6-type*[d]	*CesA4-type*[e]	*CesA7-type*[e]	*CesA8-type*[e]
Arabidopsis thaliana	Richmond and Somerville (2000)	*AtCesA1* *AtCesA10*	*AtCesA3*	*AtCesA2* *AtCesA5* *AtCesA6* *AtCesA9*	*AtCesA4*	*AtCesA7*	*AtCesA8*
Populus trichocarpa	Djerbi et al. (2005) Kumar et al. (2009)	*PtiCesA1A* *PtiCesA1B*	*PtiCesA3A* *PtiCesA3B* *PtiCesA3C* *PtiCesA3D*	*PtiCesA6A* *PtiCesA6B* *PtiCesA6C* *PtiCesA6D* *PtiCesA6E* *PtiCesA6F*	*PtiCesA4*	*PtiCesA7A* *PtiCesA7B*	*PtiCesA8A* *PtiCesA8B*
Vitis vinifera	Carroll and Specht (2011)	*VvCesA-A* *VvCesA-B*	*VvCesA-C* *VvCesA-E* *VvCesA-F*	*VvCesA-M* *VvCesA-G* *VvCesA-H* *VvCesA-I*	*VvCesA-N*	*VvCesA-J* *VvCesA-K* *VvCesA-L*	-
Oryza sativa	Djerbi et al. (2005) Carroll and Specht (2011)	*OsCesA1*	*OsCesA8* *OsCesA2*	*OsCesA6* *OsCesA5* *OsCesA3*	*OsCesA7*	*OsCesA9*	*OsCesA4*
Zea mays	Holland et al. (2000) Carroll and Specht (2011)	*ZmCesA1* *ZmCesA2*	*ZmCesA4* *ZmCesA5* *ZmCesA3* *ZmCesA9*	*ZmCesA8* *ZmCesA7* *ZmCesA6*	*ZmCesA10*	*ZmCesA12*	*ZmCesA11*
Eucalyptus grandis	Ranik and Myburg (2006)	*EgCesA5*	*EgCesA4*	*EgCesA6*	*EgCesA2*	*EgCesA3*	*EgCesA1 (EgraCesA 3; Lu et al. 2008)*
Proposed *Eucalyptus* nomenclature[g]	-	*EgrCesA1A*	*EgrCesA3A*	*EgrCesA6A*	*EgrCesA4*	*EgrCesA7A*	*EgrCesA8*

[a]Previously published phylogenies showing the primary and secondary cell wall associated clades.

[b]Three groups of genes coding for the proteins of the primary wall-associated CSCs

[c]Three groups of genes coding for the proteins of the secondary wall-associated CSCs

[d]Three different CesA proteins form the CSC (discussed above). *AtCesA1*, *AtCesA3* and *AtCesA6/9/2/5* are associated with primary cell wall formation and all genes with similar associations and phylogenetic groupings across different species listed under the headings *CesA1*-type, *CesA3*-type and CesA6-type.

[e]Three different CesA proteins form the CSC (discussed above). *AtCesA4*, *AtCesA7* and *AtCesA8* are associated with SCW formation and all genes with similar associations and phylogenetic groupings across different species are listed under the headings *CesA4*-type, *CesA7*-type and *CesA8*-type.

[g]Following the naming convention proposed by Kumar et al. (2009) and using the phylogeny from Ranik et al. (2006) we propose that the *Eucalyptus CesA* genes be renamed to reflect the names proposed for *Populus*. Letters at the end of the newly proposed *Eucalyptus CesA* names (e.g. *EgrCesA7A*) indicate genes where the recently sequenced *Eucalyptus* genome indicates a presence of a putative paralog and this allows easy naming of other *Eucalyptus CesA* genes.

1.3. Transcriptional regulation of the cellulose synthase genes

1.3.1. The promoter regions of CesA genes

As discussed above the *CesA* genes have highly co-ordinated expression patterns where three genes are always expressed in tissues undergoing SCW formation and another set of at least three *CesA* genes are associated with primary cell wall deposition (Burton et al. 2004; Hamann et al. 2004). Aside from this broad classification, recent evidence has indicated that the *CesA* genes may each have a specific expression pattern during various developmental stages and in different tissues types (Ehlting et al. 2005; Stork et al. 2010; Mendu et al. 2011). The expression of the *CesA* genes is not only modulated on a developmental or tissue-specific level but has also been shown to respond to various biotic and abiotic stresses including mechanical stress, pathogen response and drought response (Bhandari et al. 2006; Hernández-Blanco et al. 2007; Cal et al. 2013). These findings suggest that the transcriptional regulation of the *CesA* genes is complex and may incorporate overlapping regulatory networks which will have to be teased apart in order to fully understand the regulation of the gene family.

Considering the complexity of *CesA* gene expression and the wealth of information on the *CesA* protein structure and function there is comparatively little information on the

transcriptional regulation of the this gene family. Promoters are stretches of DNA which are located upstream of the coding region of a gene to which the regulatory proteins bind and initiate transcription (Pierce 2010). Investigating the regulation conferred by a promoter region of a gene is generally the starting point for most investigations on transcriptional regulation and several studies have isolated *CesA* promoters from plants such as cotton, *Populus* and *Eucalyptus* (Creux et al. 2008; Wu et al. 2009; Ko et al. 2012). These promoters have been used to monitor reporter gene expression in transgenic plant models including tobacco and *Arabidopsis* showing that the expression patterns modulated by the transgenic *CesA* promoters are generally well conserved across different plant species (Creux et al. 2008; Lu et al. 2008; Wu et al. 2009). In *Eucalyptus*, Lu et al. (2008) performed a promoter deletion study where different regions of the *EgraCesA3* (*EgrCesA8*; Table 1.1) promoter were fused to the GUS (*β*-glucuronidase) reporter gene and the expression pattern was investigated in tobacco under normal growth conditions and during tension wood formation. The authors identified a number of regions in the promoter which activated, repressed or responded to mechanical stress. A similar study on cotton identified regions in the *GhCesA4* promoter that responded to auxin and sucrose levels (Wu et al. 2009). These studies have begun to dissect the *CesA* promoter regions and have identified regions of the promoter that modulate *CesA* expression. However, these initial studies have only identified large sections of the promoter (≥ 300 bp) which are important for *CesA* expression and cannot assign reporter gene expression patterns to specific sequence motifs within the promoters.

1.3.2. Core promoter elements

In general a eukaryotic promoter can be divided into two main parts. The first is the proximal promoter which is located closest to the gene and harbours the core promoter elements to

which the basal transcriptional machinery binds (reviewed in Lenhard et al. 2012). The second is, known as the distal promoter and contains *cis*-regulatory elements such as enhancers, silencers, repressors and tethering elements which contribute to the complex spatio-temporal expression of the genes they regulate (reviewed in Spitz and Furlong 2012). Most of the eukaryotic *cis*-regulatory elements have been identified in model organisms such as yeast, *Drosophila* and humans (reviewed in Boone et al. 2007; Peter and Davidson 2011; Wittkopp and Kalay 2011) and there is far less information on regulatory elements in plants. Several *in silico* plant studies have identified many putative plant *cis*-elements from various sets of co-expressed genes; however, most of these elements have yet to be experimentally validated (Raes et al. 2003; Obayashi et al. 2007; Vandepoele et al. 2009).

All components of the promoter including the transcriptional start site (TSS) and basal transcriptional elements can play an integral role in the spatio-temporal regulation of the gene (reviewed in Goodrich and Tjian 2010). Recent studies have found that the presence of a second TSS in a number of *Arabidopsis* and rice genes is key in determining the tissue specificity of the gene (Tanaka et al. 2009). The TSS further upstream in these plant promoters was closest to the canonical RNA polymerase recruitment elements, such as the TATA-box, while the second TSS was found further downstream and had very different sequence architecture, in terms of GC content and DNA stability. This would suggest that only the upstream TSS is functional, but in full-length cDNA sequencing studies alternative transcripts in different tissues of rice and *Arabidopsis* were found to be initiated from secondary TSSs (Tanaka et al. 2009). In well studied organisms such as yeast and *Drosophila* it has been shown that promoters with one TSS tend to be specifically regulated, while promoters with multiple TSSs are more often constitutively expressed (Smale and Kadonaga 2003; Carninci et al. 2006; Juven-Gershon and Kadonaga 2010), suggesting that the TSS itself can play an important role in the tissue-specificity of a particular gene. Currently,

except for an initial TSS prediction for the *Eucalyptus CesA* promoters (Creux et al. 2008), there is no information on how/if the TSS is involved in modulating the tissue specificity of the *CesA* promoters.

The TATA-box (Table 1.1) is the sequence at which the transcriptional pre-initiation complex is assembled and can recruit RNA polymerase for the initiation of transcription (Tora 2002). It has been shown that inducible-gene promoters contain TATA-boxes and often include a number of nucleosomes, while constitutive gene promoters tend to be TATA-less and have large nucleosome free regions (Huisinga and Pugh 2004; Weiner et al. 2010). These two main types of promoters (TATA-box and TATA-less promoters) have also been noted in *Arabidopsis* genome studies (Molina and Grotewold 2005; Yamamoto et al. 2007; Zuo and Li 2011). The proportions of these two promoter types were similar to those found in *Drosophila*, where about 30% of the genes in the genome had TATA-box containing promoters (Molina and Grotewold 2005). *Arabidopsis* genes with TATA-less promoters also appeared to have longer (on average 30 bp longer) 5'UTRs than the promoters with TATA-boxes and this could aid in recruiting the pre-initiation complex (PIC) and initiating transcription (Molina and Grotewold 2005). The *CesA* promoters that have been isolated to date have not conformed to one specific promoter type. Creux et al. (2008) found that the presence or absence of a TATA-box sequence in the *CesA* promoters of *Arabidopsis*, *Eucalyptus* and *Populus* was not conserved within gene sets, expression sets or species. For example, *EgCesA1* (*EgrCesA8*) and *EgCesA3* (*EgrCesA7*) promoters did not have a TATA-box but their orthologous sequences in *Arabidopsis* and *Populus* did contain putative TATA-box motifs. Wu et al. (2009) also identified a TATA-box in the SCW-associated *GhCesA4* promoter from cotton, suggesting that there is no clear rule for the presence of a TATA-box in SCW-associated *CesA* promoters and may be a gene and/or species-specific feature in the *CesA* gene family.

An alternative to the TATA-box, known as the initiator element (Inr) has been identified at the TSS of several TATA and TATA-less promoters from yeast and plant species (Smale 1997; Molina and Grotewold 2005; Yamamoto et al. 2007; Kadonaga 2012). Inr elements are one of the most common elements in eukaryotic core promoters and have a highly variable consensus sequence (YRmotif [R, +1]; Yamamoto et al. 2007). An important feature of these elements is the CT-rich region where transcription is most often initiated. These elements recruit TFIID, which in turn recruits the rest of the PIC (Pre-initiation complex) for the initiation of transcription (reviewed in Goodrich and Tjian 2010). The R (A/G) site in the initiator sequence is usually designated as +1, because it interacts with a number of upstream and downstream elements in a position dependent way and is very close to or may be the exact point of transcriptional initiation (reviewed in Juven-Gershon and Kadonaga 2010). Some of the *Eucalyptus CesA* promoters appear to show different core promoter elements, which are not conserved across homologous promoters of different species. The *EgrCesA8* (Table 1.1) promoter was shown to have an initiator element at the predicted TSS (Creux et al. 2008); suggesting this could be a mechanism for transcriptional initiation for at least one of the *Eucalyptus CesA* genes. In the study by Creux et al. (2008), the core elements were identified by manual inspection and further *in silico* or *in vivo* studies would be required to confirm these core promoter elements.

Bernard et al. (2010) identified a set of TC-motifs in the *Arabidopsis* and rice genomes which they proposed were alternative plant-specific initiator elements. These elements were identified 30 bp upstream of the TSS and could play a similar role to the TATA-box in these promoters by recruiting the PIC. The authors also noted that the short (6 bp) TC-motifs often co-occurred with longer CT-microsatellites and these together could regulate plant gene expression. A genome-wide study of different *Brassica* species revealed an over representation of CT- and GA repeats in the 5'UTR and core promoter regions of

many genes (Zhang et al. 2006), suggesting these elements may play a role in plant gene regulation. Although these repeat elements and novel core promoter elements have yet to be positively identified and characterized in the *CesA* promoters, a recent study on simple sequence repeats identified several microsatellites in the *Populus CesA* promoters that were associated with holocellulose or *α*-cellulose content (Du et al. 2013). These findings suggest that simple sequence repeats such as microsatellites in *CesA* promoters may play a role in the transcriptional regulation of cellulose synthase expression and should be further investigated.

1.3.3. Cis-regulatory elements within the CesA promoters

Aside from the core elements, promoters also harbour *cis*-regulatory elements which modulate the tissue-/stage-/response-specificity of the gene by interacting with various transcription factors in a sequence-specific manner (reviewed in Spitz and Furlong 2012). *Cis*-element detection studies generally take one of two approaches, either using *in silico* detection in promoter datasets or more labour-intensive (but accurate) *in vivo* or *in vitro* techniques such as electrophoretic mobility shift assay (EMASs), Yeast-1-hybrid and ChIP-Seq (reviewed in Stormo and Zhao 2010). Most *in silico* studies of the *CesA* promoters have made use of the PLACE database (http://www.dna.affrc.go.jp/PLACE/) which houses all previously reported plant *cis*-elements (Higo et al. 1999). Using this tool authors could identify putative elements which might explain observed promoter::GUS expression patterns. In an example of this, Lu et al. (2008) identified a mechanical stress-responsive element (MSRE) in the *EgrCesA8* promoter using reporter gene analysis. This MSRE contained several stress responsive elements (W-box, ARR1, EEC1, CCA1 and MYC) listed on the PLACE database and the authors speculated that some of these may be responsible for the regulation of tension wood formation in *Eucalyptus* (Lu et al. 2008). Ko et al. (2006) used

expression data to identify a core set of 52 genes putatively involved in secondary cell wall formation and in the promoters of these genes they identified an over-represented *cis*-element (ACAAAGAA) which they associated with SCW formation. The element was identified in at least one of the three SCW-associated *CesA* promoters (*AtCesA8*). A phylogenetic footprinting approach based on the promoters of three primary-wall associated and three SCW-associated *CesA* genes from *Arabidopsis*, *Populus* and *Eucalyptus* first searched within these data sets for short conserved sequences. These sequences were then used to search the PLACE database to classify the putative *cis*-elements (Creux et al. 2008). The study identified *cis*-elements which were previously reported to be involved in general responses to light (CRPE16), hormones (CRPE4, CRPE5 and CRPE7) and general stem development (CRPE10 and CRPE11) in a group of primary wall-associated *CesA* promoters. The same study identified mainly novel motifs which were over-represented in the promoters of *CesA* genes associated with SCW deposition highlighting a need for further analysis of these promoters to identify SCW-related *cis*-elements (Creux et al. 2008). In a similar approach, Ding et al. (2012) used the *cis*-elements listed on the PLACE database to perform a genome-wide scan in *Arabidopsis* and *Populus* and identified clusters of *cis*-elements that were conserved in the promoters of cellulose biosynthetic and other SCW-specific genes.

The PLACE database has been a reliable, freely available resource used in many different *cis*-element analyses including those of the *CesA* promoters (Haberer et al. 2006; Creux et al. 2008; Lu et al. 2008; Wang et al. 2009a; Wu et al. 2009; Wu et al. 2010; Ding et al. 2012; Mehrotra et al. 2013). Unfortunately the curator of the database has retired and the database has not been maintained since 2007 (http://www.dna.affrc.go.jp/PLACE/) and thus is missing many newly described *cis*-elements. This has created a large gap in the current tools available for plant researchers to quickly assess the types of *cis*-elements in a given sequence. TRANSFAC (http://www.gene-regulation.com) is the closest alternative database

but excludes many of the *cis*-elements listed on the PLACE database because it only lists experimentally verified elements and does not include elements which have been identified by *in silico* methods (Wingender et al. 2000).

The *cis*-elements discussed above, while prevalent in the *CesA* promoters, have little or no supporting experimental data and none of them have been linked to a specific transcription factor or implicated in specific DNA-protein interactions. GUS reporter gene analysis was used to identify the TERE (tracheary-element-regulating) *cis*-element (CTTNAAAGCNA) in the *Zinnia* cystein protease 4 (ZCP4) promoter (Pyo et al. 2007). The ZCP4 gene is co-regulated with the SCW-associated *CesA* genes during *Zinnia* protoplast transdifferentiation to form tracheary elements. To validate the TERE element, a series of mutated elements were synthesised and only the exact *cis*-element sequence could express the reporter gene in the tissue-specific manner observed for the native TERE element. Pyo et al. (2007) identified the TERE element in the promoters of *AtCesA4* and *AtCesA7* suggesting it plays a role in the SCW-associated transcriptional regulation of these genes. Another validated *cis*-element which is involved in SCW-associated gene regulation is SNBE (Secondary NAC binding element; WNNYBTNNNNNNNAMGNHW) which was shown to be bound by the SCW NAC master regulators such as VND7 a NAC family transcription factor (Zhong et al. 2010c). While the authors validated this element and identified a number of putative targets all containing the SNBE motif, the SCW-related *CesA* genes were not among them. This is a confusing finding as Yamaguchi et al. (2011) showed by EMSA and global transcriptome analysis that *At*VND7 binds to *AtCesA8* and *AtCesA4* promoters. However, they claimed that the binding action is through the TERE element. These results suggest that even for well characterized *cis*-elements there is still much to be learned about the DNA-protein interactions in the *CesA* promoters.

To date, only the AC-elements/SMRE/M46RE/MBSIIG/ACTYPE *cis*-element has consistently and conclusively been linked with transcription factors binding to the SCW-associated *CesA* promoters. This element has several names as it has been identified by a number of different studies on different SCW-associated promoters in various plant species including *Eucalyptus*, *Arabidopsis* and *Populus* (Hatton et al. 1995; Patzlaff et al. 2003; Goicoechea et al. 2005; Shen et al. 2009; Rahantamalala et al. 2010; Winzell et al. 2010; Zhong et al. 2011; Kim et al. 2013). Initially, the element was characterized as a lignin-specific element co-ordinating the genes of the lignin biosynthetic pathway in *Eucalyptus* (Patzlaff et al. 2003; Goicoechea et al. 2005; Rahantamalala et al. 2010). More recently, a similar element was identified in the promoter regions of the SCW-associated *CesA* genes of *Arabidopsis* and *Populus* (Winzell et al. 2010; Zhong et al. 2011; Kim et al. 2013). Kim et al. (2013) showed direct binding of a MYB-family transcription factor (*At*MYB46) to this element in the *AtCesA4*, *7* and *8* promoters using yeast-1-hybrid and EMSA analysis. However, this element is not a cellulose-specific element as it has been shown to directly interact with other MYB transcription factors known to regulate other pathways in the SCW-regulatory network including lignin and xylan (Goicoechea et al. 2005; Winzell et al. 2010). These results suggest that while the binding sites and some of the interacting proteins are well characterized, the mechanisms determining expression specificity to genes involved in SCW biosynthetic pathways, regulated by MYB proteins, remains elusive.

1.3.4. Transcription factors modulating cellulose biosynthesis

Before the recent advances in next-generation DNA sequencing, understanding of the mechanisms involved in cellulose biosynthesis and SCW formation were greatly aided by genome-wide transcriptome profiling techniques such as microarray and cDNA-AFLP

analysis. Over the years extensive microarray studies have been performed on the stems of woody and herbaceous plant species and these early studies provided the first clues to which transcription factors may be regulating cell wall deposition (Hertzberg et al. 2001; Demura et al. 2002; Ko et al. 2004; Schrader et al. 2004; Brown et al. 2005; Persson et al. 2005; Zhao et al. 2005; Ko et al. 2006; Friedmann et al. 2007; Pavy et al. 2008; Dharmawardhana et al. 2010; Mizrachi et al. 2010). Many of the genes currently identified as key SCW regulators including *AtNST1*, *AtNST2* (NAC secondary wall thickening promoting factor; NST), *AtSND1*, *AtMYB46* and *AtKNAT7* (Knotted-like; KNAT) where first listed in these early studies (Ko et al. 2004; Ehlting et al. 2005; Persson et al. 2005; Zhao et al. 2005; Andersson-Gunneras et al. 2006; Ko et al. 2006; Pavy et al. 2008). Many of the genes listed in these studies still remain unknown or poorly characterized and may prove to be important members of the SCW regulatory network.

In an effort to narrow down the extensive gene lists generated by microarray studies, researchers began to perform meta-analyses on different expression datasets expected to share core genes. Persson et al. (2005) and Brown et al. (2005) chose to narrow down the gene expression list and identified genes that were directly involved with cellulose biosynthesis by selecting genes that were co-expressed with the primary or secondary-wall associated *CesA* genes. Ko et al. (2006) used data from 45 different developmental stages in *Arabidopsis* and identified 52 core genes involved in xylem formation. This list contains several members of the SCW-regulatory network including *AtSND1*, *AtNST1*, *AtNST2*, *AtMYB85*, *AtMYB103* and *AtMYB52*. In this manner they identified many genes with expression patterns similar to those of the *CesA* genes and these also contained SCW transcription factors such as *AtKNAT7*, *AtSND2* and *AtMYB43* (Brown et al. 2005; Persson et al. 2005). These co-expression lists also identified transcription factors from the bZIP and

zinc finger transcription factors which may still prove to be important regulators of the *CesA* genes (Persson et al. 2005).

After the initial identification of putative SCW regulators from the early microarray experiments, *Arabidopsis* knock-out/over-expression mutants further aided in characterizing the role of these regulators (Kubo et al. 2005; Ohashi-Ito et al. 2005; Zhong et al. 2006; Ko et al. 2007; Mitsuda et al. 2007; Mitsuda and Ohme-Takagi 2008; Ko et al. 2009). These studies revealed that *At*NST1 and *At*SND1 were functionally redundant and a double knock-out mutant (*snd1*/*nst1*) had severe secondary cell wall defects suggesting a role as a master regulator (Zhong et al. 2006; Mitsuda and Ohme-Takagi 2008). Overexpression of *At*SND2 and *At*MYB103 resulted in increased secondary cell wall thickness in fiber cells indicating a tissue-specific role for these transcription factors (Zhong et al. 2008). While highly valuable for unravelling the SCW-regulatory network, *Arabidopsis* mutant studies alone could not indicate which genes were direct targets of the transcription factors being investigated. Ko et al. (2007) performed a microarray analysis of an *Arabidopsis* *At*SND1-overexpressor and could identify a number of target genes for this transcription factor. Hussey et al. (2011) later over-expressed *At*SND2 in *Arabidopsis* and did a comparative analysis of the genes over-expressed in both the 35S::*At*SND1 (Ko et al. 2007) and 35S::*At*SND2 *Arabidopsis* plants. In this way it was possible to add to the network and the authors proposed that *At*SND2 is likely a direct target of *At*SND1 and that *AtCesA8* is possibly a direct target of *At*SND2.

*At*KNAT7, which was first identified by early microarray studies (Ehlting et al. 2005; Persson et al. 2005; Andersson-Gunneras et al. 2006), is one of the few SCW regulators to be described as a negative regulator in the network (Li et al. 2011; Li et al. 2012a). *At*KNAT7 has been shown to negatively regulate SCW formation in fiber cells and positively regulate vessel wall formation. The opposite regulatory effects of this gene on fibers and vessels leads to a thickening of fiber cell walls but a typical *irx* phenotype in the vessels of the knock-out

mutants (Li et al. 2012a). *At*KNAT7 is also one of the only (the others being *At*VND7, *At*VND6 and *At*VNI2) SCW transcription factors known to form hetero- and homo-dimers and presumably operates in larger regulatory complexes with proteins such as *At*OFP4 and *At*MYB75 (Hackbusch et al. 2005; Li et al. 2011; Bhargava et al. 2013). Bhargava et al. (2013) suggested that the *At*MYB75-*At*KNAT7 complex may play a role in regulating the expression of *AtCesA5* during seed coat development and this is the first protein complex proposed to be involved in *CesA* gene regulation.

The development of a transactivation assay in *Arabidopsis* suspension cultures and a hormone inducible system (Aoyama and Chua 1997), have greatly enhanced our ability to map transcription factor-promoter interactions in the SCW-regulatory network (Zhong et al. 2007). Using these and other techniques, Ko et al. (2009) showed that *At*MYB46 could activate the expression of the SCW-associated *CesA* promoters. Independent studies later demonstrated, with EMSA and yeast-1-hybrid assays, that this interaction was due to *At*MYB46 directly binding the promoters of the SCW *CesAs* (Zhong and Ye 2012; Kim et al. 2013). Similarly, *At*VND7 has also been shown to directly bind to and activate the SCW-associated *CesA* promoters in *Arabidopsis* (Yamaguchi et al. 2011). The *At*BES1 transcription factor which is involved in response to plant hormones such as brassinosteroids has also been shown to directly interact with primary-wall associated and SCW-associated *CesA* genes in *Arabidopsis*. However, its position in the network and relationship to other genes within the network is not well understood (Xie et al. 2011).

While the transcription factors regulating SCW formation have been extensively studied in *Arabidopsis*, there have been fewer studies of the transcription factors regulating SCW genes in woody species such as *Populus* and *Eucalyptus* (Goicoechea et al. 2005; Legay et al. 2007; Legay et al. 2010; Winzell et al. 2010; Zhong et al. 2010b; Zhong et al. 2011). The master regulator of the SCW-regulatory network and its orthologs (*Ptr*WND1A,

*Ptr*WND1B, *Ptr*WND2A, *Ptr*WND2B, *Ptr*WND6A and *Ptr*WND6B) was identified in *Populus* and has been shown to activate several other *Populus* transcription factors (Zhong et al. 2010b). A transactivation experiment on the a number of *Populus* transcription factors (*Ptr*NAC157, 156, 150, *Ptr*MYB128, 18, 3, 2, 20, 21, 74, 75, 18, *Ptr*GATA8 and *Ptr*ZN1) revealed that they activated *Populus* cell wall structural genes including the SCW-associated *CesA* genes (Zhong et al. 2011). The *Populus* xylan biosynthetic genes GT43A, GT43B and Xyn10A were confirmed as direct targets of *Pt*MYB021. The AC-like *cis*-element (ACTYPE) was proposed as the sequence facilitating this interaction (Winzell et al. 2010). While these studies have begun to outline the *Populus* SCW regulatory network there is still a great deal still to be done before it can be properly compared to the extensive *Arabidopsis* network.

1.3.5. The SCW-associated CesA regulatory network

In order to review what is known about the transcription factors that regulate the SCW-associated *CesA* promoters in *Arabidopsis* and *Populus*, a regulatory model of all the transcription factors which directly interact or transactivate the SCW *CesA* promoters was constructed from the literature (Figure 1.2). Conservation of the key transcription factors (*At*VND6, *At*VND7, *At*MYB46, *At*KNAT7 and *At*SND2) which regulate the SCW-associated *CesA* gene expression in *Arabidopsis* and *Populus* is evident by the large number of homologs that have been identified in the two genera (Table 1.2). However, in some cases there are duplicate genes (e.g. *At*MYB46 and *Populus* homologs *Ptr*MYB2, *Ptr*MYB3, *Ptr*MYB20 and *Ptr*MYB21) in the *Populus* network which may be undergoing sub-functionalization in specific tissues or cell types and their specific functions are yet to be determined (Table 1.2). A dominant feature of both the *Arabidopsis* and *Populus CesA*

regulatory networks is a number of feed forward loops where a master regulator can directly activate the *CesA* promoter and also activates an intermediate transcription factor which in turn activates the *CesA* promoter. An *Arabidopsis* example of a feed forward loop is the transactivation of the *CesA* promoters by *At*VND7 (Zhong et al. 2008), and *At*VND7 can also activate *At*MYB46 (Yamaguchi et al. 2011) which in turn activates the *CesA* promoters (Figure 1.2). A similar feed forward loop is observed in the *Populus* activation network where *Ptr*WND6A can activate the *CesA* promoter directly and can activate *Ptr*NAC157 (Zhong et al. 2011) which activates *CesA* expression (Figure 1.2, brown). Further evidence of conservation of the mechanisms regulating the *CesA* genes in *Arabidopsis* and *Populus* is found in the M46RE/SMRE binding site identified for *At*MYB46. A similar AC-type binding site was identified for the orthologous transcription factors in *Populus* (Winzell et al. 2010, Zhong and Ye 2012, Kim et al. 2012). These findings suggest that a great deal of the SCW-associated *CesA* activation network is conserved across great evolutionary distances and is expected to be similar in other commercially important plant genera such as *Eucalyptus*.

Table 1.2 *Arabidopsis* transcription factors identified as interactors or activators of the SCW-associated *CesA* promoters and their *Populus* homologs

[b]*Arabidopsis* transcription factor	*Arabidopsis* identifier	[a,b]*Populus* transcription factor	*Populus* accession number
*At*VND6	At5g62380	*Ptr*WND3A	XM_002322362
		*Ptr*WND3B	XM_002318252
		*Ptr*WND4A	XM_002329829
		*Ptr*WND4B	XM_002304392
*At*VND7	At1g71930	*Ptr*WND6A	HQ215857
		*Ptr*WND6B	HQ215858
*At*MYB46	At5g12870	*Ptr*MYB2	XM_002299875
		*Ptr*MYB3	XM_002299908
		*Ptr*MYB20	XM_002313267
		*Ptr*MYB21	XM_002313298
*At*BES1	At1g19350	-[c]	-
*At*C3H14	At1g66810	-[c]	-
*At*MYB52	At1g17950	*Ptr*MYB90	XM_002329715
		*Ptr*MYB161	XM_002309735
*At*MYB54	At1g73410	*Ptr*MYB167	XM_002318288
		*Ptr*MYB175	XM_002327991
*At*KNAT7	At1g62990	*Ptr*KNAT7	XM_002299533
*At*SND2	At4g28500	*Ptr*NAC154	XM_002327995

		*Ptr*NAC156	XM_002309731
*At*SND3	At1g28470	*Ptr*NAC105 *Ptr*NAC157	XM_002329769 XM_002305738
*At*MYB103	At1g63910	*Ptr*MYB10 *Ptr*MYB128	XM_002297978 XM_002304481
AtNAC75	At4g29230	*PtrNAC150*	XM_002330488
AtC2H2	At1g26610	*PtrZF1*	XM_002324065
AtGATA8	At3g54810	*PtrGATA8*	XM_002316335
AtMYB43 *AtMYB20*	At5g16600 At1g66230	*PtrMYB18*	XM_002305179
AtMYB85	At4g22680	*PtrMYB75*	XM_002321802
AtMYB50 *AtMYB86*	At1g57560 At5g26660	*PtrMYB74*	XM_002321591
AtMYB55	At4g01680	*PtrMYB121*	XM_002302666
AtNST1	At2g46770	*PtrWND2A* *PtrWND2B*	HQ215849 HQ215858

[a]In many cases the *Populus* transcription factors regulating SCW formation have at least one duplicate gene and these are grouped together in the same row as the individual *Arabidopsis* homolog.

[b]Homolog names and numbers were obtained from Zhong et al. (2010ab) and while the *Arabidopsis* transcription factors have been associated with the SCW related CesA promoters not all of the *Populus* orthologs have been. The association of the *Populus* transcription factors with SCW CesA promoters has not been described for many of the genes listed here.

[c]Currently no *Populus* orthologs have been reported for *AtBES1* and *AtC3H14*.

While much of the machinery regulating *CesA* expression appears to be conserved between *Arabidopsis* and *Populus* there also appears to be distinct differences between these two model species. In *Populus*, *Ptr*WND2B activates the *PtrCesA8*-type promoter but in *Arabidopsis* the close homolog *At*NST1 does not, suggesting that some divergence in the network has occurred (Table 1.2 and Figure 1.2). Similarly, in *Arabidopsis At*VND6 activates *AtCesA4* expression, but none of the possible *Populus* homologs activate *CesA* expression (Zhong et al. 2010b; Zhong et al. 2011). The conclusions drawn from these *CesA*-activation networks is limited because all of the *Populus* data comes from only two studies and much more research is required to confirm and further identify all the possible proteins interacting with the *Populus CesA* promoters. Even in *Arabidopsis* with its wealth of information only six interactions with the *CesA* promoters have been confirmed by EMSA or yeast-1-hybrid assays (Xie et al. 2011; Yamaguchi et al. 2011; Zhong and Ye 2012; Kim et al. 2013) and

only one *cis*-element motif has been identified in the *CesA* promoters for *AtMYB46* (Figure 1.2).

The commercially valuable plantation tree, *Eucalyptus*, is fast becoming a model woody species in its own right with the completion of the genome (http://phytozome.net/) and a great deal of RNA-seq data to support the gene models (Mizrachi et al. 2010). However there is little or no information on the SCW-regulatory network and how the *CesA* genes are regulated in *Eucalyptus*. Currently only one *Eucalyptus* master regulator (*Eg*WND1) has been isolated. This transcription factor is a close ortholog of *At*SND1, which does not interact or activate the SCW-associated *CesA* promoters in *Arabidopsis*, but limited activation or interaction studies have been performed on *Eg*WND1 and we cannot say if its function is conserved in *Eucalyptus* (Zhong et al. 2010a). There is evidence to suggest that there are some differences in the *Arabidopsis*, *Populus* and *Eucalyptus* SCW-regulatory networks. *Eg*WND1 could activate the *Populus* SCW-associated *CesA* promoters, but in *Arabidopsis* *At*SND1 does not directly activate the SCW-associated *CesA* promoters (Zhong et al. 2011). The other SCW-related transcription factors which have been identified in *Eucalyptus* have mostly had a role in regulating lignin biosynthetic genes. *Eg*MYB2 (an ortholog of *At*MYB83) was found to bind directly to the promoters of lignin biosynthetic genes at the MBSIIG/SMRE/M46RE like element. Previous studies of *At*MYB83 suggested that it acts redundantly with *At*MYB46 and as such is a master regulator which is likely to initiate many structural genes including the *CesA* genes (McCarthy et al. 2009; Zhong and Ye 2012). However, the capabilities of *Eg*MYB2 as a master regulator and *CesA* activator in *Eucalyptus* are yet to be established. *Eg*MYB1 has also been shown to bind to the promoters of lignin genes and appears to negatively regulate its targets and no link to *CesA* regulation has been established to date (Legay et al. 2007; Legay et al. 2010). While the *EgrCesA8* promoter was the focus of a promoter deletion study and broad regulatory regions were identified (Figure

1.2) no transcription factors could be assigned as interacting partners for these regions (Lu et al. 2008). Apart from these three transcription factors, no other *Eucalyptus* transcription factors have been isolated or characterized and there is no information on the conservation or variation of the SCW-regulatory network in this economically important species.

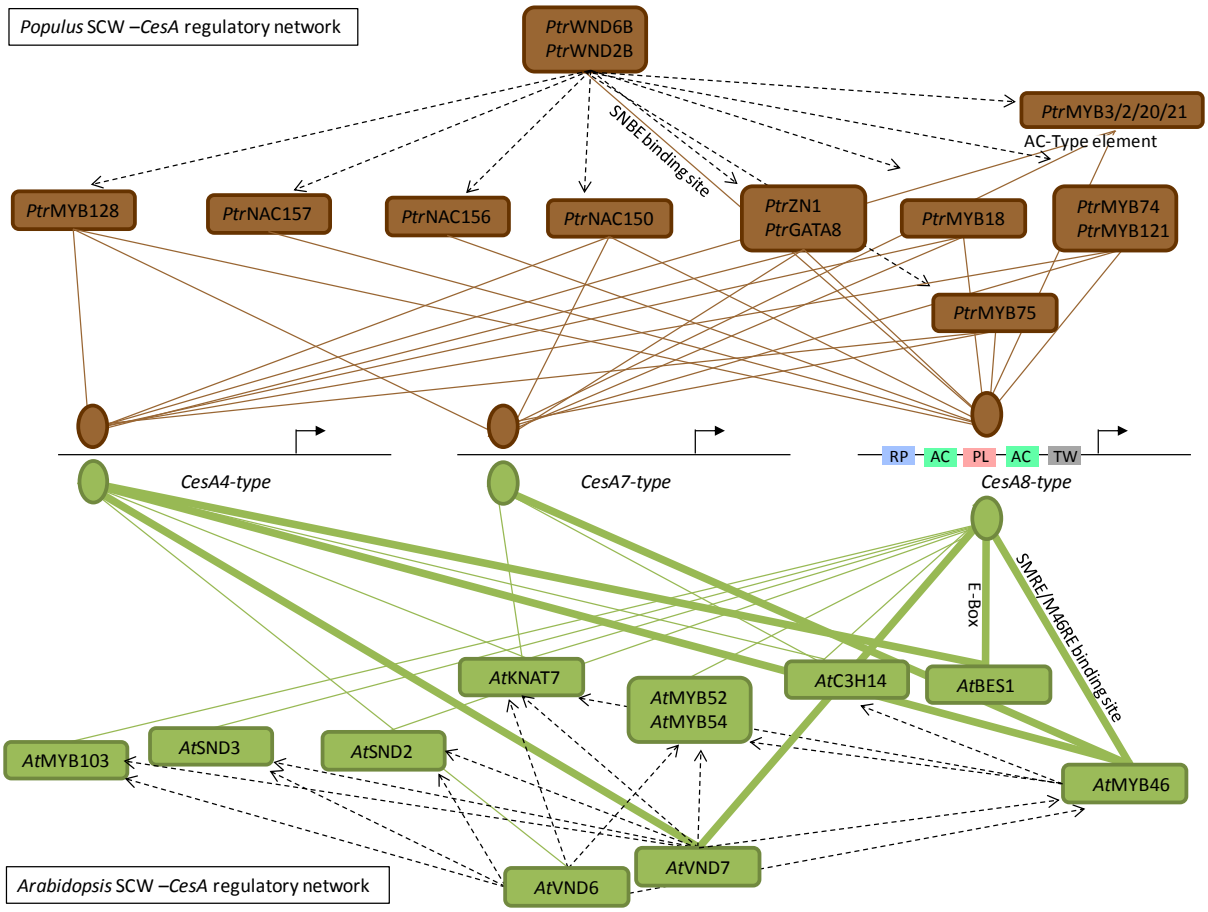


Figure 1.2 Model of SCW-associated *CesA* promoter interactions with the transcription factors of the SCW-regulatory network in woody (*Populus*) and herbaceous (*Arabidopsis*) model plants based on literature. The *CesA*-type suffix indicates the type of *CesA* gene as *Populus* has duplicated *CesA* genes which are similar to each other and the *Arabidopsis* genes (Table 1.1). Green and brown lines connecting to circles indicate transcription factors which activate the *CesA* genes in *Arabidopsis* (green) and *Populus* (brown) respectively and circles indicate little or no positional information is available for interaction. Bold lines indicate interactions confirmed by EMSA or yeast-1-hybrid. Black dashed arrows indicate which other genes in the SCW regulatory network are activated by the master regulators that also interact with the *CesA* promoters creating feed forward loops. Labels on the coloured lines indicate where a transcription factor has a known binding site which has been identified and experimentally verified. Coloured blocks on the *CesA8*-type promoter indicates the

results from the only *CesA* promoter truncation study in *Eucalyptus* where functional regions where identified (TW-tension wood response element, AC-activator element, PL-phloem repressor element, RP-repressor element).

1.3.6. Other mechanisms regulating cellulose biosynthesis

Other than the protein-DNA interactions discussed above there are several other regulatory mechanisms involved in modulating transcript and protein abundance including post-transcriptional modification and post-translational modifications (Vaucheret 2006; Spoel et al. 2010; Syed et al. 2012). There is very little information on the extent to which these alternate regulatory mechanisms affect *CesA* gene expression, although there are several studies that suggest that this is a key form of control for some *CesA* genes. Held et al. (2008) identified small interfering RNAs (siRNAs) in barley that were processed from a *HvCesA6*-antisense transcript and led to the down-regulation of *CesA* genes involved in primary cell wall formation during leaf elongation. Whether the siRNA regulation is specific to barley *CesA* genes or to primary-wall associated *CesAs* remains to be seen. Next generation sequencing technology is making great headway into understanding the level of non-coding RNA control within the genome (Wang et al. 2009b). In *Vitis*, over 100 microRNAs were identified by high throughput sequencing and were differentially expressed in different plant tissues including leaves, stems and inflorescences (Wang et al. 2012). A recent study identified several *Populus* microRNAs that appear to target xylan biosynthetic genes during xylem formation (Puzey et al. 2012). This same study identified a microRNA targeted to *PtCesA1*, but it is unclear what effect this microRNA will have on the gene's expression. In the future similar RNA sequencing studies could focus on the non-coding RNAs expressed during cellulose biosynthesis which could further elucidate the roles of non-coding RNAs in modulating *CesA* gene expression.

RNA-seq studies have also revealed that tissue-specific alternate transcript expression is highly prevalent in eukaryotic genomes (Syed et al. 2012). In apple an alternative transcript for *CesA8A* was identified which lacked the zinc finger domain. During pathogen infection, this alternate transcript is transcribed as a bicistronic pre-mRNA which is then spliced to produce the alternate *CesA8A* transcript. This arrangement of genes appears to be conserved across different species including *Populus*, *Medicago* and *Vitis*, suggesting this is a conserved pathogen response mechanism (Guerriero et al. 2012). This newly identified *CesA* regulatory mechanism should be further investigated to assess the extent to which it can affect *CesA* regulation during pathogen infection.

Post-translational modification can have an effect on how a protein functions and can affect a specific phenotype or outcome dictated by that specific protein (Spoel et al. 2010). A common form of post-translational modification is phosphorylation of proteins. An *Arabidopsis* membrane phosphor-proteomics study revealed that CesA proteins have a number of phosphorylated sites (Nühse et al. 2004). Some of these sites fall within the class-specific regions of the CesA proteins. Taylor (2007) identified phosphorylation sites in *At*CesA7 that were linked to protein degradation and the results suggested that phosphorylation and degradation of the CesAs may regulate the relative levels of cellulose synthase in the cell. When these phosphorylation sites are mutated the movement of the CSCs becomes asymmetrical and results in cell growth deformities similar to the phenotype displayed by the *rsw1* (*rootswelling1*) mutant (Chen et al. 2010). These results suggest that post-translational modifications such as phosphorylation play an important role in CesA functioning and should be kept in mind when investigating CesA regulation.

1.4. Conclusion

Plant *CesA* genes have been the focus of many studies (reviewed in Taylor 2008) and *CesA* genes have been isolated from a host of different plant species. Currently, *CesA* genes have been named according to the order in which they were isolated for each plant species. In some cases this naming convention has led to confusion with some *CesA* genes acquiring multiple names. The strong phylogenetic relationships of the *CesA* genes from different species (Carroll and Specht 2011) and the distinct expression groups identified for *CesA* genes of a specific phylogenetic clade allows for accurate identification of orthologous *CesA* genes in most species. Using this information Kumar et al. (2009) proposed a new naming convention for the *Populus CesA* genes which aligns it with the convention first reported by the *Arabidopsis* community (Richmond and Somerville 2000). This review proposes that the same naming convention be applied to the *Eucalyptus CesA* genes as these genes are also gaining multiple names as research progresses. Given the strong phylogenetic and expression conservation between the *CesA* genes of all seed plants (Yin et al. 2009) it may be useful to consider renaming all the currently identified *CesA* genes in this way as it would simplify and clarify discussions on these genes.

Eucalyptus is a valuable commercial crop producing cellulose for many industrial applications. However, *Eucalyptus* also has a number of model-plant properties such as the completed genome (http://phytozome.net/) and the availability of RNA-seq datasets (Mizrachi et al. 2010), which makes it an attractive option for cellulose regulatory studies. There is little information on the SCW-regulatory network of *Eucalyptus* and while much can be inferred from the partially conserved networks of *Arabidopsis* and *Populus* (Zhong et al. 2010a), it is important to confirm these regulatory mechanisms in *Eucalyptus* as well. Identifying the *Eucalyptus* transcription factors will also allow us to directly test and add to previous findings on the *Eucalyptus CesA* promoters. Studies performed on the *Eucalyptus*

CesA promoters and the SCW-regulatory network has led to a number of questions which need to be addressed. Firstly, are previously identified putative *cis*-elements conserved within the *CesA* promoters across different *Eucalyptus* species? Secondly, are the *Eucalyptus CesA* promoter regions, which harbour conserved *cis*-elements, functionally active *in planta*? Finally, do orthologous *Eucalyptus* transcription factors from the SCW-regulatory network bind the conserved regions of the SCW-associated *CesA* promoter? Answering these questions will give a better understanding of *Eucalyptus CesA* promoter architecture and function. Investigating the *CesA* promoter and transcription factor interaction in *Eucalyptus* may also add insight to the overall network by identifying novel interactions and confirming the conservation of previously identified interactions.

1.5. References

Andersson-Gunneras S, Mellerowicz EJ, Love J, Segerman B, Ohmiya Y, Coutinho PM, Nilsson P, Henrissat B, Moritz T, Sundberg B (2006) Biosynthesis of cellulose-enriched tension wood in *Populus*: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. Plant J. 45: 144-165

Aoyama T, Chua N-H (1997) A glucocorticoid-mediated transcriptional induction system in transgenic plants. Plant J. 11: 605-612

Appenzeller L, Doblin M, Barreiro R, Wang HY, Niu XM, Kollipara K, Carrigan L, Tomes D, Chapman M, Dhugga KS (2004) Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (*CesA*) gene family. Cellulose 11: 287-299

Atanassov II, Pittman JK, Turner SR (2009) Elucidating the mechanisms of assembly and subunit interaction of the cellulose synthase complex of *Arabidopsis* secondary cell walls. J. Biol. Chem. 284: 3833-3841

Bernard V, Brunaud V, Lecharny A (2010) TC-motifs at the TATA-box expected position in plant genes: a novel class of motifs involved in the transcription regulation. BMC Genomics 11: 166

Bhandari S, Fujino T, Thammanagowda S, Zhang D, Xu F, Joshi CP (2006) Xylem-specific and tension stress-responsive coexpression of KORRIGAN endoglucanase and three secondary wall-associated cellulose synthase genes in aspen trees. Planta 224: 828-837

Bhargava A, Ahad A, Wang S, Mansfield SD, Haughn GW, Douglas CJ, Ellis BE (2013) The interacting MYB75 and KNAT7 transcription factors modulate secondary cell wall deposition both in stems and seed coat in *Arabidopsis*. Planta 237:1199-1211

Boone C, Bussey H, Andrews BJ (2007) Exploring genetic interactions and networks with yeast. Nature Rev. Genet. 8: 437-449

Bringmann M, Li E, Sampathkumar A, Kocabek T, Hauser M-T, Persson S (2012) POM-POM2/cellulose synthase interacting1 is essential for the functional association of cellulose synthase and microtubules in *Arabidopsis*. Plant Cell 24: 163-177

Brown DM, Zeef LAH, Ellis J, Goodacre R, Turner SR (2005) Identification of novel genes in *Arabidopsis* involved in secondary cell wall formation using expression profiling and reverse genetics. Plant Cell 17: 2281-2295

Burton RA, Shirley NJ, King BJ, Harvey AJ, Fincher GB (2004) The *CesA* gene family of barley. Quantitative analysis of transcripts reveals two groups of co-expressed genes. Plant Physiol. 134: 224-236

Cal AJ, Liu D, Mauleon R, Hsing Y-IC, Serraj R (2013) Transcriptome profiling of leaf elongation zone under drought in contrasting rice cultivars. PloS ONE 8: e54537

Campbell J, Davies G, Bulone V, Henrissat B (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem. J. 326: 929-942

Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engström PG, Frith MC (2006) Genome-wide analysis of mammalian promoter architecture and evolution. Nature Genet. 38: 626-635

Carpita NC (2011) Update on mechanisms of plant cell wall biosynthesis: how plants make cellulose and other (1→ 4)-β-D-glycans. Plant Physiol. 155: 171-184

Carroll A, Specht CD (2011) Understanding plant cellulose synthases through a comprehensive investigation of the cellulose synthase family sequences. Front. Plant Sci. 2: 5

Chen S, Ehrhardt DW, Somerville CR (2010) Mutations of cellulose synthase (CESA1) phosphorylation sites modulate anisotropic cell expansion and bidirectional mobility of cellulose synthase. Proc. Natl. Acad. Sci USA 107: 17188-17193

Creux NM, Ranik M, Berger DK, Myburg AA (2008) Comparative analysis of orthologous cellulose synthase promoters from *Arabidopsis*, *Populus* and *Eucalyptus*: evidence of conserved regulatory elements in angiosperms. New Phytol. 179: 722-737

Delmer DP (1999) Cellulose biosynthesis: exciting times for a difficult field of study. Annu. Rev. Plant Phys. Biology 50: 245-276

Demura T, Tashiro G, Horiguchi G, Kishimoto N, Kubo M, Matsuoka N, Minami A, Nagata-Hiwatashi M, Nakamura K, Okamura Y, Sassa N, Suzuki S, Yazaki J, Kikuchi S, Fukuda H (2002) Visualization by comprehensive microarray analysis of gene expression programs during transdifferentiation of mesophyll cells into xylem cells. Proc. Natl. Acad. Sci. USA 99: 15794-15799

Demura T, Ye Z-H (2010) Regulation of plant biomass production. Curr. Opin. Plant Biol. 13: 298-303

Desprez T, Juraniec M, Crowell EF, Jouy H, Pochylova Z, Parcy F, Hofte H, Gonneau M, Vernhettes S (2007) Organization of cellulose synthase complexes involved in primary cell wall synthesis in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 104: 15572-15577

Dharmawardhana P, Brunner A, Strauss S (2010) Genome-wide transcriptome analysis of the transition from primary to secondary stem development in *Populus trichocarpa*. BMC Genomics 11: 150

Ding J, Hu H, Li X (2012) Thousands of *cis*-regulatory sequence combinations are shared by *Arabidopsis* and poplar. Plant Physiol. 158: 145-155

Diotallevi F, Mulder B (2007) The cellulose synthase complex: a polymerization driven supramolecular motor. Biophys. J. 92: 2666-2673

Djerbi S, Lindskog M, Arvestad L, Sterky F, Teeri TT (2005) The genome sequence of black cottonwood (*Populus trichocarpa*) reveals 18 conserved cellulose synthase (*CesA*) genes. Planta 221: 739-746

Doblin MS, Kurek I, Jacob-Wilk D, Delmer DP (2002) Cellulose biosynthesis in plants: from genes to rosettes. Plant Cell Physiol. 43: 1407-1420

Du Q, Pan W, Xu B, Li B, Zhang D (2013) Polymorphic simple sequence repeat (SSR) loci within cellulose synthase (*PtoCesA*) genes are associated with growth and wood properties in *Populus tomentosa*. New Phytol. 197: 763-776

Ehlting J, Mattheus N, Aeschliman DS, Li E, Hamberger B, Cullis IF, Zhuang J, Kaneda M, Mansfield SD, Samuels L (2005) Global transcript profiling of primary stems from

Arabidopsis thaliana identifies candidate genes for missing links in lignin biosynthesis and transcriptional regulators of fiber differentiation. Plant J. 42: 618-640

Friedmann M, Ralph SG, Aeschliman D, Zhuang J, Ritland K, Ellis BE, Bohlmann J, Douglas CJ (2007) Microarray gene expression profiling of developmental transitions in Sitka spruce (*Picea sitchensis*) apical shoots. J. Exp. Bot. 58: 593-614

Gardiner JC, Taylor NG, Turner SR (2003) Control of cellulose synthase complex localization in developing xylem. Plant Cell 15: 1740-1748

Goicoechea M, Lacombe E, Legay S, Mihaljevic S, Rech P, Jauneau A, Lapierre C, Pollet B, Verhaegen D, Chaubet-Gigot N, Grima-Pettenati J (2005) *Eg*MYB2, a new transcriptional activator from *Eucalyptus* xylem, regulates secondary cell wall formation and lignin biosynthesis. Plant J. 43: 553-567

Goodrich JA, Tjian R (2010) Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. Nature Rev. Genet. 11: 549-558

Guerriero G, Spadiut O, Kerschbamer C, Giorno F, Baric S, Ezcurra I (2012) Analysis of cellulose synthase genes from domesticated apple identifies collinear genes WDR53 and *CesA8A*: partial co-expression, bicistronic mRNA, and alternative splicing of *CESA8A*. J. Exp. Botany 63: 6045-6056

Haberer G, Mader MT, Kosarev P, Spannagl M, Yang L, Mayer KF (2006) Large-scale *cis*-element detection by analysis of correlated expression and sequence conservation between *Arabidopsis* and *Brassica oleracea*. Plant Physiol. 142: 1589-1602

Hackbusch J, Richter K, Müller J, Salamini F, Uhrig JF (2005) A central role of *Arabidopsis thaliana* ovate family proteins in networking and subcellular localization of 3-aa loop extension homeodomain proteins. Proc. Natl. Acad. Sci. USA 102: 4908-4912

Hamann T, Osborne E, Youngs HL, Misson J, Nussaume L, Somerville C (2004) Global expression analysis of *CESA* and *CSL* genes in *Arabidopsis*. Cellulose 11: 279-286

Hatton D, Sablowski R, Yung MH, Smith C, Schuch W, Bevan M (1995) Two classes of *cis* sequences contribute to tissue-specific expression of a PAL2 promoter in transgenic tobacco. Plant J. 7: 859-876

Hazen SP, Scott-Craig JS, Walton JD (2002) Cellulose synthase-like genes of rice. Plant Physiol. 128: 336-340

Held MA, Penning B, Brandt AS, Kessans SA, Yong W, Scofield SR, Carpita NC (2008) Small-interfering RNAs from natural antisense transcripts derived from a cellulose synthase gene modulate cell wall biosynthesis in barley. Proc. Natl. Acad. Sci. USA 105: 20534-20539

Hernández-Blanco C, Feng DX, Hu J, Sánchez-Vallet A, Deslandes L, Llorente F, Berrocal-Lobo M, Keller H, Barlet X, Sánchez-Rodríguez C (2007) Impairment of cellulose synthases required for *Arabidopsis* secondary cell wall formation enhances disease resistance. Plant Cell 19: 890-903

Hertzberg M, Aspeborg H, Schrader J, Andersson A, Erlandsson R, Blomqvist K, Bhalerao R, Uhlen M, Teeri TT, Lundeberg J, Sundberg B, Nilsson P, Sandberg G (2001) A transcriptional roadmap to wood formation. Proc. Natl. Acad. Sci. USA 98: 14732-14737

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

Holland N, Holland D, Helentjaris T, Dhugga KS, Xoconostle-Cazares B, Delmer DP (2000) A comparative analysis of the plant cellulose synthase (*CesA*) gene family. Plant Physiol. 123: 1313-1323

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

Huisinga KL, Pugh BF (2004) A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in *Saccharomyces cerevisiae*. Mol. Cell 13: 573-585

Hussey SG, Mizrachi E, Spokevicius AV, Bossinger G, Berger DK, Myburg AA (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 Biol. 11: 173

Joshi CP, Bandari S, Ranjan P, Kalluri UC, Liang X, Fujino T, Samuga A (2004) Genomics of cellulose biosynthesis in poplars. New Phytol. 164: 53-61

Joshi CP, Mansfield SD (2007) The cellulose paradox—simple molecule, complex biosynthesis. Curr. Opin. Plant Biol. 10: 220-226

Juven-Gershon T, Kadonaga JT (2010) Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev. Biol. 339: 225-229

Kadonaga JT (2012) Perspectives on the RNA polymerase II core promoter. Wiley Interdiscip. Rev. Dev. Biol. 1: 40-51

Kim W-C, Ko J-H, Kim J-Y, Kim J, Bae H-J, Han K-H (2013) MYB46 directly regulates the gene expression of secondary wall-associated cellulose synthases in *Arabidopsis*. Plant J. 73: 26-36

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

Ko J-H, Beers EP, Han K-H (2006) Global comparative transcriptome analysis identifies gene network regulating secondary xylem development in *Arabidopsis thaliana*. Mol. Gen. Genomics 276: 517-531

Ko J-H, Kim H-T, Hwang I, Han K-H (2012) Tissue-type-specific transcriptome analysis identifies developing xylem-specific promoters in poplar. Plant Biotechnol. J. 10: 587-596

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

Ko JH, Han KH, Park S, Yang JM (2004) Plant body weight-induced secondary growth in *Arabidopsis* and its transcription phenotype revealed by whole-transcriptome profiling. Plant Physiol. 135: 1069-1083

Ko JH, Kim WC, Han KH (2009) Ectopic expression of MYB46 identifies transcriptional regulatory genes involved in secondary wall biosynthesis in *Arabidopsis*. Plant J. 60: 649-665

Kubo M, Udagawa M, Nishikubo N, Horiguchi G, Yamaguchi M, Ito J, Mimura T, Fukuda H, Demura T (2005) Transcription switches for protoxylem and metaxylem vessel formation. Genes & Dev 19: 1855-1860

Kumar M, Thammannagowda S, Bulone V, Chiang V, Han K-H, Joshi CP, Mansfield SD, Mellerowicz E, Sundberg B, Teeri T, Ellis BE (2009) An update on the nomenclature for the cellulose synthase genes in *Populus*. Trends Plant Sci. 14: 248-254

Kurek I, Kawagoe Y, Jacob-Wilk D, Doblin M, Delmer D (2002) Dimerization of cotton fiber cellulose synthase catalytic subunits occurs via oxidation of the zinc-binding domains. Proc. Natl. Acad. Sci. USA 99: 11109-11114

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

Legay S, Sivadon P, Blervacq AS, Pavy N, Baghdady A, Tremblay L, Levasseur C, Ladouce N, Lapierre C, Séguin A, Hawkins S, Mackay J, Grima-Pettenati J (2010) *Eg*MYB1, an R2R3 MYB transcription factor from *Eucalyptus* negatively regulates secondary cell wall formation in *Arabidopsis* and poplar. New Phytol. 188: 774-786

Lei L, Li S, Gu Y (2012) Cellulose synthase complexes: composition and regulation. Front. Plant Sci. 3:75

Lenhard B, Sandelin A, Carninci P (2012) Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nature Rev. Genet. 13: 233-245

Li E, Bhargava A, Qiang W, Friedmann MC, Forneris N, Savidge RA, Johnson LA, Mansfield SD, Ellis BE, Douglas CJ (2012a) The Class II KNOX gene KNAT7 negatively regulates secondary wall formation in *Arabidopsis* and is functionally conserved in *Populus*. New Phytol. 194: 102-115

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

Li S, Lei L, Somerville CR, Gu Y (2012b) Cellulose synthase interactive protein 1 (CSI1) links microtubules and cellulose synthase complexes. Proc. Natl. Acad. Sci. USA 109: 185-190

Liang X, Joshi CP (2004) Molecular cloning of ten distinct hypervariable regions from the cellulose synthase gene superfamily in aspen trees. Tree Physiol. 24: 543-550

Lu S, Li L, Yi X, Joshi CP, Chiang VL (2008) Differential expression of three *Eucalyptus* secondary cell wall-related cellulose synthase genes in response to tension stress. J. Exp. Bot. 59: 681-695

McCarthy RL, Zhong R, Ye Z-H (2009) MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. Plant Cell Physiol. 50: 1950-1964

Mehrotra R, Sethi S, Zutshi I, Bhalothia P, Mehrotra S (2013) Patterns and evolution of ACGT repeat *cis*-element landscape across four plant genomes. BMC Genomics 14: 203

Mendu V, Griffiths JS, Persson S, Stork J, Downie AB, Voiniciuc C, Haughn GW, DeBolt S (2011) Subfunctionalization of cellulose synthases in seed coat epidermal cells mediates secondary radial wall synthesis and mucilage attachment. Plant Physiol. 157: 441-453

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

Mitsuda N, Ohme-Takagi M (2008) NAC transcription factors NST1 and NST3 regulate pod shattering in a partially redundant manner by promoting secondary wall formation after the establishment of tissue identity. Plant J. 56: 768-778

Mizrachi E, Hefer C, Ranik M, Joubert F, Myburg A (2010) *De novo* assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by Illumina mRNA-Seq. BMC Genomics 11: 681

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

Morgan JL, Strumillo J, Zimmer J (2012) Crystallographic snapshot of cellulose synthesis and membrane translocation. Nature 493: 181-186

Nühse TS, Stensballe A, Jensen ON, Peck SC (2004) Phosphoproteomics of the *Arabidopsis* plasma membrane and a new phosphorylation site database. Plant Cell 16: 2394-2405

Obayashi T, Kinoshita K, Nakai K, Shibaoka M, Hayashi S, Saeki M, Shibata D, Kazuki S, Ohta H (2006) ATTED-II: a database of co-expressed genes and *cis*-elements for identifying co-regulated gene groups in *Arabidopsis*. Nucleic Acids Res. 35:D863-D869

Ohashi-Ito K, Kubo M, Demura T, Fukuda H (2005) Class III homeodomain leucine-zipper proteins regulate xylem cell differentiation. Plant Cell Physiol. 46: 1646-1656

Ohashi-Ito K, Oda Y, 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

Paredez AR, Somerville CR, Ehrhardt DW (2006) Visualization of cellulose synthase demonstrates functional association with microtubules. Science 312: 1491-1495

Patzlaff A, Newman LJ, Dubos C, Whetten RW, Smith C, McInnis S, Bevan MW, Sederoff RR, Campbell MM (2003) Characterisation of *Pt*MYB1, an R2R3-MYB from pine xylem. Plant Mol. Biol. 53: 597-608

Pavy N, Boyle B, Nelson C, Paule C, Giguère I, Caron S, Parsons LS, Dallaire N, Bedon F, Bérubé H, Cooke J, Mackay J (2008) Identification of conserved core xylem gene sets: conifer cDNA microarray development, transcript profiling and computational analyses. New Phytol. 180: 766-786

Pear JR, Kawagoe Y, Schreckengost WE, Delmer DP, Stalker DM (1996) Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. Proc. Natl. Acad. Sci. USA 93: 12637-12642

Persson S, Pardez A, Carroll A, Palsdottir H, Doblin M, Poindexter P, Khitrov N, Auer M, Somerville CR (2007) Genetics evidence if three unique components in primary cell-wall cellulose synthase complexes in *Arabidopsis*. Proc. Natl. Acad. Sci. USA 104: 15566-15571

Persson S, Wei HR, Milne J, Page GP, Somerville CR (2005) Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. Proc. Natl. Acad. Sci. USA 102: 8633-8638

Peter IS, Davidson EH (2011) Evolution of gene regulatory networks that control embryonic development of the body plan. Cell 144: 970-985

Pierce BA (2010) Genetics: A Conceptual Approach. W. H. Freeman and Company, New York

Popper ZA, Michel G, Hervé C, Domozych DS, Willats WG, Tuohy MG, Kloareg B, Stengel DB (2011) Evolution and diversity of plant cell walls: from algae to flowering plants. Annu. Rev. Plant Biol. 62: 567-590

Puzey JR, Karger A, Axtell M, Kramer EM (2012) Deep annotation of *Populus trichocarpa* microRNAs from diverse tissue sets. PLoS ONE 7(3): e33034

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

Raes J, Rhode A, Christensen JH, Vand dePeer Y, Boerjan W (2003) Genome-wide characterization of the lignification toolbox in *Arabidopsis*. Plant Phyiol. 133: 1051-1071

Rahantamalala A, Rech P, Martinez Y, Chaubet-Gigot N, Grima-Pettenati J, Pacquit V (2010) Coordinated transcriptional regulation of two key genes in the lignin branch pathway-CAD and CCR-is mediated through MYB-binding sites. BMC Plant Biol. 10: 130

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

Richmond TA, Somerville CR (2000) The cellulose synthase superfamily. Plant Physiol. 124: 495-498

Robinson D, Quader H (1981) Structure, synthesis, and orientation of microfibrils. IX. A freeze-fracture investigation of the Oocystis plasma membrane after inhibitor treatments. Euro. J. Cell Biol. 25: 278-288

Samuga A, Joshi CP (2002) A new cellulose synthase gene (*PtrCesA2*) from aspen xylem is orthologous to *Arabidopsis AtCesA7* (irx3) gene associated with secondary cell wall synthesis. Gene 296: 37-44

Saxena IM, Brown RM, Jr., Dandekar T (2001) Structure-function characterization of cellulose synthase: relationship to other glycosyltransferases. Phytochemistry 57: 1135-1148

Saxena IM, Brown RM, Jr., Fevre M, Geremia RA, Henrissat B (1995) Multidomain architecture of β-glycosyl transferases: implications for mechanism of action. J. Bacteriol. 177: 1419-1424

Schrader J, Nilsson J, Mellerowicz E, Berglund A, Nilsson P, Hertzberg M, Sandberg G (2004) A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. Plant Cell 16: 2278-2292

Sethaphong L, Haigler CH, Kubicki JD, Zimmer J, Bonetta D, DeBolt S, Yingling YG (2013) Tertiary model of a plant cellulose synthase. Proc. Natl. Acad. Sci. 110: 7512-7517

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

Smale ST (1997) Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. BBA-Gene Struct. Expr. 1351: 73-88

Smale ST, Kadonaga JT (2003) The RNA polymerase II core promoter. Annu. Rev Biochem. 72: 449-479

Somerville C (2006) Cellulose synthesis in higher plants. Annu. Rev. Cell Dev. Biol. 22: 53-78

Song D, Shen J, Li L (2010) Characterization of cellulose synthase complexes in *Populus* xylem differentiation. New Phytol. 187: 777-790

Spitz F, Furlong EE (2012) Transcription factors: from enhancer binding to developmental control. Nature Rev. Genet. 13: 613-626

Spoel SH, Tada Y, Loake GJ (2010) Post-translational protein modification as a tool for transcription reprogramming. New Phytol. 186: 333-339

Stork J, Harris D, Griffiths J, Williams B, Beisson F, Li-Beisson Y, Mendu V, Haughn G, DeBolt S (2010) CELLULOSE SYNTHASE9 serves a nonredundant role in secondary cell wall synthesis in *Arabidopsis* epidermal testa cells. Plant physiol. 153: 580-589

Stormo GD, Zhao Y (2010) Determining the specificity of protein–DNA interactions. Nature Rev. Genet. 11: 751-760

Stracke R, Werber M, Weisshaar B (2001) The R2R3-MYB gene family in *Arabidopsis thaliana*. Curr. Opin. Plant Biol. 4: 447-456

Suzuki S, Li LG, Sun YH, Chiang VL (2006) The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthase-like genes in *Populus trichocarpa*. Plant Physiol. 142: 1233-1245

Syed NH, Kalyna M, Marquez Y, Barta A, Brown JW (2012) Alternative splicing in plants–coming of age. Trends Plant Sci. 17: 616-623

Tanaka T, Koyanagi KO, Itoh T (2009) Highly diversified molecular evolution of downstream transcription start sites in rice and *Arabidopsis*. Plant Physiol. 149: 1316-1324

Taylor NG (2007) Identification of cellulose synthase *At*CesA7 (IRX3) *in vivo* phosphorylation sites – a potential role in regulating protein degridation. Plant Mol. Biol. 64: 161-171

Taylor NG (2008) Cellulose biosynthesis and deposition in higher plants. New Phytol. 178: 239-252

Taylor NG, Howells RM, Huttly AK, Vickers K, Turner SR (2003) Interactions among three distinct CESA proteins essential for cellulose synthesis. Proc. Natl. Acad. Sci. USA 100: 1450-1455

Taylor NG, Laurie S, Turner SR (2000) Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in *Arabidopsis*. Plant Cell 12: 2529-2539

Tora L (2002) A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. Genes & Dev. 16: 673-675

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A,

Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, de Peer YV, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313: 1596-1604

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

Vaucheret H (2006) Post-transcriptional small RNA pathways in plants: mechanisms and regulations. Genes & Dev. 20: 759-771

Vergara CE, Carpita NC (2001) *β*-D-Glycan synthases and the *CesA* gene family: lessons to be learned from the mixed-linkage (1→ 3),(1→ 4) *β*-D-glucan synthase. Plant Mol. Biol. 47: 145-160

Wang C, Han J, Liu C, Kibet KN, Kayesh E, Shangguan L, Li X, Fang J (2012) Identification of microRNAs from Amur grape (*vitis amurensis* Rupr.) by deep sequencing and analysis of microRNA variations with bioinformatics. BMC Genomics 13: 122

Wang J, Elliott JE, Williamson RE (2008) Features of the primary wall CESA complex in wild type and cellulose-deficient mutants of *Arabidopsis thaliana*. J. Exp. Bot. 59: 2627-2637

Wang X, Haberer G, Mayer KF (2009a) Discovery of *cis*-elements between sorghum and rice using co-expression and evolutionary conservation. BMC Genomics 10: 284

Wang Z, Gerstein M, Snyder M (2009b) RNA-Seq: a revolutionary tool for transcriptomics. Nature Rev. Genet. 10: 57-63

Weiner A, Hughes A, Yassour M, Rando OJ, Friedman N (2010) High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. Genome Res. 20: 90-100

Wightman R, Marshall R, Turner SR (2009) A cellulose synthase-containing compartment moves rapidly beneath sites of secondary wall synthesis. Plant Cell Physiol. 50: 584-594

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

Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Prüß M, Reuter I, Schacherer F (2000) TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Res. 28: 316-319

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

Wittkopp PJ, Kalay G (2011) *Cis*-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nature Rev. Genet. 13: 59-69

Wu A-M, Hu JS, Liu J-Y (2009) Functional analysis of a cotton cellulose synthase A4 gene promoter in transgenic tobacco plants. Plant Cell Rep. 28: 1539-1548

Wu H, Zhang S, Wang Z, Su J, Du K (2010) *Cis*-acting elements analysis and function detection of xylem specific promoter from *Populus*. Frontiers Agric. China 4: 226-231

Xie L, Yang C, Wang X (2011) Brassinosteroids can regulate cellulose biosynthesis by controlling the expression of *CESA* genes in *Arabidopsis*. J. Exp Bot 62: 4495-4506

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

Yamamoto YY, Ichida H, Matsui M, Obokata J, Sakurai T, Satou M, Seki M, Shinozaki K, Abe T (2007) Identification of plant promoter constituents by analysis of local distribution of short sequences. BMC Genomics 8: 67

Yin Y, Huang J, Xu Y (2009) The cellulose synthase superfamily in fully sequenced plants and algae. BMC Plant Biol. 9: 99

Zhang L, Zuo K, Zhang F, Cao Y, Wang J, Zhang Y, Sun X, Tang K (2006) Conservation of noncoding microsatellites in plants: implication for gene regulation. BMC Genomics 7: 323

Zhao C, Craig JC, Petzold HE, Dickerman AW, Beers EP (2005) The xylem and phloem transcriptomes from secondary tissues of the *Arabidopsis* root-hypocotyl. Plant Physiol. 138: 803-818

Zhong R, Demura T, 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, Ye Z-H (2010a) Evolutionary conservation of the transcriptional network regulating secondary cell wall biosynthesis. Trends Plant Sci. 15: 625-632

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

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

Zhong R, Lee C, Zhou J, McCarthy RL, 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 RL, Lee C, Ye Z-H (2011) Dissection of the transcriptional program regulating secondary wall biosynthesis during wood formation in poplar. Plant Physiol. 157: 1452-1468

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

Zhong RQ, Richardson EA, Ye ZH (2007) Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of *Arabidopsis*. Planta 225: 1603-1611

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

Zuo Y-C, Li Q-Z (2011) Identification of TATA and TATA-less promoters in plant genomes by integrating diversity measure, GC-Skew and DNA geometric flexibility. Genomics 97: 112-120

Chapter 2

Diversity and *cis*-element architecture of the promoter regions of cellulose synthase genes in *Eucalyptus*

N.M. Creux, M.H. de Castro, M. Ranik, F.M. Maleka, A.A. Myburg

Department of Genetics, Faculty of Natural and Agricultural Sciences, Forestry Agricultural Biotechnology Institute (FABI), University of Pretoria, 2000, South Africa

This chapter is published in *Tree Genetics and Genomes*, but was reformatted to the *Planta Journal* format to match the rest of the dissertation. This chapter was published with the previously published *Eucalyptus CesA* names (Ranik and Myburg 2006), however, for continuity in the dissertation the new *Eucalyptus CesA* names proposed in Chapter 1 will be used. I performed all the *cis*-element mapping and analysis on the 13 *Eucalyptus* promoters for each of the six *CesA* genes. I also wrote the manuscript and prepared it for submission. Ms M.H. De Castro isolated and cloned the *CesA* promoters from the 12 other *Eucalyptus* species. She also performed the nucleotide diversity analysis and aided in the manuscript preparation. Mr F. Maleka provided the *E. urophylla* population data for the *CesA1* promoter and provided input during the manuscript preparation. Mr M. Ranik also contributed to hypothesis generation, analysis and preparation of this manuscript. Prof. A.A. Myburg provided advice, direction and supervision in the planning of the project. He also provided direction in the interpretation of the results and critical revision of the manuscript. All other assistance is acknowledged at the end of the chapter.

2.1 Abstract

Lignocellulosic biomass from fast-growing plantation trees is composed of carbohydrate-rich materials deposited into plant cell walls in a coordinated manner during wood formation. The diversity and evolution of the transcriptional networks regulating this process have not been studied extensively. We investigated patterns of species-level nucleotide diversity in the promoters of cellulose synthase (*CesA*) genes from different *Eucalyptus* tree species and assessed the possible roles of DNA sequence polymorphism in the gain or loss of *cis*-elements harboured within the promoters. Promoter regions of three primary and three secondary cell wall-associated *CesA* genes were isolated from 13 *Eucalyptus* species and were analysed for nucleotide and *cis*-element diversity. Species-level nucleotide diversity (π) ranged from 0.014 to 0.068 and different *CesA* promoters exhibited distinct patterns of sequence conservation. A set of 22 putative *cis*-elements were mapped to the *CesA* promoters using *in silico* methods. Forty two percent of the mapped *cis*-element occurrences contained singleton polymorphisms which resulted in either the gain or loss of a *cis*-element in a particular *Eucalyptus* species. The promoters of *Eucalyptus CesA* genes contained regions that are highly conserved at the species (*Eucalyptus*) and genus (with *Arabidopsis* and *Populus*) level, suggesting the presence of regulatory modules imposing functional constraint on such regions. Nucleotide polymorphisms in the *CesA* promoters more frequently created new *cis*-element occurrences than disrupted existing *cis*-element occurrences when compared to the consensus sequence, a process which may be important for the maintenance and evolution of cellulose gene regulation in plants.

Keywords: *Cis*-element conservation, promoter evolution, secondary cell wall, wood formation, *CesA*, woody biomass

2.2 Introduction

With the current worldwide focus on renewable energy production and carbon sequestration (Ragauskas et al. 2006; Piao et al. 2009), lignocellulosic biomass from fast-growing plantation trees is being targeted as a renewable source of carbon for biofuels and biomaterials (Regalbuto 2009; Rathmann et al. 2010). The bulk of this biomass is comprised of cellulose, hemi-cellulose and lignin contained in the secondary cell walls of wood fibre cells (Gorshkova et al. 2012). The biochemical and structural complexity of wood is determined largely by the coordinated expression of hundreds of regulatory, structural and biosynthetic genes (Aspeborg et al. 2005; Mellerowicz and Sundberg 2008). Several transcription factor genes have already been identified as key regulatory components of these pathways in the model plant *Arabidopsis* (Zhong et al. 2008; Zhong et al. 2010) and in woody plants such as *Populus and Eucalyptus* (Hu et al. 2010; Legay et al. 2010; McCarthy et al. 2010). Despite the emerging understanding of the molecular machinery underlying major biosynthetic pathways in wood-forming tissues, the nature and evolution of transcriptional networks regulating these pathways are not well described in woody plants such as *Eucalyptus*.

Eucalyptus is a richly diverse genus (Brooker 2000), containing over 700 species of woody plants, many of which are rapid and prolific producers of cellulose-rich biomass. Some of these species and their fast-growing hybrids form the basis of the most widely cultivated hardwood plantation crop in the world (Eldridge et al. 1994; Grattapaglia et al. 2009). The genus is divided into several subgenera (Pryor and Johnson 1971; Steane et al. 1999; Brooker 2000), of which the subgenus *Symphyomyrtus* is the largest and most diverse containing most of the commercially grown *Eucalyptus* species (Eldridge et al. 1994). The high phenotypic diversity in the subgenus is reflected at the DNA sequence level with nucleotide diversity values ranging from moderately diverse (π = 0.0186) for *E. grandis* (Novaes et al. 2008) to

highly diverse (π = 0.063) for *E. loxophleba* (with single nucleotide polymorphisms every 16 to 33 positions, Kulheim et al. 2009). These values are similar to the levels of nucleotide diversity reported in other outbred forest tree genera such as *Pinus* (π = 0.01-0.02) and *Populus* (π = 0.005-0.01) as reviewed by Neale and Ingvarsson (2008). The high nucleotide diversity in *Eucalyptus* and the recently completed *E. grandis* reference genome sequence (DOE JGI, http://www.phytozome.net) provide opportunities to investigate the evolution and diversity of regulatory networks underlying wood development.

In eukaryotes, gene regulatory networks comprise *cis*-acting sequence elements in promoters, trans-acting elements or transcription factors which bind to promoter sequences and the genes regulated by both kinds of factors. *Cis*-elements are often found clustered together in promoter regions where transcription factors can bind as hetero- or homo-dimers to modulate the transcription of the gene (Reviewed in Farnham 2009). *Cis*-regulatory elements are often shared by the promoters of co-expressed genes due to common trans-regulation (Vandepoele et al. 2009). Another feature of *cis*-element sequences is their conservation in orthologous promoters from different species and genera, in comparison to flanking non-coding sequences (Freeling and Subramaniam 2009). These unique features of *cis*-regulatory elements are employed in a host of computational algorithms for the *in silico* detection of *cis*-elements (Tompa et al. 2005; Das and Dai 2007; Wijaya et al. 2008; Kim et al. 2009). Software programs, such as FootPrinter (Blanchette and Tompa 2003; Fang and Blanchette 2006), PHYLONET (Than et al. 2008) and PhyloScan (Carmack et al. 2007) are based on phylogenetic footprinting algorithms, which use the evolutionary relationships among promoter sequences to identify conserved *cis*-regulatory motifs (Blanchette et al. 2002; Blanchette and Tompa 2002). Phylogenetic footprinting approaches have been used to identify *cis*-elements in a number of plant species including *Populus* and *Eucalyptus* (Creux et al. 2008; Shi et al. 2010).

While cellulose deposition during secondary cell wall formation has been extensively studied in herbaceous and woody plants (Reviewed in Taylor 2008; Popper et al. 2011; Gorshkova et al. 2012), the promoters and *cis*-element composition of the cellulose synthase (*CesA*) genes have been the focus of only a small number of studies (Creux et al. 2008; Lu et al. 2008; Wu et al. 2009). CESA proteins, encoded by the *CesA* gene family, form large membrane-embedded complexes depositing cellulose into plant cell walls (Mutwil et al. 2008). In embryophytes, this gene family forms a distinct clade consisting of 8-18 members per species (Hamann et al. 2004; Roberts and Bushoven 2007; Kumar et al. 2009; Yin et al. 2009). In plants there are two distinct *CesA* gene expression groups, one associated with primary cell wall formation and the other with secondary cell wall formation (Taylor et al. 2004; Desprez et al. 2007). Independent functional analyses of the *CesA* genes in woody and herbaceous species have revealed that orthologous *CesA* genes are functionally conserved in diverse plant species (Tanaka et al. 2003; Taylor et al. 2003; Samuga and Joshi 2004; Ranik and Myburg 2006; Kumar et al. 2009). This conservation likely includes a set of shared *cis*-regulatory sequences and transcription factors since *CesA* gene expression profiles are also highly conserved across different plant genera (Burton et al. 2004; Ranik and Myburg 2006; Creux et al. 2008).

In this study we hypothesize that *cis*-regulatory elements will coincide with regions of lower species-level nucleotide diversity in the promoters of evolutionary distinct *Eucalyptus* tree species. In addition, we hypothesize that different sets of *cis*-elements are conserved in the promoters of primary or secondary cell wall-related *CesA* genes. The objectives were: i) to quantify and assess patterns of species-level nucleotide diversity in the promoters of six *Eucalyptus* cellulose synthase genes, ii) to assess the effects of nucleotide polymorphism on putative *cis*-element occurrences and iii) to identify putative *cis*-elements that are differentially conserved in promoters of primary and secondary cell wall-related *CesA* genes.

This study is the first to characterize the species-level nucleotide diversity and *cis*-element architecture of *CesA* gene promoters in a plant genus and adds to our understanding of the transcriptional regulation of this important plant gene family.

2.3 Materials and Methods

2.3.1 Plant material and DNA isolation

Leaf material was obtained from *Eucalyptus* species conservation hedges established with seed collected from natural stands in Australia (CSIRO, Supplementary file 2.1). High quality genomic DNA was extracted from leaf material using the DNeasy® Plant Mini Kit (Qiagen, Valencia, CA). Genomic DNA was isolated from a single tree from each of 13 *Eucalyptus* species, including *E. fastigata* (subgenus *Eucalyptus*), and 12 species of the commercially important subgenus *Symphyomyrtus* representing three sections, *Latoangulatae*, *Maidenaria* and *Exsertaria* (Supplementary file 2.1). In addition, genomic DNA was also isolated from nineteen *E. urophylla* individuals originating from seed collected on seven Indonesian islands (Timor, Flores, Alor, Pantar, Adonara, Lomblen and Wetar), broadly representing the geographical range of the species (Payn et al. 2008).

2.3.2 Promoter isolation and sequencing

Primer Designer software (version 5, Scientific and Educational Software, Durham, NC) was used to design primers (Supplementary file 2.2) for the amplification of gene and promoter regions based on previously published *E. grandis CesA* gene and promoter sequences (Ranik and Myburg 2006; Creux et al. 2008). For *CesA8* (*EgCesA1*; Table 1.1), PCR amplification of a single gene fragment from each of the 13 *Eucalyptus* species, extending from the

promoter to the end of intron 1, was performed (Supplementary file 2.2). Only the promoter regions (approximately 1 kb upstream of the ATG) were amplified for the other five *CesA* genes. PCR was performed in 20 µl reaction volumes with 30 ng of genomic DNA, 0.4 µM of each primer, 0.20 mM of each dNTP and 0.15 U of ExSel DNA polymerase (Supertherm) with proofreading capability (4-fold lower error rate than standard Taq polymerase, according to manufacturers) using the following conditions: 30 cycles of denaturation at 94°C for 30 seconds, annealing at 56°C for 30 seconds and primer extension at 72°C for 2 minutes. The amplified fragments were cloned (InsT/Aclone, MBI Fermentas, Hanover, MD) and a cloned copy of each promoter fragment was sequenced using overlapping Sanger reads (Macrogen Inc.), representing a single allele of each promoter from each species (Genbank accession numbers: JN573683 - JN573751).

The chromatograms were visually inspected to check the sequence quality of each base. When conflicting bases were identified at the same site the clone was re-sequenced and re-analyzed until a consensus was reached. At least 1000 bp of upstream sequence was analysed for *CesA1*, *3*, *4*, *7* and *8* (for naming convention refer to Table 1.1) while approximately 800 nucleotides of the *CesA6* promoter were used. The 5' upstream regions isolated from each gene started at most 25 bp upstream of the start codon (ATG) and contained the 5' UTR and a minimum of 500 bp of each promoter.

2.3.3 Orthologous gene and promoter sequences of Arabidopsis and Populus

Kumar et al. (2009) proposed a new phylogeny-based nomenclature for the *CesA* genes in *Populus*. Their naming convention allows for direct comparison of the *Arabidopsis* and *Populus CesA* genes (Supplementary file 2.3). This change in nomenclature has been discussed in Chapter 1 and Table1.1 lists the previously published (Ranik and Myburg 2006)

and new naming conventions. Supplementary file 2.3 lists the *Eucalyptus CesA* genes and their orthologs in *Populus* and *Arabidopsis*. Promoter sequences of the *Arabidopsis thaliana CesA* orthologs (*AtCesAs 1-10*) were obtained from The *Arabidopsis* Information Resource (TAIR9, www.arabidopsis.org), and those of the *Populus trichocarpa* orthologs (*PtiCesAs*, Kumar et al. 2009) were obtained from the *Populus* Genome Browser (DOE JGI http://www.phytozome.net, Supplementary file 2.3). The same set of promoter sequences were used as in the study by Creux et al. (2008). Additionally, the orthologous gene sequences of *Eucalyptus CesA8*, *AtCesA8* (At4g18780) and *PtiCesA8-A* (Pti235238), were obtained from the same databases. To ensure that the regions of the *Eucalyptus CesA8* gene and its orthologs could be compared, the promoter regions were trimmed to equal length (297 bp) in all orthologs. The first intron region was also trimmed (to 68 bp), resulting in an analysis region that included part of the proximal promoter, 5' UTR (untranslated region), exon 1 and part of intron 1 (Accession numbers JN573752 - JN573783).

2.3.4 DNA sequence analysis

After removal of vector sequences, DNA sequences of the *Eucalyptus CesA* promoters were assembled using the Vector NTI software package (version 9.1.0, Invitrogen). The sequences were aligned using the Clustal W (Thompson et al. 1994) function of BioEdit (version 7.0.9, Hall 1999). DNA sequence analysis and nucleotide diversity (π; Nei and Li 1979) and θ_W; (Watterson 1975) calculations were performed using DnaSP (DNA Sequence Polymorphism, version 4.50.3, Rozas et al. 2003). The distribution of nucleotide diversity in the sequences was graphically represented using per site sliding windows (50 bp or 100 bp) of π.

2.3.5 Cis-element selection

We focused on 22 previously reported, putative *cis*-elements, most of which have not been functionally characterized, and we refer to these from this point forward as *cis*-elements without ascribing any functional annotation (Table 2.1). First, putative *cis*-elements were selected based on their over-representation in *CesA* promoters associated with primary (CRPE17, CRPE 12, CRPE 11, CRPE 10, CRPE8, CRPE6) or secondary (CRPE31, CRPE28, CRPE26, CRPE25) cell wall formation, as identified in a comparative study of *Eucalyptus*, *Arabidopsis* and *Populus CesA* promoters (Creux et al. 2008). A second set of *cis*-elements was identified by using the PLACE database homology search tool (http://www.dna.affrc.go.jp/PLACE/, Higo et al. 1999). *Cis*-elements with motif lengths of greater than five base pairs were selected if they were present in the promoters of all three primary or all three secondary cell wall-related *CesA* genes (Table 2.1). The xylem-specific promoter element identified by Ko et al. (2006) and the tracheary element-specific motif identified by Pyo et al. (2007) were also added to the set of *cis*-elements as they have been suggested to play a role in secondary cell wall formation in *Arabidopsis*.

2.3.6 Cis-element mapping

Pattern Matching in RSA-Tools (http://rsat.scmbb.ulb.ac.be/rsat/; Thomas-Chollier et al. 2008) was used to map occurrences of the selected *cis*-elements onto the six *CesA* promoter sequences of each of 13 *Eucalyptus* species. Pattern matching and image generation were conducted using the default settings. A random dataset was generated using RSA-Tools Random Sequence Generator, which calibrated the sequences on *Arabidopsis* non-coding upstream sequences. An *Arabidopsis*-specific Markov model was used to generate the random sequences. To visualise patterns of *cis*-element conservation and variation across

related species and sections (Supplementary file 2.1), we constructed a neighbour joining tree in MEGA 5 (Tamura et al. 2011) based on the concatenated promoter sequences from each species. The resulting species order was used to arrange the promoters for *cis*-element mapping.

Promoter sequences containing the mapped *cis*-elements were divided into discrete 100 bp intervals (i.e. -1 to -100, -101 to -200, etc.), and *cis*-element occurrences in each section were counted and graphically represented in comparison with the random data set. A two-tailed t-test assuming equal variance was performed, to identify regions that showed significant differences from the random dataset with $\alpha = 0.01$ and $\alpha = 0.001$. Motif logos were generated using the output sequences from RSA-Tools in the online motif logo tool, Weblogo (http://weblogo.berkeley.edu/logo.cgi), with all default settings (Schneider and Stephens 1990; Crooks et al. 2004). These were compared to the *cis*-element consensus sequences that were generated across all three genera (*Arabidopsis*, *Populus* and *Eucalyptus*) using the same tools (Supplementary file 2.4).

2.3.7 Cis-element conservation analysis

Cis-element conservation was estimated by manually counting the number and type of nucleotide changes that occurred within *cis*-element occurrences (Supplementary files 2.5 and 2.6). The *cis*-element occurrences could be grouped into three categories: conserved *cis*-element occurrences, where no changes were observed in any of the *Eucalyptus* species analysed; moderately conserved *cis*-element occurrences, where only a single position in the *cis*-element sequence was changed in one or more species; and non-conserved *cis*-element occurrences, where more than one position in the *cis*-element sequence was changed in one or more of the species analysed. The types of nucleotide changes were also classified as

singletons (occurring in only one of the 13 species), or polymorphisms (occurring in two or more species). The different *cis*-element and mutation counts were entered into Excel (Microsoft Office 2007), where the averages and percentages of the different polymorphism affecting *cis*-element occurrences were calculated (Supplementary file 2.6).

2.4 Results

2.4.1 Sequence divergence and nucleotide diversity of the Eucalyptus CesA8 gene and promoter at the population, species and genus levels

We compared the nucleotide diversity of the proximal promoter and 5' regions (ATG was the anchor point placed at position 0 bp) of the secondary cell wall-related *Eucalyptus CesA8* gene (Ranik and Myburg 2006), including representative portions of the 5' UTR, first exon and first intron to the orthologous cellulose synthase genes of *Arabidopsis* and *Populus* (Figure 2.1, Supplementary file 2.1: *CesA8* in *Populus* and *Arabidopsis*; Kumar et al. 2009). Nucleotide diversity in these regions was calculated for three sets of sequences: (1) the genus-level comparison of the *E. urophylla CesA8* gene to its *Arabidopsis* and *Populus* orthologs (Supplementary file 2.3), (2) the species-level comparison of the *CesA8* gene sequences from thirteen *Eucalyptus* species (Supplementary file 2.1) and finally, (3) a population-level comparison which included sequences from a population sample of 19 *E. urophylla* trees (Figure 2.1). As expected, the average nucleotide diversity over the whole 611 bp region was highest for the genus-level comparison (π = 0.461), which was close to a value expected for unrelated sequences (Figure 2.1, green trend line). The *Eucalyptus* species- and population-level comparisons exhibited significantly lower average diversity (π = 0.015 and π = 0.006, respectively) in the same region. The observed nucleotide diversity at the population level (Figure 2.1, black trend line) was generally below 1% and the diversity

observed in parts of the promoter and first exon was due to single nucleotide polymorphisms (SNPs). The nucleotide diversity of the promoter was higher at the species-level (Figure 2.1, red trend line), but distinct regions of relative sequence conservation were observed within the promoter region (Figure 2.1, black arrows). These results demonstrate that at the genus-level, the promoters are too divergent structurally to observe conserved sequence elements by direct sequence comparison. Conversely, the population-specific comparison suggested that the sequences were too similar to identify defined regions of conservation. The species-level comparison, however, did reveal conserved regions in the promoters, which may contain clusters of *cis*-regulatory elements and the other five *CesA* promoters (*CesA1*, 3, *4*, 6 and *7*) were consequently investigated at this level.

2.4.2 Isolation and analysis of CesA promoter regions of 13 Eucalyptus species

The upstream regions of six *CesA* genes (*EgCesA1*, *3*, *4*, *6*, *7* and *8*; Table 1.1) were isolated from one individual of each of 13 *Eucalyptus* species and used to investigate detailed patterns of sequence and spatial conservation among the *Eucalyptus* species. Creux et al. (2008) showed that *E. grandis CesA3* had an intron in the 5'UTR and this intron was also present in the *CesA3* promoter of the other 12 *Eucalyptus* species investigated here, indicating conservation in the 5'UTR of the gene. It is well documented that 5' UTR sequences and the associated introns can play important roles in *cis*-regulation (Karthikeyan et al. 2009; Livny and Waldor 2009). For this reason, we disregarded promoter/ UTR boundaries and in all instances, the entire upstream regions including the 5' UTR (and intron in the case of *CesA3*) were analysed for regulatory element occurrence and conservation.

Despite several rounds of primer optimisation and design, we were unable to isolate the upstream regions of the *CesA1* gene from *E. camaldulensis*, *E. tereticornis* and *E. dunnii*

presumably due to high sequence divergence in the upstream priming sites. Amplification of the *E. fastigata CesA1* promoter was only achieved when the primers were moved immediately upstream of the original binding sites, which were subsequently found to contain *E. fastigata*-specific sequence polymorphisms (primer sequences, Supplementary file 2.2). As a result, the *CesA1* promoter dataset only contained sequences from ten *Eucalyptus* species. Furthermore, two distinct sequence haplotypes of the *CesA1* promoter were observed among the ten *Eucalyptus* species analysed (Supplementary files 2.4 and 2.5). Species from the subgenus *Eucalyptus* (*E. fastigata*) and *Symphyomyrtus* section *Latoangulatae* (*E. grandis, E. urophylla* and *E. saligna*) shared a single haplotype. The second haplotype was only observed in species of the section *Maidenaria* (*E. macarthurii, E. globulus maidenii, E. globulus globulus, E. globulus bicostata, E. smithii* and *E. nitens*, Supplementary files 2.4 and 2.5). PCR amplification with haplotype-specific primers confirmed that the two putative haplotypes do not co-occur within any of the *Eucalyptus* species analysed and are therefore not likely to be derived from paralogous sequences in the *Eucalyptus* genome (data not shown).

There are 11 expressed *CesA* genes in *Eucalyptus* (Mizrachi et al. 2010) and to ensure that no duplicate promoters were isolated all of the *CesA* promoter sequences analysed in this study were compared to the *E. grandis* genome sequence (DOE JGI, http://www.phytozome.net). In all cases the sequences matched a single region directly upstream of the corresponding *CesA* gene, supporting our inference that no paralogous promoters were isolated. Sequence comparison of the two *CesA1* promoter haplotypes with the genome sequence also confirmed that the two haplotypes correspond to the promoter regions of a single *CesA1* gene locus in the *E. grandis* genome (data not shown, Phytozome gene ID: Eucgr.C02801.1).

2.4.3 Species-level sequence diversity in the promoter regions of the Eucalyptus CesA genes

The average species-level nucleotide diversity (π) of the six *Eucalyptus CesA* promoter regions varied from π = 0.014 for *CesA6* to π = 0.068 for *CesA1* (Table 2.2). The high species-level nucleotide diversity observed in the *CesA1* promoter regions could be ascribed to the presence of the two distinct haplotypes (Supplementary files 2.5 and 2.7). In all of the promoters a local decrease in nucleotide diversity across species was observed at the transcriptional start site (Figure 2.2, shaded boxes). Additionally, in the *CesA3*, 4, 6, 7 and *8* promoters there were regions further upstream where local species-level diversity was below π = 0.02, which is the nucleotide diversity expected for conserved coding regions and could indicate functional constraints within these regions (Figure 2.2).

2.4.4 Cis-element position and frequency in the promoter regions of the Eucalyptus CesA genes

We investigated the positional conservation of *cis*-elements in the *CesA* promoter regions by mapping occurrences of 22 previously identified putative *cis*-elements (Table 2.1) in the promoter sequences of 13 *Eucalyptus* species (Figure 2.3 and Supplementary file 2.5). Three of the *cis*-elements (CRPE25, CRPE26 and TERE; Table 2.1) could not be found in any of the 13 *Eucalyptus CesA* promoter sequences, even when mismatches were allowed, but they were detected in the *Arabidopsis* and *Populus* orthologs (Figure 2.3). No occurrences of CRPE11 could be identified in any of the sequences analysed, including the *Eucalyptus* sequences, possibly due to the stringent search parameters used in the present study compared to Creux et al. (2008). *Cis*-element occurrence counts revealed regions with significantly higher/lower number of occurrences, while an even distribution was observed in the random dataset (Figure 2.4). The occurrences of each *cis*-element in the *Eucalyptus CesA* promoters

were used to generate *Eucalyptus*-specific *cis*-element consensus sequences which were very similar at the species- and genus-level, even in sites allowing for alternative bases (Supplementary file 2.4).

A phylogeny based on the promoter sequences was constructed so that the promoters could be ordered accordingly. This phylogeny reflected the extensive phylogeny constructed by Steane et al. (2011). Ordering the promoters in this way allowed for the identification of *cis*-element occurrences that were variable across the different Symphyomyrtus sections. An example of this is the occurrence of the MYB1AT element at -600 bp in the *CesA6* promoters of species only in the Exsertaria section (*E. tereticornis* and *E. camaldulensis*) and is not present in the *CesA6* promoters of the other *Eucalyptus* species (Figure 2.3). This may indicate that some *cis*-element occurrences are dependent on the species phylogeny, however, to understand the complexities of this relationship further investigation would be required.

Mapping the *cis*-element occurrences to the promoter regions (Figure 2.3 and Supplementary file 2.5) allowed us to evaluate the *cis*-element content of conserved promoter regions (Figure 2.2). In each of the *Eucalyptus* promoters, a set of co-occurring elements (Figure 2.3, transparent grey blocks) was identified near the previously predicted TSSs (Creux et al. 2008). This set of putative *cis*-elements comprised multiple occurrences of the CRPE31 (GNGNAGNA, Figure 2.3: orange) and CTRMCAMV35S (TCTCTCTCT, Figure 2.3: purple) motifs with the exception of *CesA7* where only CRPE31 occurred approximately 200 bp upstream of the predicted TSS (Figure 2.3). These were not conserved in the *Arabidopsis* and *Populus* promoters (Figure 2.3). In *CesA6*, however, the cluster and region of lower species-level nucleotide diversity was located further upstream (-450 to –300) of the TSS and may indicate that the predicted TSS for this gene should be re-evaluated. The TATABOX 5 element (TTATTT, Figure 2.3: pink), from the PLACE database, was the only

TATA-box-like sequence identified in the *Eucalyptus CesA* promoters. In the *CesA8*, 1 and 6 promoters, the putative TATA-box-like motif was found far upstream (400 to 1000 bp) from the TSS (Figure 2.3). The *cis*-element maps also enabled us to observe positional conservation of *cis*-element combinations. MYB1AT (WAACCA, Figure 2.3: black) and NODCON1GM (AAAGAT, Figure 2.3: brown) were found to co-occur in the secondary cell wall-related *CesA* promoters within 200 bp of each other (Figure 2.3 A, B and C). These two putative *cis*-elements also appeared to be positionally conserved within these promoters as they were always observed in the region between -200 to -600 bp upstream of the TSS in *CesA4*, 7 and 8 (Figure 2.3 D, E and F).

2.4.5 Cis-element evolution in the promoter regions of the Eucalyptus CesA genes

To investigate the potential effect of DNA sequence evolution on *cis*-element occurrences in *Eucalyptus*, all of the putative *cis*-element occurrences in the *Eucalyptus CesA* promoters were investigated for sequence conservation. The individual *cis*-element occurrences were scored as conserved if no polymorphism occurred in the region across all 13 promoter sequences (Figure 2.5 A and Supplementary file 2.6). Putative *cis*-element occurrences that had a single nucleotide polymorphism (present in two or more cloned sequences) relative to the *Eucalyptus* consensus sequence (Supplementary file 2.4) and those that had two or more nucleotide changes away from the consensus sequence were counted separately (Figure 2.5 A and Supplementary file 2.6). Overall, only 29% of the *cis*-element occurrences investigated were fully conserved in the 13 *Eucalyptus* promoter sequences. The *Eucalyptus CesA8* promoters had the highest number of fully conserved *cis*-element occurrences (59%), while *CesA1* had the lowest at 13% (Figure 2.5 A).

We identified 89 instances (42% of all occurrences) where a putative *cis*-element occurrence was present or absent in all but one of the 13 promoters (Supplementary file 2.6). These instances were classified as singleton (gain or loss) occurrences and 70% of these resulted from single nucleotide changes, while 30% were due to indels (Figure 2.5 B). Some of the single nucleotide changes could have resulted from cloned PCR errors, however the observed frequency of singletons was approximately ten times higher than would be expected from polymerase induced errors (average $1x10^{-4}$ per base pair; Keohavong and Thilly 1989; Ling et al. 1991). We investigated the frequency of singleton mutations that resulted in a change towards the consensus sequence (i.e. a change that rendered the *cis*-element identifiable by the software in that promoter) or away from the consensus sequence (i.e. a change rendering the consensus sequence unidentifiable by the software). The majority of sequence changes (72% of the SNPs and 71% of the indels, Figure 2.5 B) resulted in a change towards the *cis*-element consensus sequence (Supplementary file 2.4) and therefore may indicate a gain (or maintenance) of the *cis*-element occurrence in that species. However, to accurately investigate this, more individuals should be sequenced for each species and experimentally verified *Eucalyptus cis*-elements should be used.

2.5 Discussion

CesA gene family members have conserved roles in the deposition of primary and secondary cell walls in all seed plants studied to date (Burn et al. 2002; Burton et al. 2004; Hamann et al. 2004; Ranik and Myburg 2006). This suggests that the major clades of the gene family differentiated early during Spermatophyte evolution (Sarkar et al. 2009; Yin et al. 2009; Popper et al. 2011). A distinctive set of expression patterns characterize the major clades with *CesA* genes involved in primary and secondary cell wall deposition exhibiting unique, developmentally regulated expression profiles (Hamann et al. 2004). The conserved nature of

these genes and their highly coordinated, but differential, expression patterns point to the action of a conserved network of *cis*- and trans-regulatory factors in plants (Demura and Fukuda 2007; Zhong et al. 2010). The aim of this study was to characterize the architecture and diversity of *cis*-elements in six *CesA* promoters across *Eucalyptus* tree species in terms of sequence and *cis*-element conservation.

Nucleotide diversity levels are generally higher in promoter regions than in other genic regions, most likely due to lower overall functional constraints on promoter regions (Nei 2007). That said, the maintenance of functional *cis*-elements in promoter regions and the modular nature of transcription factor binding suggests that there should be localised regions that are more conserved than the rest of the promoter and contain clusters of *cis*-elements (Maniatis et al. 1987; Ho et al. 2009). To determine whether this is the case for *CesA* genes, we compared nucleotide diversity levels in different *CesA8* gene regions at the population, species and genus levels. While the nucleotide diversity in the genus and population levels were at the highest and lowest extremes of the scale respectively, the diversity at the species level showed a large range of values from 1.4% to 6.8% (Figure 2.1). At the species level the nucleotide diversity plot profile revealed that the higher nucleotide diversity in the upstream regions was interrupted by areas with distinctly lower nucleotide diversity (Figure 2.1 - red line and arrows). In a similar study on the *Drosophila* bithorax complex, Ho et al (2009) also identified regions of conservation across orthologous promoters in different *Drosophila* species and suggest that short conserved sequences are a feature of related eukaryotic promoters.

The conservation of promoter sequences has been documented for a number of different plant species, but sequence conservation in promoter regions may not always indicate *cis*-element conservation (Reineke et al. 2011). In light of this, we evaluated the occurrences of previously identified *cis*-elements (Table 2.1) in the cloned *Eucalyptus*

promoter sequences. We found that 76% of the *cis*-element occurrences were either fully conserved (29%) in the 13 *Eucalyptus* promoters or only varied from the consensus sequence by a single nucleotide change (47%, Figure 2.5 A and Supplementary file 2.6). In this study the *cis*-element occurrences which are affected by one nucleotide change are likely also to be conserved elements because *cis*-element consensus sequences often contain some ambiguous bases and transcription factors can still bind to these variable sites. The *cis*-element occurrences with a single change could also be more conserved than estimated here since the changes may be due to allelic variation within the different species, but this variation would have to be further investigated at the population level within each species when a robust set of experimentally tested *cis*-elements are available for *Eucalyptus*.

Identifying conserved *cis*-elements or regions in the promoter does not fully describe the spatial arrangement of these elements, which is important because *cis*-elements are often found clustered together, rather than evenly distributed across the length of the promoter (Maniatis et al. 1987; Hansen et al. 2010). We found that there was strong clustering of *cis*-elements at particular intervals in the promoters when compared to neighbouring intervals or to a random dataset (Figure 2.4). A significant cluster of *cis*-elements was observed at the position of the TSS in a number of the *CesA* promoters (Figure 2.4: *CesA1*, *3*, *6* and *8*), and the position of this *cis*-element cluster (TSS-associated cluster) also coincided with a highly conserved promoter region (Figure 2.2; transparent grey blocks). In a genome-wide comparative study of *Populus* and *Arabidopsis* a number of conserved *cis*-element clusters were also identified in the promoters of genes associated with cellulose deposition (Ding et al. 2012) and this could indicate functional constraints acting on the particular *CesA* promoter regions.

The TSS-associated cluster of *cis*-elements identified in five of the six *Eucalyptus* *CesA* promoters (except *CesA6*) contained multiple occurrences of two *cis*-elements CRPE31

and CTRMCAMV35S (Figure 2.3: purple). CRPE31 (GNGNAGNG, Figure 2.3: orange) was the most abundant *cis*-element detected in the *Eucalyptus* promoters as well as in a previous *cis*-element study of *CesA* promoters (Creux et al. 2008). The reverse compliment sequence of this element (CNCTNCNC) could be an initiator element similar to the elements described in *Arabidopsis* (Bernard et al. 2010). The initiator element has been shown to co-occur with the TSS in the promoters of different plant species (Yamamoto et al. 2007) and this may suggest a function for these putative CRPE31 occurrences in the *Eucalyptus* promoters.

The other main element in the TSS-associated cluster was the CTRMCAMV35S element (TCTCTCTCT, Figure 2.3: purple). In the PLACE database, this element is listed as an enhancer found in the commonly used Cauliflower Mosaic Virus (CAMV) 35S viral promoter (Pauli et al. 2004). This element is highly over-represented in the *CesA* promoters in the form of TC-repeats (Supplementary file 2.5). Multiple copies of the CTRMCAMV35S element could enhance the expression of the *CesA* genes which would be of interest because these are some of the most highly expressed *Eucalyptus* genes (Mizrachi et al. 2010). Alternatively, the CTRMCAMV35S repeats could represent plant-specific regulatory regions known as Y-patches (CT or TC repeats) which have been associated with the TSSs of other plant promoters in genome-wide studies of *Arabidopsis* and rice (Molina and Grotewold 2005; Yamamoto et al. 2007). The putative CRPE31 and CTRMCAMV35S element occurrences in this conserved TSS-associated cluster may play a role in the initiation of transcription of a number of the *Eucalyptus CesA* genes.

The suggestion that transcriptional initiation of the *Eucalyptus CesA* genes could be reliant on the presence of putative initiator (CRPE31) and Y-patch (CTRMCAMV35S) elements is further bolstered by the lack of a TATA-box in many of the *Eucalyptus CesA* promoters (Figure 2.3). The only over-represented element resembling a TATA-box (PLACE ID: TATABOX5; TTATTT) was found in a subset of the *CesA3* promoter sequences in the

correct location (Figure 2.3: pink). This TATA-box-like element was also identified in the *Eucalyptus CesA1*, *3*, *6* and *8* promoters, but its position in each instance was ≥ 600 bp upstream of the TSS (Figure 2.3, pink) and did not coincide with any major decrease in species-level nucleotide diversity. This suggests that these sequences do not function as TATA-boxes, because in the original description of this element it was functional when located 50 bp upstream of the TSS (Tjaden et al. 1995). Additional support for this TSS-associated cluster comes from our previous study which reported some of the *Eucalyptus grandis CesA* promoters, which contain the TSS-associated cluster, to be TATA-less. However, the promoter fragments were still able to drive tissue-specific expression of the GUS reporter gene in *Arabidopsis* and had experimentally verified TSSs (Creux et al. 2008).

Further upstream of the TSSs, other putative *cis*-elements could be observed as singletons, pairs or clusters in the *CesA* promoters (Figure 2.3). One example of this is the co-occurrence of the NODCON1GM (AAAGAT, Figure 2.3: dark brown) and MYB1AT (WAACCA, Figure 2.3: black) elements in the *CesA* promoters of the genes associated with secondary cell wall deposition. In *CesA4*, *7* and *8*, the elements co-occurred in the region between -200 and -600 bp, and this positional conservation was not observed in the promoters of the *CesA* genes associated with primary cell wall formation (Figure 2.3). This suggests that NODCON1GM and MYB1AT may play a role in secondary cell wall-specific expression of these genes.

The putative MYB1AT occurrences in *CesA4*, 7 and 8 were of special interest for a number of reasons. Yazaki et al. (2003) found the MYB1AT element to be involved in *Arabidopsis* dehydration stress response while investigating gibberellic acid (GA) and abscisic acid (ABA) responses in rice. GA has also been found to play a role in cambial cell differentiation and xylem development (Love et al. 2009) pointing to the co-regulation of genes involved in xylogenesis. This putative MYB-like element may also be important as a

number of MYB transcription factors have been identified as key members of the transcriptional network regulating the secondary cell wall formation (Goicoechea et al. 2005; Demura and Fukuda 2007; Legay et al. 2010; Zhong et al. 2010).

While many of the putative *cis*-element occurrences investigated in this study were conserved in the promoters of the different *Eucalyptus* species, a small number of occurrences varied among species (Figure 2.5 B). We counted 89 singleton changes (i.e. loss or gain of an occurrence in only one of the 13 sequences) in the *CesA* promoters and 66% of the sequence polymorphisms changed the promoter sequence towards a known *cis*-element consensus sequence (Supplementary file 2.6). One hypothesis for this *cis*-element variation is known as *cis*-element buffering, which ensures the maintenance of a particular *cis*-element sequence in faster evolving sequences such as promoters (Tanay et al. 2005). It proposes that the promoter sequence may change to abolish a particular *cis*-element occurrence, but a second mutation in that promoter maintains the binding site of a particular transcription factor. This mechanism of *cis*-element maintenance has been observed in the *cis*-regulatory modules of a number of *Drosophila* genes (Hare et al. 2008; Ho et al. 2009) and could be a mechanism of maintenance in the *CesA* gene promoters. However, this can only be confirmed by further investigation on both the population-wide and genome-wide level with experimentally verified *cis*-elements from *Eucalyptus*.

An interesting observation form this study was that, while most of the *Eucalyptus CesA* promoters were similar in terms of GC content and number of indels, the *CesA1* promoters were distinct. The *CesA1* promoter dataset had the lowest GC content (36%) and more than double the number of indels than the other promoters (Table 2.2). *CesA1* also presented with the lowest number of conserved *cis*-element occurrences (Figure 2.5 A). Sequence analysis revealed at least two *CesA1* promoter haplotypes in the *Eucalyptus* dataset and these were congruent with the different sections of *Symphyomyrtus* in which they

occurred (Supplementary files 2.1 and 2.5). The promoter regions of species belonging to the section *Exsertaria* (*E. camaldulensis* and *E. tereticornis*) could not be amplified and might represent yet another haplotype. It appears that the distal part of the *CesA1* promoter is undergoing rapid divergence, and the pattern of divergence is in keeping with the phylogeny of the *Eucalyptus* species. Since haplotype 1 is shared among representatives of both subgenera, it is likely to be ancestral and the alignment of the *CesA1* promoter sequences (Supplementary files 2.4 and 2.5) suggests that multiple insertional events have occurred. The high sequence divergence observed in the distal region (upstream from -390 bp) suggests that the proximal region of the *CesA1* promoter may harbour most of the essential *cis*-elements for this gene.

2.6 Conclusions

This is one of only a few studies that have investigated sequence diversity in the promoters of a plant gene family (Koch et al. 2001; de Meaux et al. 2006; Tanaka et al. 2009; Zhao et al. 2009). We identified regions within the *CesA* promoters that are conserved across *Eucalyptus* species and coincided with the putative occurrences of *cis*-elements, suggesting that they have important biological functions. Overall, we found that 29% of the investigated *cis*-element occurrences were fully conserved in *Eucalyptus CesA* genes. Only 30% of the singleton changes were away from the consensus sequence (Figure 2.5 B), suggesting there are functional constraints on some sequences within the promoter regions. The *CesA* promoters of *Eucalyptus* appear to be TATA-less and a highly conserved region in these promoters was identified in the vicinity of the TSSs, suggesting that the basal transcriptional machinery for this *Eucalyptus* gene family relies on other basal *cis*-elements such as a putative initiator element and Y-patch to initiate transcription. Conserved *cis*-elements were also found in the promoters of the *CesA* genes associated with secondary cell wall formation.

These were not present in the primary cell wall-associated *CesA* promoters. The study provides insight into the diversity and evolution of *cis*-regulatory sequences underlying the unique expression profiles of this important plant gene family and will lay the foundation for future studies on the function of these promoter regions and for comparative genomic analysis of promoter elements across multiple *Eucalyptus* genomes.

2.7 Acknowledgements

The authors would like to acknowledge Alisa Postma for her contribution to the preliminary study leading to this research and Dr. Albe van der Merwe and Dr. Christine Martiz-Olivier for comments and guidance during the preparation of the manuscript. This work was supported with funding provided by Mondi and Sappi, through the Forest Molecular Genetics (FMG) Programme, the Technology and Human Resources for Industry Programme (THRIP) and the National Research Foundation of South Africa (NRF).

Table 2.1 Details of 22 *cis*-regulatory elements selected from literature and PLACE database scans and used for DNA pattern matching

Source [a]	Motif Identity [b]	Motif Sequence [c]	PLACE Annotation [d]
Primary cell wall-associated motifs[e]			
Creux et al. (2008)	CRPE 17	GTCKGT	Unknown
Creux et al. (2008)	CRPE12	ATNWATTA	Phosphate response domain
Creux et al. (2008)	CRPE11	GACNGTSNGTGGGC	Stem enhancer element
Creux et al. (2008)	CRPE10	CCNCMCCC	Vascular-specific expression
Creux et al. (2008)	CRPE8	GGNGGTGG	Anthocyanin regulatory element
Creux et al. (2008)	CRPE6	NMTTCTGTC	Iron deficiency responsive element
Place DB	CAREOSREP1	CAACTC	Gibberellin up-regulated proteinase expression
Place DB	DRE2COREZMRAB17	ACCGAC	Drought-responsive element
Place DB	RBCSCONSENSUS	AATCCAA	Light-regulated expression
Place DB	TATABOX5	TTATTT	Functional TATA element
Place DB	SEF4MOTIFGM7S	RTTTTTR	Beta-conglycinin enhancer
Secondary cell wall-associated motifs[e]			
Creux et al. (2008)	CRPE31	GNGNAGNG	Unknown
Creux et al. (2008)	CRPE28	NNGCATGC	Iron deficiency response element
Creux et al. (2008)	CRPE26	TCCTGCYG	Unknown
Creux et al. (2008)	CRPE25	RCYSTGCCC	Phloem-specific expression
Place DB	CTRMCAMV35S	TCTCTCTCT	Enhancer of gene expression
Place DB	REALPHALGLHCB21	AACCAA	Phytochrome regulatory elements
Place DB	PYRIMIDINEBOXOSRAMY1A	CCTTTT	Pyrimidine box involved in sugar repression
Place DB	NODCON1GM	AAAGAT	Organ-specific element
Place DB	MYB1AT	WAACCA	Activation drought and ABA-induced expression
Pyo et al. (2007)	TERE	CTTNAAAGCNA	Tracheary element-specific expression
Ko et al.(2006)	XYLAT	ACAAAGAA	Xylem-specific

[a] Original source of the *cis*-element

[b] Published name or identity of the *cis*-element

[c] Published consensus sequences for the *cis*-element motifs with ambiguous bases represented as IUPAC codes where W = A/T, M = A/C, R = A/G, K = T/G, S = G/C, Y = C/T and N represents any of the four bases.

[d] Putative function of the *cis*-elements as reported in literature or the PLACE database.

[e] The *cis*-element motifs were assigned as primary or secondary cell wall-associated based on the study in which they were first identified and/ or the description on the PLACE database.

Table 2.2 Species-level nucleotide diversity in the promoter regions of six cellulose synthase (*CesA*) genes cloned from 13 *Eucalyptus* tree species

	CesA8	***CesA4***	***CesA7***	***CesA3***	***CesA1***	***CesA6***
Number of species analysed	13	13	13	13	10[b]	13
Length of aligned sequence (including gaps)	1132	1284	1317	1286	1970	863
G+C content (%)	43	47	45	47	36	53
Total number of sites (excluding gaps)	989	1048	1238	1168	1260	730
Number of polymorphic sites	66	134	93	105	253	51
Total number of singleton sites	72	138	95	110	261	52
Nucleotide diversity (π)	0.018	0.029	0.018	0.021	0.068	0.014
Nucleotide diversity (θ_W)	0.023	0.042	0.025	0.030	0.073	0.023
Total number of insertions and deletions (indels)[a]	4	6	3	6	15	9

[a] Including repeat regions and indels of varying lengths occurring in more than one species analysed.

[b] The *CesA1* promoter region could only be isolated from 10 species.

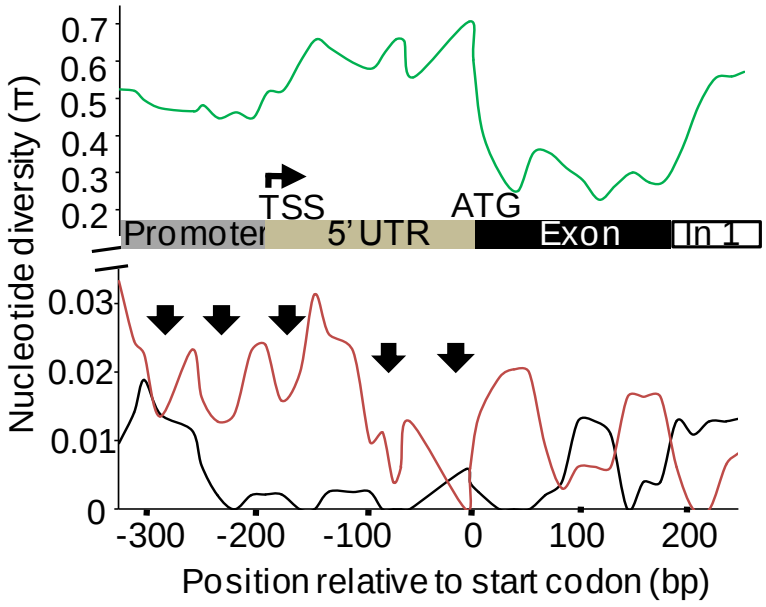


Figure 2.1 Nucleotide diversity (π) in the proximal promoter and gene regions surrounding the translational start site of the *Eucalyptus CesA8* gene (orthologous to *CesA8* in *Arabidopsis* and *Populus*) at the genus, species, and population levels. Gene regions are depicted by the horizontal bar at the centre of the graph. Nucleotide diversity among *Eucalyptus urophylla CesA8* and its *Arabidopsis thaliana* (*AtCesA8*) and *Populus trichocarpa* (*PtiCesA8*-A) orthologs is shown in green. Species-level nucleotide diversity in the corresponding regions of the *CesA8* gene from thirteen *Eucalyptus* species is shown in red, while nucleotide diversity in the *CesA8* gene from a population of *E. urophylla* trees is shown in black. Black arrows indicate promoter and 5' UTR localized regions with lower nucleotide diversity. TSS - transcriptional start site, ATG - translational start codon

---

Figure 2.2 Species-level nucleotide diversity profiles of the promoter regions from six cellulose synthase (*CesA*) genes in 13 *Eucalyptus* species. Nucleotide diversity (π) per site is represented by a grey line and the moving average by a black line. Nucleotide position is indicated relative to the start of translation (ATG, +1) with the transcriptional start site indicated by an asterisk. The line and coloured blocks at the bottom of each graph show the position of the mapped *cis*-element occurrences in the *E. grandis* reference sequence. A *cis*-element colour key is at the bottom of the figure. Transparent grey blocks indicate the position of the TSS-associated *cis*-elements coinciding with regions of lower species-level nucleotide diversity

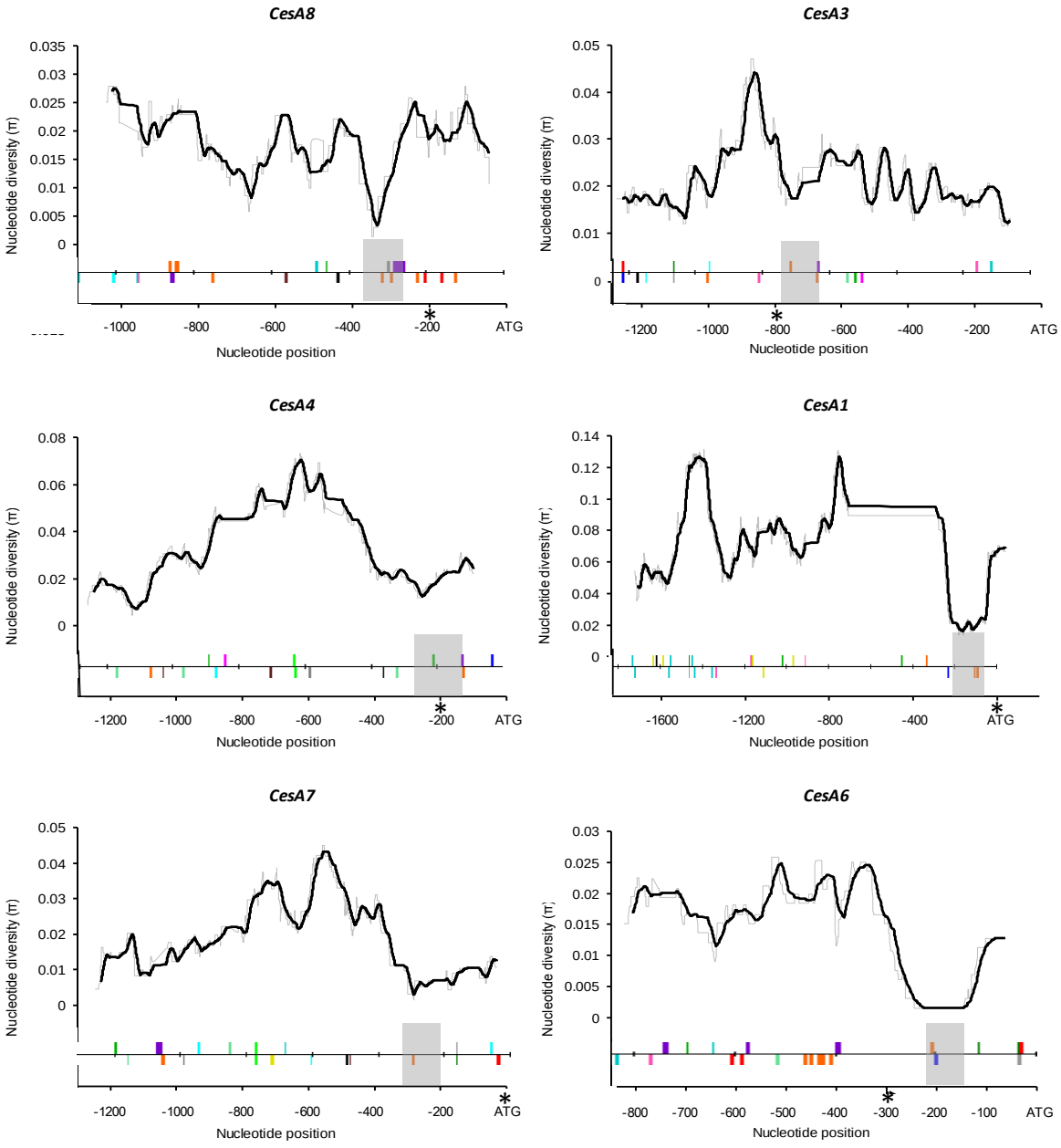
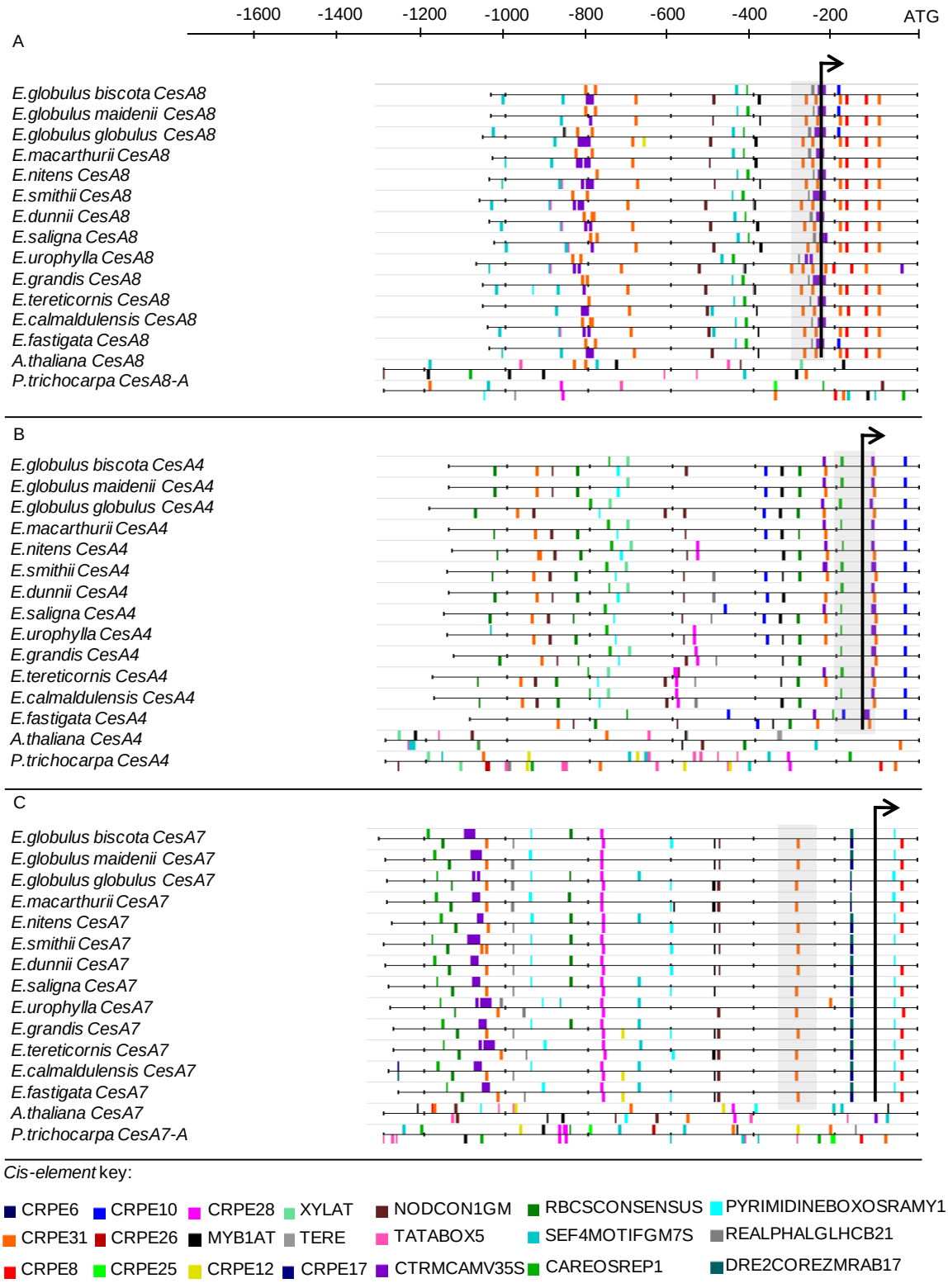
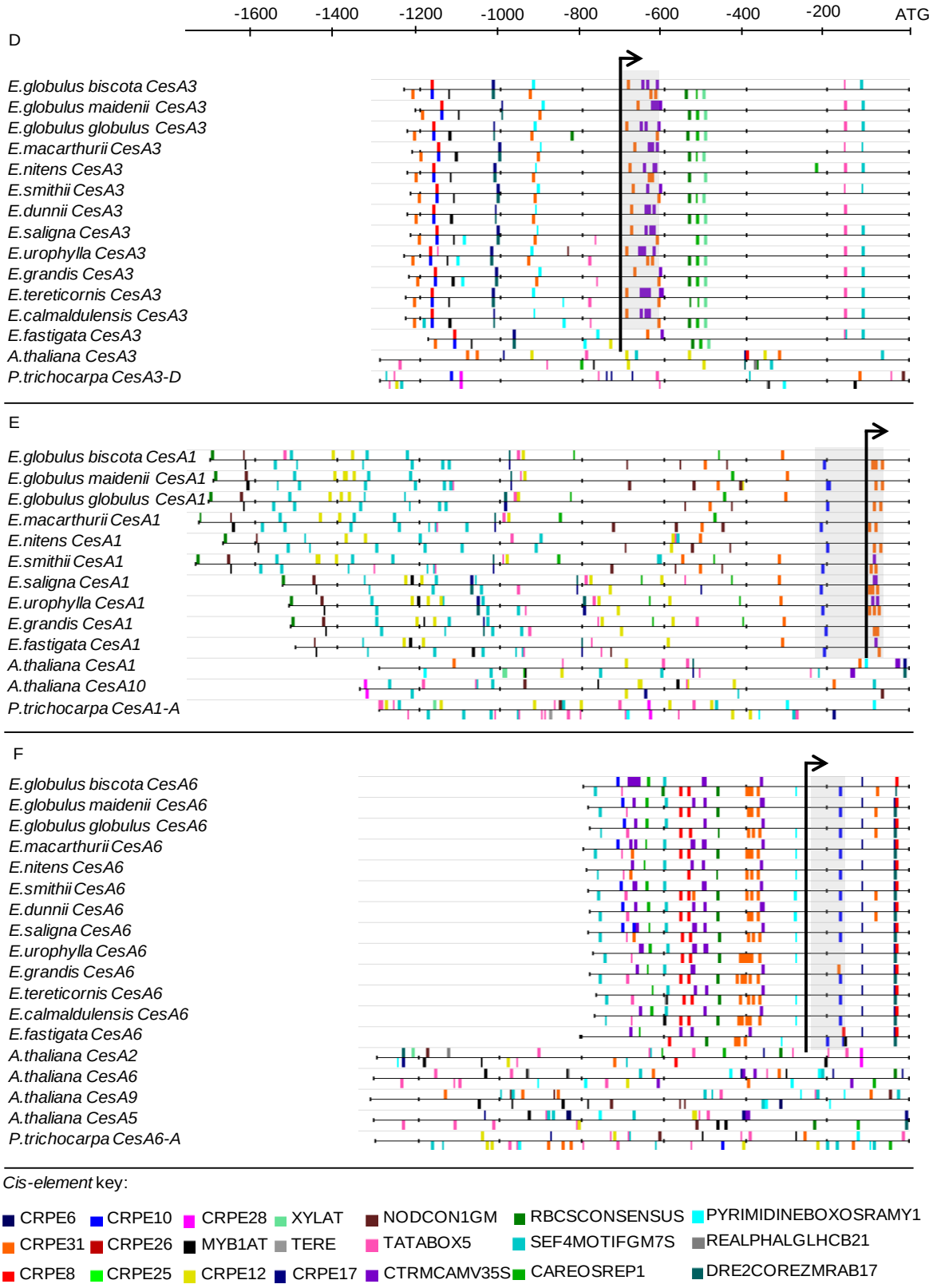


Figure 2.3 Occurrences of 21 putative *cis*-regulatory elements mapped in the promoters of six orthologous groups of *CesA* genes in 13 *Eucalyptus* species, *Arabidopsis thaliana* and *Populus trichocarpa*. The size of each promoter region and relative positions of mapped *cis*-elements in relation to the start codon (ATG) are indicated by the ruler at the top. A colour key of *cis*-elements is given at the bottom of the image. The left hand margin of the figure shows the name and species of each promoter. A-C represents the SCW-associated *CesAs* (4, 7 and 8) and D-F represent the primary cell wall associated *CesAs* (*CesA1*, *3* and *6*). Horizontal black lines in each block represent the promoter sequences for each species and coloured squares show the position of the mapped *cis*-element occurrences found in sense (above line) or anti-sense orientation (below line). Predicted transcriptional start sites (TSSs) are indicated by tailed arrows and transparent grey blocks show the region of the putative TSS-associated *cis*-element cluster.





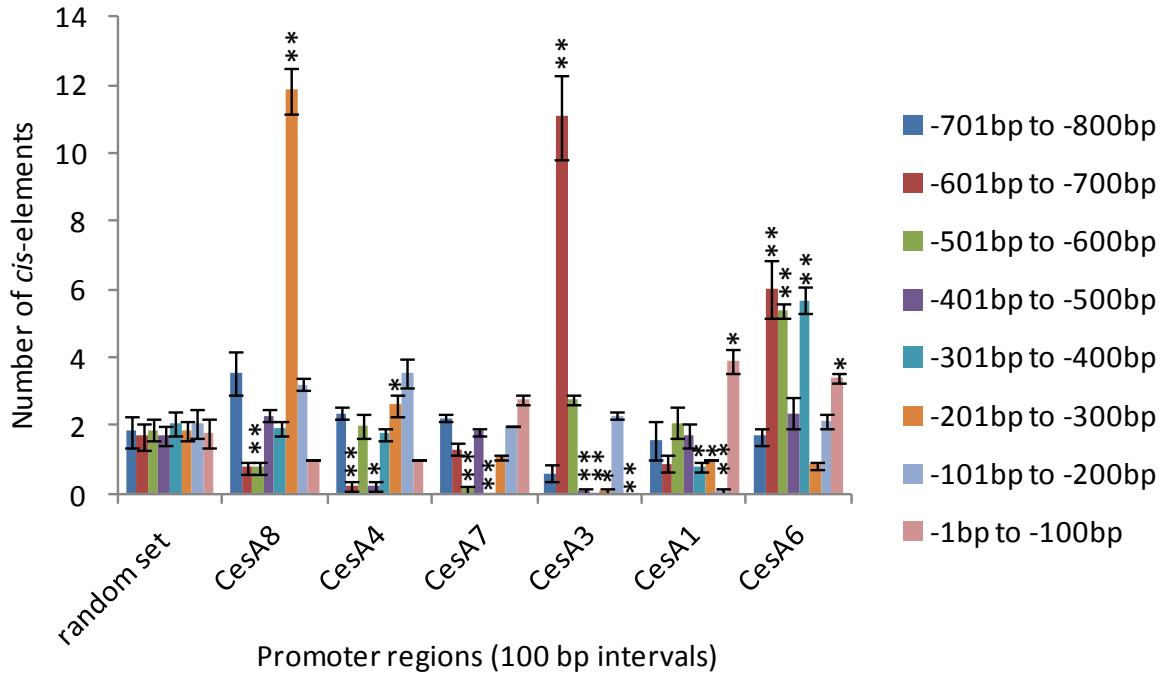


Figure 2.4 Frequency of *cis*-element occurrences along the length of the six *CesA* promoter regions averaged across the 13 *Eucalyptus* species in 100 bp intervals, compared to a randomly generated sequence dataset. The frequency of *cis*-element occurrences in each of the *CesA* promoter sets (*CesA1*, *3*, *4*, *6*, *7* and *8*) are also indicated. The y-axis gives the number of *cis*-elements in each 100 bp interval and the x-axis represents the *CesA* promoter regions being analysed. The error bars show the standard error of the frequency measurement where $n = 13$ (except for *CesA1* where $n = 10$). * and ** indicate significant differences from the random dataset ($p = 0.01$ and 0.001 respectively; two tailed t-test assuming equal variance)

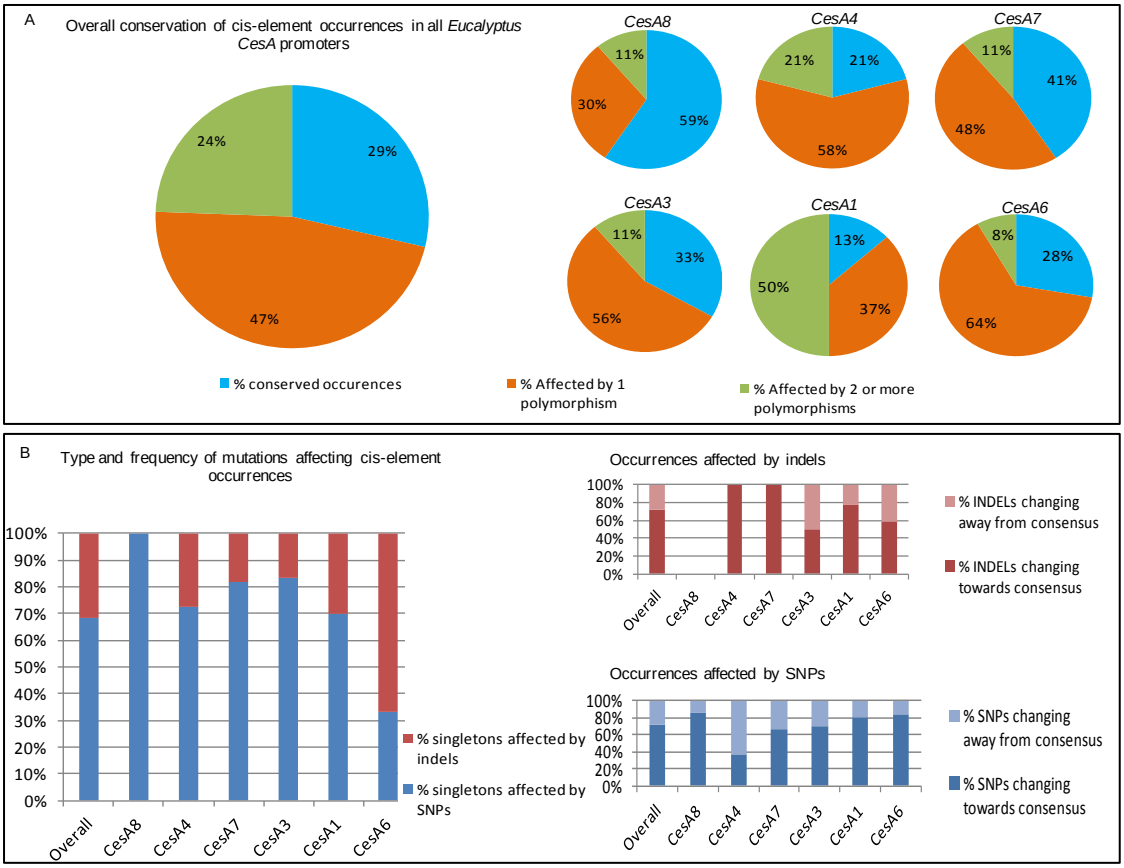


Figure 2.5 Evaluation of the types and frequency of polymorphisms observed within *cis*-element occurrences in the *CesA* promoter regions of 13 *Eucalyptus* tree species. A) Conservation of the *cis*-element consensus sequences across *Eucalyptus* species. The large pie chart (left) depicts the conservation in all promoters across all six genes, while the six smaller pie charts (right) show the *cis*-element sequence conservation within the promoters of each gene (singleton changes were not counted as polymorphisms). B) The frequency of single nucleotide polymorphisms and indels in the promoter regions of 13 *Eucalyptus CesA* promoters. The large bar graph (left) represents the percentage of *cis*-element occurrences affected by SNPs and indels that only affect a single promoter of the 13 species (also referred to as singleton gains or losses). The smaller red bar graph (top right) shows *cis*-element occurrences affected by an indel. The blue bar graph (bottom right) indicates cases where SNPs changed the *cis*-element sequence towards or away from the consensus sequence.

2.10 References

Aspeborg H, Schrader J, Coutinho PM, Stam M, Kallas A, Djerbi S, Nilsson P, Denman S, Amini B, Sterky F, Master E, Sandberg G, Mellerowicz E, Sundberg B, Henrissat B, Teeri TT (2005) Carbohydrate-active enzymes involved in the secondary cell wall biogenesis in hybrid aspen. Plant Physiol. 137: 983-997

Bernard V, Brunaud V, Lecharny A (2010) TC-motifs at the TATA-box expected position in plant genes: a novel class of motifs involved in the transcription regulation. BMC Genomics 11: 166

Blanchette M, Schwikowski B, Tompa M (2002) Algorithms for phylogenetic footprinting. J. Comput. Biol. 9: 211-223

Blanchette M, Tompa M (2002) Discovery of regulatory elements by a computational method for phylogenetic footprinting. Genome Res. 12: 739-748

Blanchette M, Tompa M (2003) FootPrinter: a program designed for phylogenetic footprinting. Nucl. Acids Res. 31: 3840-3842

Brooker MIH (2000) A new classification of the genus *Eucalyptus* L'Her. (Myrtaceae). Aust. Syst. Bot. 13: 79-148

Burn JE, Hocart CH, Birch RJ, Cork AC, Williamson RE (2002) Functional analysis of the cellulose synthase genes *CesA1*, *CesA2*, and *CesA3* in *Arabidopsis*. Plant Physiol. 129: 797-807

Burton RA, Shirley NJ, King BJ, Harvey AJ, Fincher GB (2004) The *CesA* gene family of barley. Quantitative analysis of transcripts reveals two groups of co-expressed genes. Plant Physiol. 134: 224-236

Carmack CS, McCue L, Newberg L, Lawrence C (2007) PhyloScan: identification of transcription factor binding sites using cross-species evidence. Algorithm Mol. Biol. 2: 1

Creux NM, Ranik M, Berger DK, Myburg AA (2008) Comparative analysis of orthologous cellulose synthase promoters from *Arabidopsis*, *Populus* and *Eucalyptus*: evidence of conserved regulatory elements in angiosperms. New Phytol. 179: 722-737

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res. 14: 1188- 1190

Das M, Dai HK (2007) A survey of DNA motif finding algorithms. BMC Bioinformatics 8: S21

de Meaux J, Pop A, Mitchell-Olds T (2006) *Cis*-regulatory evolution of chalcone-synthase expression in the genus *Arabidopsis*. Genetics 174: 2181-2202

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

Desprez T, Juraniec M, Crowell EF, Jouy H, Pochylova Z, Parcy F, Hofte H, Gonneau M, Vernhettes S (2007) Organization of cellulose synthase complexes involved in primary cell wall synthesis in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 104: 15572-15577

Ding J, Hu H, Li X (2012) Thousands of *cis*-regulatory sequence combinations are shared by *Arabidopsis* and poplar. Plant Physiol. 158: 145-155

Eldridge K, Davidson J, Harwood C, Wyk GV (1994) Eucalypt domestication and breeding Oxford University Press, New york

Fang F, Blanchette M (2006) FootPrinter3: phylogenetic footprinting in partially alignable sequences. Nucleic Acids Res. 34: W617-W620

Farnham PJ (2009) Insights from genomic profiling of transcription factors. Nature Rev. Genet. 10: 605-616

Freeling M, Subramaniam S (2009) Conserved noncoding sequences (CNSs) in higher plants. Curr. Opin. Plant Biol. 12: 126-132

Goicoechea M, Lacombe E, Legay S, Mihaljevic S, Rech P, Jauneau A, Lapierre C, Pollet B, Verhaegen D, Chaubet-Gigot N, Grima-Pettenati J (2005) *EgMYB2*, a new transcriptional activator from *Eucalyptus* xylem, regulates secondary cell wall formation and lignin biosynthesis. Plant J. 43: 553-567

Gorshkova T, Brutch N, Chabbert B, Deyholos M, Hayashi T, Lev-Yadun S, Mellerowicz EJ, Morvan C, Neutelings G, Pilate G (2012) Plant fiber formation: state of the art, recent and expected progress, and open questions. CRC Cr. Rev. Plant Sci. 31: 201-228

Grattapaglia D, Plomion C, Kirst M, Sederoff RR (2009) Genomics of growth traits in forest trees. Curr. Opin. Plant Biol. 12: 148-156

Hall A (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser: 95-98

Hamann T, Osborne E, Youngs HL, Misson J, Nussaume L, Somerville C (2004) Global expression analysis of *CESA* and *CSL* genes in *Arabidopsis*. Cellulose 11: 279-286

Hansen L, Marino-Ramirez L, Landsman D (2010) Many sequence-specific chromatin modifying protein-binding motifs show strong positional preferences for potential

regulatory regions in the *Saccharomyces cerevisiae* genome. Nucleic Acids Res. 38: 1772-1779

Hare EE, Peterson BK, Iyer VN, Meier R, Eisen MB (2008) Sepsid *even-skipped* enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. PLoS Genet. 4: e1000106

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

Ho MCW, Johnsen H, Goetz SE, Schiller BJ, Bae E, Tran DA, Shur AS, Allen JM, Rau C, Bender W, Fisher WW, Celniker SE, Drewell RA (2009) Functional evolution of *cis*-regulatory modules at a homeotic gene in *Drosophila*. PLoS Genet. 5: e1000709

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

Karthikeyan AS, Ballachanda DN, Raghothama KG (2009) Promoter deletion analysis elucidates the role of *cis*-elements and 5'UTR intron in spatiotemporal regulation of *AtPht1;4* expression in *Arabidopsis*. Plant Physiol. 136: 10-18

Keohavong P, Thilly WG (1989) Fidelity of DNA polymerases in DNA amplification. Proc. Natl. Acad. Sci. USA 86: 9253-9257

Kim HD, Shay T, O'Shea EK, Regev A (2009) Transcriptional regulatory circuits: predicting numbers from alphabets. Science 325: 429- 432

Ko J-H, Beers E, Han K-H (2006) Global comparative transcriptome analysis identifies gene network regulating secondary xylem development in *Arabidopsis thaliana*. Mol. Genet. Genomics 276: 517-531

Koch MA, Weisshaar B, Kroymann J, Haubold B, Mitchell-Olds T (2001) Comparative genomics and regulatory evolution: conservation and function of the *Chs* and *Apetala3* promoters. Mol. Biol. Evol. 18: 1882-1891

Kulheim C, Hui Yeoh S, Maintz J, Foley W, Moran G (2009) Comparative SNP diversity among four *Eucalyptus* species for genes from secondary metabolite biosynthetic pathways. BMC Genomics 10: 452

Kumar M, Thammannagowda S, Bulone V, Chiang V, Han K-H, Joshi CP, Mansfield SD, Mellerowicz E, Sundberg B, Teeri T, Ellis BE (2009) An update on the nomenclature for the cellulose synthase genes in *Populus*. Trends Plant Sci. 14: 248-254

Legay S, Sivadon P, Blervacq A-S, Pavy N, Baghdady A, Tremblay L, Levasseur C, Ladouce N, Lapierre C, Séguin A, Hawkins S, Mackay J, Grima-Pettenati J (2010) *EgMYB1*, an

R2R3 MYB transcription factor from *Eucalyptus* negatively regulates secondary cell wall formation in *Arabidopsis* and poplar. New Phytol. 188: 774-786

Ling LL, Keohavong P, Dias C, Thilly WG (1991) Optimization of the polymerase chain reaction with regard to fidelity: modified T7, Taq, and vent DNA polymerases. Genome Res 1: 63-69

Livny J, Waldor MK (2009) Mining regulatory 5'UTRs from cDNA deep sequencing datasets. Nucleic Acids Res. 38: 1504- 1514

Love J, Björklunda S, Vahala J, Hertzberg M, Kangasjärvib J, Sundberg B (2009) Ethylene is an endogenous stimulator of cell division in the cambial meristem of *Populus*. Proc. Natl. Acad. Sci. USA 106: 5984-5989

Lu S, Li L, Yi X, Joshi CP, Chiang VL (2008) Differential expression of three *Eucalyptus* secondary cell wall-related cellulose synthase genes in response to tension stress. J. Exp. Bot. 59: 681-695

Maniatis T, Goodbourn S, Fischer JA (1987) Regulation of inducible and tissue-specific gene expression. Science 236: 1237-1245

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

Mellerowicz EJ, Sundberg B (2008) Wood cell walls: biosynthesis, developmental dynamics and their implications for wood properties. Curr. Opin. Plant Biol. 11: 293-300

Mizrachi E, Hefer C, Ranik M, Joubert F, Myburg A (2010) *De novo* assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by Illumina mRNA-Seq. BMC Genomics 11: 681

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

Mutwil M, Debolt S, Persson S (2008) Cellulose synthesis: a complex complex. Curr. Opin. Plant Biol. 11: 252-257

Neale DB, Ingvarsson PK (2008) Population, quantitative and comparative genomics of adaptation in forest trees. Curr. Opin. Plant Biol. 11(2):149-155.

Nei M (2007) The new mutation theory of phenotypic evolution. Proc. Natl. Acad. Sci. USA 104: 12235-12242

Nei M, Li W (1979) Mathematical model for studying genetic variation in term of restriction endonucleases. Proc. Natl. Acad. Sci. USA 76: 5267-5273

Novaes E, Drost D, Farmerie W, Pappas G, Grattapaglia D, Sederoff R, Kirst M (2008) High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. BMC Genomics 9: 312

Pauli S, Rothnie HM, Chen G, He X, Hohn T (2004) The cauliflower mosaic virus 35S promoter extends into the transcribed region. J. Virol. 78: 12120-12128

Payn K, Dvorak W, Janse B, Myburg A (2008) Microsatellite diversity and genetic structure of the commercially important tropical tree species *Eucalyptus urophylla*, endemic to seven islands in eastern Indonesia. Tree Genet. Genomes 4: 519-530

Piao SL, Fang JY, Ciais P, Peylin P, Huang Y, Sitch S, Wang T (2009) The carbon balance of terrestrial ecosystems in China. Nature 458: 1009-1013

Popper ZA, Michel G, Herve C, Domozych DS, Willats WGT, Tuohy MG, Kloareg B, Stengel DB (2011) Evolution and diversity of plant cell walls: from algae to flowering plants. Annu. Rev. Plant Biol. 62: 567-590

Pryor LD, Johnson LAS (1971) A classification of the Eucalypts. Australian National University Press, Canberra

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

Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ, Jr., Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T (2006) The path forward for biofuels and biomaterials. Science 311: 484-489

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

Rathmann R, Szklo A, Schaeffer R (2010) Land use competition for production of food and liquid biofuels: An analysis of the arguments in the current debate. Renew. Energ. 35: 14-22

Regalbuto JR (2009) Cellulosic biofuels: Got gasoline? Science 325: 822- 824

Reineke AR, Bornberg-Bauer E, Gu J (2011) Evolutionary divergence and limits of conserved non-coding sequence detection in plant genomes. Nucleic Acids Res. 39: 6029-6043

Roberts AW, Bushoven JT (2007) The cellulose synthase (CESA) gene superfamily of the moss *Physcomitrella patens*. Plant Mol. Biol. 63: 207 - 219

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19: 2496-2497

Samuga A, Joshi CP (2004) Differential expression patterns of two new primary cell wall-related cellulose synthase cDNAs, *PtrCesA6* and *PtrCesA7* from aspen trees. Gene 334: 73-82

Sarkar P, Bosneaga E, Auer M (2009) Plant cell walls throughout evolution: towards a molecular understanding of their design principles. J. Exp. Bot. 60: 3615-3635

Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 18: 6097-6100

Shi R, Sun Y-H, Li Q, Heber S, Sederoff R, Chiang VL (2010) Towards a systems approach for lignin biosynthesis in *Populus trichocarpa*: transcript abundance and specificity of the monolignol biosynthetic genes. Plant Cell Physiol. 51: 144-163

Steane DA, McKinnon GE, Vaillancourt RE, Potts BM (1999) ITS sequence data resolve higher level relationships among the eucalypts. Mol. Phylogenet. Evol. 12: 215-223

Steane DA, Nicolle D, Sansaloni CP, Petroli CD, Carling J, Kilian A, Myburg AA, Grattagalia D, Vaillancourt RE (2011) Population genetic analysis and phylogeny reconstruction in *Eucalyptus* (Myrtaceae) using high-throughput, genome-wide genotyping. Mol. Phylogenet. Evol. 59: 206-224

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28: 2731-2739

Tanaka K, Murata K, Yamazaki M, Onosato K, Miyao A, Hirochika H (2003) Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall. Plant Physiol. 133: 73-83

Tanaka T, Koyanagi KO, Itoh T (2009) Highly diversified molecular evolution of downstream transcription start sites in rice and *Arabidopsis*. Plant Physiol. 149: 1316 - 1324

Tanay A, Regev A, Shamir R (2005) Conservation and evolvability in regulatory networks: The evolution of ribosomal regulation in yeast. Proc. Natl. Acad. Sci. USA 102: 7203-7208

Taylor NG (2008) Cellulose biosynthesis and deposition in higher plants. New Phytol 178: 239-252

Taylor NG, Gardiner JC, Whiteman R, Turner SR (2004) Cellulose synthesis in the *Arabidopsis* secondary cell wall. Cellulose 11: 329-338

Taylor NG, Howells RM, Huttly AK, Vickers K, Turner SR (2003) Interactions among three distinct CESA proteins essential for cellulose synthesis. Proc. Natl. Acad. Sci. USA 100: 1450-1455

Than C, Ruths D, Nakhleh L (2008) PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. BMC Bioinformatics 9: 322

Thomas-Chollier M, Sand O, Turatsinze J-V, Janky Rs, Defrance M, Vervisch E, Brohee S, van Helden J (2008) RSAT: regulatory sequence analysis tools. Nucleic Acids Res. 36: W119- W127

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22: 4673-4680

Tjaden G, Edwards JW, Coruzzi GM (1995) *Cis*-elements and trans-acting factors affecting regulation of a nonphotosynthetic light-regulated gene for chloroplast glutamine synthetase. Plant Physiol. 108: 1109-1117

Tompa M, Li N, Bailey TL, Church GM, De Moor B, Eskin E, Favorov AV, Frith MC, Fu Y, Kent WJ, Makeev VJ, Mironov AA, Noble WS, Pavesi G, Pesole G, Regnier M, Simonis N, Sinha S, Thijs G, van Helden J, Vandenbogaert M, Weng Z, Workman C, Ye C, Zhu Z (2005) Assessing computational tools for the discovery of transcription factor binding sites. Nat. Biotechnol. 23: 137-144

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

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. 7: 256-276

Wijaya E, Yiu S-M, Son NT, Kanagasabai R, Sung W-K (2008) MotifVoter: a novel ensemble method for fine-grained integration of generic motif finders. Bioinformatics 24: 2288-2295

Wu A-M, Hu J, Liu J-Y (2009) Functional analysis of a cotton cellulose synthase A4 gene promoter in transgenic tobacco plants. Plant Cell Rep. 28: 1539-1548

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

Yazaki J, Kishimoto N, Nagata Y, Ishikawa M, Fujii F, Hashimoto A, Shimbo K, Shimatani Z, Kojima K, Suzuki K, Yamamoto M, Honda S, Endo A, Yoshida Y, Sato Y, Takeuchi

K, Toyoshima K, Miyamoto C, Wu J, Sasaki T, Sakata K, Yamamoto K, Iba K, Oda T, Otomo Y, Murakami K, Matsubara K, Kawai J, Carninci P, Hayashizaki Y, Kikuchi S (2003) Genomics approach to abscisic acid- and gibberellin-responsive genes in rice. DNA Res. 10: 249-261

Yin Y, Huang J, Xu Y (2009) The cellulose synthase superfamily in fully sequenced plants and algae. BMC Plant Biol. 9: 99

Zhao L, Lu L, Zhang L, Wang A, Wang N, Liang Z, Lu X, Tang K (2009) Molecular evolution of the E8 promoter in tomato and some of its relative wild species. J. Biosci. 34: 71-83

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

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

Chapter 3

Eucalyptus CesA8 promoter regulatory model revealed by promoter deletion and reporter gene expression in *Eucalyptus* and *Arabidopsis*

N.M. Creux[1,2], A. Spokevicius[3], G. Bossinger[3], C. Maritz-Olivier[1], A.A. Myburg[1,2]

Department of Genetics[1], Faculty of Natural and Agricultural Sciences, Forestry Argicultural Biotechnology Institute (FABI)[2], University of Pretoria, Pretoria, 2000, South Africa

Department of Forest and Ecosystem Science[3], Melbourne School of Land and Environment, University of Melbourne, Creswick, VIC, Australia

This chapter was prepared in the format of a manuscript for a research journal (e.g. *Planta*). I performed all of the vector constructions, *cis*-element analysis, *Arabidopsis* transformations, selections and GUS assays. I worked in collaboration with Prof. Gerd Bossinger and Dr. Antanas Spokevicius at the University of Melbourne, Australia to optimize the induced somatic sector analysis (ISSA) for promoter::GUS constructs using the full-length *Eucalyptus CesA* promoters as a case study (Appendix A). They also aided in the ISSA analysis of the *EgrCesA8* promoter truncations. Prof. Zander Myburg and Dr. Christine Maritz-Olivier provided advice, direction, critical revision and supervision in planning of the project. All other technical assistance is acknowledged at the end of the chapter. Appendix A appears exactly as it was published in the *Planta Journal* and therefore contains the older previously published *Eucalyptus CesA* gene names; please refer to Table 1.1 for new naming convention proposed for these genes.

3.1 Abstract

The process of cellulose synthesis and deposition in plants has been the focus of a number of studies due its high commercial value. Cellulose is deposited into the plant cell wall by large membrane bound protein complexes consisting of different cellulose synthase (CesA) subunits. The composition of the CesA protein complex varies depending on whether the primary or secondary cell wall is being deposited. In all angiosperms three *CesA* genes are associated with secondary cell wall (SCW) formation while a range of different *CesAs* are involved in primary cell wall formation. The regulation of the *CesA* genes associated with primary or secondary cell wall formation is currently not well characterized. In this study we used promoter::GUS truncation analysis to dissect the SCW *EgrCesA8* promoter and assess various regions for reporter gene expression in *Eucalyptus*, *Populus* and *Arabidopsis*. An *in silico cis*-element analysis identified clusters of known SCW-associated *cis*-elements in regions modulating GUS expression. A cluster of SCW-associated *cis*-elements was located in a region displaying tissue-specific expression and could represent an important *cis*-regulatory module (CRM) in this promoter. A $CT_{(11)}$-microsatellite at the transcriptional start site activated strong non-specific expression in *Arabidopsis* and *Eucalyptus* suggesting it is a key element in the core promoter of *EgrCesA8*. These findings shed light on the regulatory mechanisms governing the SCW-related *EgCesA8* gene expression in *Eucalyptus*.

Keywords: Cellulose synthase, Promoter, *β*-glucuronidase, *cis*-element, *cis*-regulatory module (CRM), Wood formation

3.2 Introduction

Cellulose is a major commodity used for many different industrial applications including food additives, lignocellulosic biofuels, wood and paper products. One of the major sources of cellulose pulp is *Eucalyptus* plantations producing 40% of all hardwood pulp per annum (http://www.eucalyptusfacts.org/). Cellulose is found in all plant cell walls where the bulk is deposited in the secondary cell walls of fiber and vessel cells, which are the major cell type in woody angiosperm stems (Plomion et al. 2001). Previous studies have identified between 8 and 18 different cellulose synthase (*CesA*) genes in many distantly related plant species (Hamann et al. 2004; Roberts and Bushoven 2007; Kumar et al. 2009; Yin et al. 2009). Regardless of the evolutionary history of these plants there is strong conservation of *CesA* gene expression patterns with orthologs of *CesA4*, *7* and *8* being consistently expressed in tissues forming secondary cell walls (Burton et al. 2004; Hamann et al. 2004; Samuga and Joshi 2004; Ranik and Myburg 2006). While many studies have investigated the genes involved in cellulose deposition and the process by which this occurs, much remains to be elucidated about the mechanisms regulating these important genes.

Gene regulation is a complex process that occurs on many different levels from basic transcription factor-DNA interactions to less well described mechanisms such as chromatin structure and chromosome interactions (reviewed in Lelli et al. 2012). Much of the plant research to date deals with the identification of transcription factors involved in gene regulation under different conditions or at different developmental states and tissues (Jiao et al. 2007; Nakashima et al. 2009; Kaufmann et al. 2010). A number of transcription factors have been implicated in secondary cell wall (SCW) formation and cellulose deposition in different plant species including *Eucalyptus*, *Populus* and *Arabidopsis* (Reviewed in Demura and Ye 2010; Zhong et al. 2010a). This SCW regulatory network comprises transcription factors with known DNA binding domains, mainly from the NAC (NAM/ATAF/CUC) and

MYB (MYELOBLASTOSIS) transcription factor families. While many studies have focused on the proteins involved in the SCW regulatory network there is a distinct lack of information on the actual DNA binding sites of these important proteins in SCW genes.

The SCW regulatory network contains many different transcription factors that have direct evidence for DNA binding, but few of these have been linked to a conserved *cis*-regulatory element. To date, only the SNBE and M46RE *cis*-elements have been identified in the SCW-related *AtCesA8* promoter (Zhong et al. 2010b; Kim et al. 2012; Zhong and Ye 2012). However, Hazen et al. (2010) suggest from their yeast-1-hybrid results that over 30 different transcription factors are expected to bind to the *AtCesA8* promoter and therefore many more *cis*-element sites are anticipated. There are only a few studies that have investigated the functional and sequence conservation of these binding sites in woody plant species such as *Eucalyptus* and *Populus* (Creux et al. 2008; Winzell et al. 2010; Zhong et al. 2011; Ding et al. 2012 and Chapter 2). Goicoechea et al. (2005) identified an MBSIIG element in the promoters of lignin biosynthetic genes of *Eucalyptus* which interacted with *Eg*MYB2. Similarly, the *Populus* ortholog of *Eg*MYB2 (*Pt*MYB21) interacts with the ACTYP element identified in the xylan biosynthetic genes (Winzell et al. 2010). Both the ACTYP and MBSIIG elements share sequence similarity with the M46RE sequence identified in *Arabidopsis* (Kim et al. 2012, Zhong et al. 2012).

Currently identifying and testing *cis*-regulatory elements is a difficult and laborious task and as such has contributed to the limited *cis*-element information available for the SCW regulatory network. The main approaches currently being used to identify and characterize *cis*-regulatory elements are *in silico* analysis and *in vitro* or *in vivo* analysis. *In silico* analysis is based on a number of different algorithms which identify short conserved sequences in large non-coding DNA datasets (reviewed in Altobelli 2012). The *cis*-element identification algorithms use co-expression data, positional data and sequence conservation data to identify

putative *cis*-elements in promoter regions. It is well known that *cis*-elements are found close to each other in groups or clusters along the length of the promoter. Ding et al. (2012) developed a search algorithm based on *cis*-element clustering and *cis*-element conservation across species to identify putative elements associated with SCW formation. However, the short sequence length of a typical plant *cis*-element and the degenerate nature of *cis*-regulatory motifs are difficult to overcome during *in silico* analyses and so the results often contain a number of false positives (Van Loo and Marynen 2009). It is predicted that as more sequence data is obtained and a deeper understanding of transcriptional regulation is reached the accuracy of the algorithms will be improved.

Many *in vitro* and *in vivo* techniques such as promoter deletion studies, reporter gene expression studies, elecrotphoretic mobility shift assays (EMSA), yeast-1-hybrid (Y1H) assays, protein pull-down assays and more recently various high-throughput sequencing methods have been developed to identify *cis*-regulatory elements and analyze gene promoter function (reviewed in Stormo and Zhao 2010; Hughes 2011). In plants many of the molecular techniques employed when identifying DNA-protein interactions are low throughput or are still in the optimization phase. A technique which is often employed is promoter::reporter gene analysis in native or model species. Here the promoter region of interest is inserted upstream of a reporter gene in a plasmid and is stably transformed into the model organism of choice (Rosellini 2012). A frequently used reporter gene is *β*-glucuronidase (GUS), which interacts with the 5-bromo4-chloro-3-indoyl *β*-D-glucuronide (X-Gluc) substrate to produce a blue colour which can be tracked in different plant tissues (Jefferson et al. 1987; Alwen et al. 1992). Reporter gene analysis proves more difficult in non-model plants such as *Eucalyptus* which are often recalcitrant to transformation. Previously we optimized a method to monitor GUS expression directly in the woody tissues of *Eucalyptus* using induced somatic sector

analysis (ISSA; Van Beveren et al. 2006, Appendix A), which helps to extend the use of reporter genes in woody tissues.

Eucalyptus is an economically important species synthesizing vast quantities of cellulose. With the completion of the *Eucalyptus* genome sequence it is fast becoming a good model for the investigation of SCW deposition and cellulose synthesis. However, to fully utilize the current genome sequence and to perform future comparative genome studies it is important to identify and validate different regulatory regions, including those involved in the modulation of SCW deposition. Using the *Eucalyptus cellulose synthase 8* gene promoter as a case study *(EgrCesA8)*, we addressed three main questions. Firstly, which of the known SCW-associated *cis*-elements occur in the *EgrCesA8* promoter? Secondly, which regions of the *EgrCesA8* promoter regulate reporter gene expression? And lastly, do any of the known SCW *cis*-elements occur within regions of the *EgrCesA8* promoter, which modulate reporter gene expression? In this study SCW-specific *cis*-elements were mapped to the *EgrCesA8* promoter and a tight cluster of *cis*-element occurrences was identified between -900 bp and -700 bp of the promoter. Promoter truncation analysis indicated that this region activated tissue-specific reporter gene expression, while a region between -1500 bp and -1200 bp appeared to repress gene expression. A $CT_{(11)}$-microsatellite at the transcriptional start site (TSS) directed strong non-specific expression of the GUS reporter gene in both *Arabidopsis* and *Eucalyptus*. Overall, a series of activators, repressors and tissues-specific elements were identified in the *EgrCesA8* promoter, which will lay the foundations for further studies on how cellulose deposition is regulated and conserved in this important forest tree species.

3.3 Materials and Methods

3.3.1 Cis-element mapping and analysis

Previous studies performed detailed *in silico* analyses on the conserved sequences and putative *cis*-elements harboured within the *EgrCesA8* and other *Eucalyptus CesA* promoters (Creux et al. 2008, Chapter 2). We compiled a list (Table 3.1) of the conserved *cis*-elements identified in these studies and supplemented the list with verified *cis*-elements from the literature (Goicoechea et al. 2005; Ko et al. 2006; Pyo et al. 2007; Winzell et al. 2010; Zhong et al. 2010b; Ding et al. 2012; Kim et al. 2012; Zhong and Ye 2012) and TRANSFAC database (http://www.gene-regulation.com). Ding et al. (2012) identified a number of *cis*-elements from the PLACE database (http://www.dna.affrc.go.jp/PLACE/; Higo et al. 1999) that occurred in clusters in *Arabidopsis* and *Populus* promoters. The *cis*-element composition of the cluster was dependent on the type of promoters being analysed, for example the *Arabidopsis* and *Populus* promoters of genes related to cellulose deposition all shared a similar cluster of *cis*-elements. In the current study, the datasets generated by Ding et al. (2012) were used to identify the *cis*-elements from the cellulose-associated or SCW-associated *cis*-element clusters that were also highly abundant (≥ five occurrences per promoter set) in the *Arabidopsis* and *Populus* promoters of cellulose biosynthetic genes. The highly abundant *cis*-elements that were also part of the cellulose-associated and SCW-associated clusters were included in the list. All of the *cis*-element consensus sequences (Table 3.1) were used in a pattern matching search with default settings (no mismatches allowed) on the Regulatory Sequence Analysis (RSA) Tools website (http://rsat.ulb.ac.be/; Thomas-Chollier et al. 2008) in order to identify occurrences in the *EgrCesA8* promoter and to map the *cis*-element positions within the promoter (Figure 3.1).

3.3.2 Promoter truncation

The *EgrCesA8* promoter was isolated by genome walking and cloned into the pMDC162 vector (Invitrogen, CA, USA) in a previous study (Creux et al. 2008). The original 2 kbp promoter fragment was considered to be the full-length promoter. From this full-length promoter construct five regions of incrementally smaller lengths from the start codon (1411 bp, 1036 bp, 853 bp, 588 bp, and 334 bp) were PCR amplified and used to construct the truncated *EgrCesA8* promoter::GUS reporter vectors. In Chapter 1, a TC-repeat element was identified at the transcriptional start site of the *EgrCesA8* promoter and two vectors were constructed to investigate the function of this repetitive region. The first construct contained the -228 bp 5' UTR (from the start codon to the transcriptional start site; Creux et al. 2008) and the second construct contained the 5' UTR with the TC-repeat element attached to it at the 5' end in the native position. All the promoter truncations of the *EgrCesA8* promoter were produced by PCR amplification where forward primers were designed at specific intervals along the promoter and a single reverse primer was situated at the translation start site (Figures 3.2, 3.3 and Table 3.2).

3.3.3 Sequence analysis

The *Eucalyptus grandis* SCW-associated *CesA* (*EgrCesA3*) promoter sequence reported by Lu et al. (2008) was obtained from NCBI (accession number: EU165713). A nucleotide blast search was performed to investigate sequence similarity to the sequence for the *EgrCesA8* promoter (accession number: EU737100) reported in Creux et al. (2008). Vector NTI (Invitrogen, CA, USA) was used to generate *in silico* versions of the promoter deletions and truncations studied by Lu et al. (2008). Using the vector NTI alignment tool (Invitrogen, CA, USA) the Lu et al. (2008) promoter fragments were compared to the fragments generated for

EgrCesA8 in the current study. Constructs that were similar or different between the two studies were identified and this made possible the comparison of the results from the two studies (Figure 3.1). All DNA sequences were analyzed and managed using Vector NTI (Invitrogen, CA, USA)

3.3.4 Plasmid construction

All promoter truncations were PCR amplified (Primers: TableS3.1) using high fidelity Takara Ex Taq DNA polymerase (Takara, Shiga, Japan) and T/A cloned into the GW/8PCRTOPO entry vector (Invitrogen, CA, USA) which were chemically transformed into DH5*α* *E. coli* cells. The sequence and orientation of promoter fragments were confirmed by colony PCR and plasmid sequencing (Macrogen, Seoul, Korea). Colony PCR was performed by selecting a colony and re-suspending it in 20 µl SABAX water (Adcock-Ingram, Midrand, South Africa) which was then incubated at 95°C for 5 min. The suspension was briefly centrifuged and 5 µl was used as template for amplification. M13 forward primer (5'CACGACGTTGTAAAACGAC3') was used in the PCR with either the forward (Table 3.1) or the reverse (*EgrCesA8P*_ATG_reverse) gene-specific primer to determine orientation. Promoter fragments were recombined into the pMDC162 destination vector (Invitrogen, CA, USA) using Clonase (Invitrogen, CA, USA) and chemically transformed into DH5*α* cells using standard transformation protocols. Five clones of each truncate were sequenced (Macrogen, Seoul, Korea). After sequence confirmation the destination vectors harbouring different promoter truncations were chemically transformed into *Agrobacterium* strain LBA4404 by a standard transformation protocol described in the *Arabidopsis* handbook (Weigel and Glazebrook 2002).

3.3.5 Arabidopsis transformation and selection

Columbia-O (Col-O) wild type *Arabidopsis* plants were transformed by a standard floral dipping method into an *Agrobacterium* sucrose solution (*Arabidopsis* handbook) and seeds were harvested (Weigel and Glazebrook 2002). Seeds were surface sterilized and sown on soft plant agar containing 1 mg/ml ceflotaxamine (Austell, JHB South Africa) and 50 mg/ml HyClone hygromycin B solution (Utah USA) in 90 mm petridishes. Five to eight independent transformed lines for each construct (2 kb, 1411 bp, 1036 bp, 853 bp, 588 bp, 334 bp, TC-UTR and UTR) were selected and transplanted to peat moss bags (Jiffy Products InternationalAS, Norway). Positive transformants were grown and allowed to produce seed that was harvested. The harvested seed was re-planted and seed was harvested again, this was repeated 2 times to bulk seed stocks before the experimental trial was initiated on T3 seed stocks. All plants were grown under standard long day conditions where light was maintained for 16 h a day using artificial light at a temperature of 22°C and 75% humidity (Hussey et al. 2011). At least five independent transgenic lines were selected for each of the *EgrCesA8* promoter truncation constructs, the 35S positive control construct and the full-length (2018 bp upstream ATG) *EgrCesA8* promoter.

3.3.6 PCR confirmation of positive Arabidopsis transformants

Transgenic and wild type leaves were harvested from the *Arabidopsis* plants during the 5th week of growth and stored at -20°C until DNA was extracted with the Nucleospin plant extract II kit (Machery-Nagel GMBH, Duren, Germany). Each independent *Arabidopsis* line was tested for insert by PCR analysis. To test for insertion of constructs -2018, -1411, -1036, -853, -588 and -340 a reverse GUS-specific primer (B_R) and a -334 specific primer (D_F) was used (Table 3.2 and Figure 3.3). Similarly, to test for the 35S::GUS construct the same GUS

reverse primer was used with a 35S-specific forward primer (E_F; Table 3.2 and Figure 3.3). For the UTR construct, a 5'UTR-specific primer (C_F) was used with the GUS specific primer. The TC-UTR construct could not be amplified with the GUS-specific primer and therefore forward and reverse TC-UTR-specific primers (G_F and F_R) were used (Table 3.2 and Figure 3.3). A pair of primers (A_F and A_R) which amplified a region in the endogenous *AtCesA7* of *Arabidopsis* was used as positive PCR control. The insert-specific PCR and the positive control PCR were performed as a multiplex PCR to increase throughput. Amplification was performed using ExSel taq (Jain Biologicals, Haryana, India) and the amplification protocol was as follows: Denaturation at 95°C for 2 min, annealing at 58°C for 30 sec, elongation at 72°C for 90 sec and a final extension step at 72°C for 10 min. Products were resolved on a 1% agarose gel stained with ethidium bromide (Sigma, MO, USA) and visualized under UV with the Gel Doc XR+ imaging system (BioRad, CA, USA).

3.3.7 Arabidopsis harvesting and GUS staining

Three different tissues (roots, hypocotyl and leaves) were harvested from three different plants for each of 4-5 independently transformed lines (Supplementary files 3.1 and 3.2). Plants were six weeks of age when harvested for reporter gene analysis. GUS staining was conducted as described by Van Beveren et al. (2006). Tissues were de-stained by multiple wash steps in absolute ethanol and then stored in 70% ethanol for further analysis (Weigel and Glazebrook 2002).

3.3.8 GUS histochemical assays

Arabidopsis root and leaf tissues were imaged with a digital Cannon EOS camera and a 60 mm macro lens (Cannon, Tokyo, Japan). The hypocotyls were removed from 70% ethanol, briefly dried on absorbent paper and placed in a petridish on a white paper background. After imaging, hypocotyls and roots were returned to 70% ethanol for storage. The leaf tissue was first placed on a slide with a cover slip suspended in 70% ethanol. Various shades of grey and white paper backgrounds were used to obtain the correct contrast so that all GUS stained regions, including fine veins, were visible.

3.3.9 Induced Somatic Sector Analysis

ISSA (Induced Somatic Sector Analysis) was performed as described in Van Beveren et al. (2006) and Appendix A. Briefly, three-month-old *Eucalyptus* clones (*Eucalyptus camaldulensis x globulus* and *Eucalyptus camaldulensis x grandis*) were potted in premium potting mix. In the greenhouse the day temperature was maintained at 21-25°C and the night temperature between 14°C and 17°C with a photoperiod of 16 hours. The same constructs generated for the analysis of GUS expression in the *Arabidopsis* plants were used to transform the AGL1 *Agrobacterium* strain (Lazo et al. 1991) by electroporation as described in Sambrook and Russell (2001). Bacteria were grown for 48 hours at 28°C in LB medium containing 25 µg/ml rifampicin and 50 µg/ml kanamycin (Sigma, Mo, USA), diluted 1:20 with fresh LB and grown to OD_{600} of 0.4 to 0.6 after which the cells were recovered by centrifugation (1150 x *g* for 15 min) and re-suspended in 1 ml of Murashige–Skoog (MS) media prior to inoculation (Van Beveren et al. 2006 and Appendix A). Each 1 cm^2 cambial window was inoculated with 5 µl of *Agrobacterium* suspension (containing one of the

EgrCesA8 promoter truncation constructs or control under investigation; one promoter per window). Stem sections harbouring cambial windows were excised from the main stem during harvesting (Van Beveren et al. 2006). Un-inoculated stem tissue (outside the window area) was removed and the remaining cambial tissue was cut transversely into 1 mm half-discs. Cambial discs were incubated in a water bath in GUS solution at 55°C for 10 minutes prior to being placed upright in the dark on a rotary shaker (150 rpm) at 37°C overnight. The GUS solution was then replaced with 70% ethanol and samples were stored at 4°C (Van Beveren et al. 2006, Appendix A).

3.4 Results

3.4.1 SCW-associated cis-element clusters identified in the Eucalyptus cellulose synthase 8 (CesA8) promoter

A list of previously identified *cis*-elements was compiled from the literature (Table 3.1). Out of the 55 elements investigated only 16 (29%) elements mapped to the *EgrCesA8* promoter with a high stringency (no miss-matches from the consensus). Of the elements that mapped to the *EgrCesA8* promoter (Figure 3.1), five (MYB1AT, NODCON1GM, REALPHALGLHCB21, CTRMCAMV35S and CRPE31) have previously been mapped to this promoter (Creux et al. 2008 and Chapter 2). Seven of the *cis*-elements (BPC1, CBNAC, C1, DOF, DREB1B, PEND class and PBF-PBF) were identified using the *EgrCesA8* promoter sequence to search the TRANSFAC database for experimentally verified elements. None of the latter has any information in terms of their involvement in SCW formation or cellulose deposition. The final four elements (SNBE, ACTYP, MRNA3ENDTAH3 and M46RE) were identified in the promoters of genes shown to be involved in SCW formation or cellulose deposition (Goicoechea et al. 2005; Winzell et al. 2010; Zhong et al. 2010b; Ding

et al. 2012; Kim et al. 2012). A total of 44 *cis*-element occurrences were counted. The mapped *cis*-element occurrences in the *EgrCesA8* promoter were not evenly distributed throughout the promoter but clustered in discrete regions (Figure 3.1; e.g. UTR, -300 to -500, -700 to -900 and -1400 to -1700). In regions of *cis*-element clusters a *cis*-element occurred every 20 bp, while in regions with a low number of *cis*-element occurrences a *cis*-element site only occurred every 100 bp (e.g. region -1000 to -1200). These findings are similar to the clustering observed when *cis*-elements were mapped to the *CesA8* promoter sequences from 13 different *Eucalyptus* species (Chapter 2).

3.4.2 Eucalyptus CesA8 promoter truncations show variation of reporter gene expression across different tissues in Arabidopsis

Promoter truncation analysis was performed to identify regions within the *EgrCesA8* promoter that may influence the gene's expression. Promoter truncations were generated based on extensive *cis*-element maps (Creux et al. 2008, Chapter 2 and Figure 3.1). Lu et al. (2008) previously performed truncation studies on this promoter sequence from a different *E.grandis* genotype. Using these results and the *cis*-element data as a guideline, we could select regions in the promoter that were not yet investigated and produced a series of promoter truncations by PCR amplification (Figure 3.2 and Table 3.2). This allowed for comparison of current results to those of Lu et al. (2008), and the position of previously identified functional regions in the promoter could be more accurately described.

A multiplex PCR was performed to confirm that the transgenic *Arabidopsis* lines contained each of the *EgrCesA8* promoter truncates and the GUS reporter gene (Table 3.2 and Figure 3.3). Only one fragment (*AtCesA7* positive marker) was observed in wild-type negative control plants (Figure 3.4; lanes 1-3) because they do not carry a promoter-GUS

construct and thus cannot be amplified by the construct-specific primers (Figure 3.3). Lanes where the 450 bp fragment (*AtCesA7* positive marker) was absent indicated a failed PCR reaction (Figure 3.4; lanes 35, 36 and 51) and these were repeated (data no shown). PCR negative control reactions also have no bands indicating no contamination of reagents (Figure 3.4 lane 136).

The second set of primers in each reaction was specific for the promoter fragment being amplified (Figure 3.3). In most cases (-2018, -1411, -1036, -853, -588 and -344) a GUS-specific reverse primer (B_R) and the *EgrCesA8P*_340_forward (D_F) were used to amplify across part of the promoter region and the GUS gene which produced a fragment of 2300 bp (Figure 3.3). Lanes that contained both the -450 bp and -2300 bp fragments indicated plants that were positive for the *EgrCesA8* promoter truncate and the GUS gene in the correct orientation (Figure 3.4; lanes 49, 50, 52-75, 77, 79-113, 115-135). Similarly, for the positive CAMV 35S promoter, the B_F primer was used in conjunction with a CAMV 35S-specific forward primer (E_F) and resulted in a fragment of 2500 bp (Figure 3.3). Lanes that contained a 2500 bp and 450 bp fragment indicated plants that were positive for the CAMV 35S promoter and GUS gene (Figure 3.4; lanes 4-16 and 18). The *EgrCesA8* UTR promoter fragment was shorter than -340 and therefore the D_F primer could not be used, instead the *EgrCesA8*_UTR (C_F) forward primer and the B_R primer were used to amplify a 2280 bp fragment from plants containing the UTR-GUS construct (Figure 3.3 and Figure 3.4; lanes 19-28). The TC-UTR-specific forward primer (*EgrCesA8*_UTR-TC_forward, Figure 3.3; G_F) was not compatible with the GUS-specific reverse primer (B_F) and therefore we used the TC-UTR specific reverse primer (*EgrCesA8P*_ATG, Figure 3.3; F_R) that produced a 300 bp fragment (Figure 3.3). Plants which were positive for the TC-UTR construct produced the -450 bp fragment and just below this the 300 bp TC-UTR-specific fragment (Figure 3.4 lanes 34, 37-48). Some plants only showed amplification of the 450 bp fragment even though they

were selected for on hygromycin selective media as positive transformants, indicating these plants may not contain the promoter::GUS construct and cannot be analyzed further (Figure 3.4, lanes 29-33, 41, 76 and 78).

Anomalies were identified during the PCR analysis of the transgenic *Arabidopsis* plants. In the first case, the UTRCC line gave mixed amplification results (Figure 3.4) with one of the three plants producing a promoter-GUS fragment (Figure 3.4; lanes 28-30). A similar result was obtained for a single plant in line TC-UTR1 and 588X, respectively (Figure 3.4).These results suggests that three of the 45 seed stocks collected, contained a mix of transgenic and non-transgenic seed or may not be stable lines (Das and Joshi 2011). Careful PCR screening of the plants was performed so that only plants which showed a clear band for the promoter fragment (results not shown) were assayed for GUS expression and uniform results could be obtained (Supplementary 3.1 and 3.2). In the second instance (line UTRII) no promoter-GUS amplicons could be produced (Figure 3.4, lanes 31-33) even when 5-10 different plants were screened, and does not contain the insert and reporter gene of interest. This could be due to a loss of the insert over generations or unsuccessful intergration and no further analysis was performed on line UTRII. There were three instances where the amplification failed and both the promoter fragments and the marker fragment were absent (Figure 3.4; lane 35, 51 and 114). These instances were repeated and bands were visualized (data not shown), indicating the promoter and reporter gene were present in all plants of these lines.

To accurately assay the GUS expression patterns modulated by the *EgrCesA8* promoter truncates, a number of controls (full-length *EgrCesA8* promoter, 35S promoter and wild-type negative control) where included in the analysis. The full-length (2018 bp) promoter fragment produced the expected GUS expression pattern as observed in the previous study (Creux et al. 2008) with GUS concentrated in the veins of the leaves and in

the hypocotyls where SCW formation is occurring (Figure 3.5O and T: -2018). This is in contrast to the positive control (2xCAMV35S promoter) where strong ubiquitous expression was observed in all parts of the leaf, hypocotyls and roots (Figure 3.5 B and G: 35S). To ensure results were not due to intra- or inter-line specific variation, three plants from at least four independent transgenic lines were tested for each of the eight *EgrCesA8* promoter constructs and the 35S control (Supplementary files 3.1 and 3.2). The wild type *Arabidopsis* plants showed no endogenous expression in any of the tissues assayed (Figure 3.5 A and F: WT).

The *EgrCesA8* promoter truncates produced a range of different expression patterns depending on the region that was truncated. The *EgrCesA8* promoter truncation -1411, which spanned -1411 bp upstream of the ATG (Figure 3.2), showed a loss of GUS expression in leaf veins and hypocotyls when compared to the full-length promoter (Figure 3.5 N, O, S and T). However, in truncation -1036 the GUS expression was reinstated in the leaf veins and hypocotyls (Figure 3.5 M and R). Limited GUS expression was also observed in the -334 and UTR truncates in *Arabidopsis* (Figure 3.6 C, E, H and J). These results suggest that the promoter truncate from -1036 to the ATG contains the minimal *EgrCesA8* promoter which is necessary and sufficient to confer tissue-specific expression (similar to full-length promoter) to the GUS reporter gene in *Arabidopsis*.

While the previously mentioned promoter truncations showed simple presence and absence of GUS expression in all *Arabidopsis* tissues (leaf and hypocotyl), the -853, -588 and TC-UTR promoter truncations showed a deviation from the tissue specificity observed for the *EgrCesA8* full-length (-2018) (Figure 3.5 D, K, L, I, P and Q). The -853 promoter truncation appeared to retain GUS expression in hypocotyls while GUS expression in the leaf veins was almost completely lost (Figure 3.5 L and Q), which implies that the promoter region between -1036 and -853 may contain an element which is required for leaf expression. -588 regained

some GUS expression in the leaf but this was not confined to the vessels and appeared non-specific, suggesting the loss of important tissue-specific elements (Figure 3.5 K and P). In Chapter 2, a highly conserved TC-repeat element was observed at the TSS of a number of different *CesA* genes across different *Eucalyptus* species. The TC-UTR truncate contained the TC repeats from the *Eucalyptus grandis CesA8* gene linked to the 5'UTR in the position of the TSS (Figure 3.2). The TC-UTR truncation showed strong GUS expression, but the expression was not as confined to the xylogenic tissues and was similar to the expression pattern observed for the 35S promoters (Figure 3.5 B, D, G and I).

3.4.3 Eucalyptus CesA8 promoter truncations show variable reporter gene expression in Eucalyptus stem tissues

Spatio-temporal GUS expression was assessed in *Eucalyptus* woody stem tissues using Induced Somatic Sector Analysis (ISSA; Van Beveren et al. 2006 and Appendix A). GUS expression sectors were counted and sector frequencies were calculated for phloem (P), immature xylem (X1) and mature xylem (X2). The typical sectors observed in *Eucalyptus* tissues are represented in Figure 3.6 A-F. Using the sector profiles, generated in a previous study for the full-length *EgrCesA* promoters as a baseline for comparison (Appendix A), we analyzed the *EgrCesA8* promoter truncation constructs from the *Arabidopsis* assay in *Eucalyptus* woody stem tissue (Figure 3.7 A). All the promoter truncates, except for TC-UTR, produced a similar sector profile to that of the SCW-associated full-length *CesA* promoters (Figures 3.6 A-C and 3.7 A). Similarities were observed when the *Eucalyptus* ISSA results were compared to the *Arabidopsis* leaf and hypocotyl GUS results (Figure 3.5). The TC-UTR truncation produced a sector frequency profile most similar to the profile generated for the full-length primary cell wall associated *CesA* promoters (approx. 35% P,

60% X1 and 5% X2), which suggests that the TC-repeat element may enhance general gene expression without tissue specificity. Secondly, -1411 promoter truncation only produced 4 (*n* = 4) sectors (Figure 3.7 A) and this could be due to the repressive/inactive nature of the promoter fragment as observed in *Arabidopsis* (Figure 3.5 N and S).

3.4.4 Eucalyptus CesA8 promoter truncations show different reporter gene expression in Populus stem tissues when compared to Arabidopsis and Eucalyptus

The *EgrCesA8* promoter constructs were also investigated using ISSA in *Populus* stem tissues. The constitutive 35S control expression pattern was similar to that produced in *Eucalyptus* with GUS expression in the P, X1 and X2 regions (Figure 3.7 A and B). The analysis of the promoter truncates in *Populus* stem tissues revealed an increased variability in GUS expression, much of which did not agree with findings in *Eucalyptus* and *Arabidopsis* (Figure 3.5 and Figure 3.7 A). The -1411 and -1036 truncations produced sector profiles similar to those observed for the full-length primary cell wall-associated *CesA* promoters (Figure 3.7 A). The sector profile produced by -853 *EgrCesA8* promoter truncation in *Populus* was the only truncation that corresponded to the profile produced by the full-length *EgrCesA8* promoter in *Populus* (Figure 3.7 B). This result is similar to the *Eucalyptus* sector analysis, where the -853 promoter truncate produced the same expression profile as the full-length *EgrCesA8* promoter in *Eucalyptus*. However, the full-length *EgrCesA8* promoter in *Populus* produced a primary cell wall-associated sector profile, whereas in *Eucalyptus* the full-length *EgrCesA8* promoter produced a secondary-cell-wall-associated sector profile, suggesting that it will be difficult to accurately assess GUS sector profiles for the *Eucalyptus* truncates in *Populus* stems (Figure 3.7 A, B). The three smallest *EgrCesA8* promoter

truncations (-340, TC-UTR and UTR) all produced fewer than ten independent sectors (4, 2 and 0, respectively) and may not be expressed in the *Populus* stems.

3.4.5 Regulatory regions of the Eucalyptus CesA8 promoter coincide with mapped cis-elements

A number of observations where made when a comparative analysis was performed using the different lines of evidence collected during the study (Figure 3.8). Firstly, regions in the *EgrCesA8* promoter have been identified that can modulate the gene's expression, however, very few known *cis*-elements mapped to these regions (Figure 3.8). This suggests that there are many elements involved in *CesA* regulation that have yet to be identified. Secondly, in a few cases the *cis*-element mapping data and the *Arabidopsis* GUS expression data revealed distinct clustering of known *cis*-elements in the active regions of the *EgrCesA8* promoter. One notable cluster occurred between -700 and -900 bp upstream of the ATG (Figure 3.1) and when overlapped with the GUS expression data it fell within a region that conferred tissue specificity (Figure 3.8; orange). Closer inspection of this *cis*-element cluster revealed a number of known SCW-associated *cis*-elements including M46RE, ACTYP and SNBE. The apparent coincidence of the *cis*-element cluster position with regions conferring tissue specificity in the *EgrCesA8* promoter suggest the presence of a putative *cis*-regulatory module (CRM) involved in *EgrCesA8* gene regulation (Figure 3.8). Finally, the TC-repeat element at the TSS of the *EgrCesA8* promoter produced similar GUS expression profiles in *Arabidopsis* and *Eucalyptus* when appended to the 5' UTR. If only the 5'UTR was present no GUS expression was observed in *Arabidopsis* (Figure 3.8). The TC-UTR GUS expression pattern was not tissue-specific in either species and strong GUS staining was observed in

xylogenic and non-xylogenic tissues. The TC-repeat element appears to activate gene expression and is likely an important core element for the *EgrCesA8* promoter.

3.5 Discussion

3.5.1 EgrCesA8 gene expression is modulated by a succession of functionally conserved upstream activators and repressors

To investigate the different mechanisms involved in the regulation of the *EgrCesA8* promoter we produced eight different *EgrCesA8* promoter constructs (Figure 3.2) and identified regions of the promoter that appear to modulate the gene's expression. In *Arabidopsis*, the GUS expression patterns of the different *EgrCesA8* promoter fragments was monitored in the leaves, roots and hypocotyls of *Arabidopsis* plants (Figure 3.5 A-T). A BLAST search (http://blast.ncbi.nlm.nih.gov) revealed that the *EgraCesA3* promoter fragment identified by Lu et al. (2008) was highly similar (E-value = 0.0 and 99% identity) to the *EgrCesA8* promoter fragment investigated in this study. It is highly likely that these are the same promoter from different *E.grandis* genotypes and will be referred to as *EgrCesA8* in this study. Lu et al. (2008) also conducted a deletion study on this promoter by monitoring GUS expression in tobacco. A comparison of the promoter truncates used in the two studies (Figure 3.1) showed some similarities (e.g. -588/A3Phinc and -853/A3phpa) and differences in the promoter constructs (Figure 3.1), which allowed us to compare the results and identify shared and novel regions of activity in the promoter regions.

The functional regions proposed by Lu et al. (2008) could be confined to shorter regions of the promoter in this study due to the similarity and differences of the constructs in the two studies. The full-length *EgrCesA8* promoter fragment (-2018 bp) produced GUS

expression localized in the veins of the leaves and the hypocotyls of *Arabidopsis* (Figure 3.5 O and T) and the differentiating (X1) xylem tissue of *Eucalyptus* (Figure 3.6 and 3.7A). The longest fragment evaluated by Lu et al. (2008) was -1865 bp upstream from the ATG and this fragment did not show strong GUS expression in the tobacco stem tissues they evaluated. This suggests that a 147 bp region from -2012 bp to -1865 bp is able to restore the expression of the promoter and suggests the position of a possible activator which has not been previously identified (Figure 3.8).

Deleting approximately 500 bp at the 5' end of the full-length promoter (fragment -1411 bp; Figure 3.2) significantly decreased GUS expression and in some cases GUS was undetectable in *Arabidopsis* (Figure 3.5, Supplementary 3.1 and 3.2) and *Eucalyptus* tissues (indicated by the low sector count; Figure 3.7 A). The *EgrCesA8* promoter fragments, -1865 bp (A3P; Lu et al. 2008) and the -1411 bp (Figure 3.1) did not express significant levels of GUS in tobacco or *Arabidopsis*, respectively. Both also contained a region previously identified by Lu et al. (2008) as a repressor region (Figure 3.1). GUS expression was regained when the next 400 bp region at the 5'end of the promoter was also deleted (Fragment -1036 bp; Figure 3.2). Similarly, Lu et al. (2008) also observed regained GUS expression with their A3Pssp fragment (Figure 3.1), suggesting that a repressor binds to the -1411 bp to -1266 bp region. In this manner the repressor region identified by Lu et al. (2008) could be decreased from 600 bp to 145 bp (Figure 3.1 and 3.8). Similarly, an activator was also identified by Lu et al. (2008) in the region -1266 bp to -831 bp (Figure 3.1) and the current study found that the -1036 bp construct still produced the expected expression pattern suggesting that the activator/enhancer region may be located in the region from -1036 bp to -831 bp which decreased the region to 205 bp (Figure 3.8). These smaller repressor and activator regions defined by this study can now be used for further analyses with techniques such as EMSA, protein pull-down or yeast-1-hybrid assays.

In tobacco, Lu et al. (2008) identified a region (-699 bp to -853 bp) in the *EgrCesA8* promoter, which they suggested was involved in the repression of the *EgrCesA8* gene in phloem tissues. In this study the -853 bp *EgrCesA8* promoter construct expressed GUS well in the hypocotyls, but weakly or not at all in the leaf veins in contrast to the -1036 bp construct which expressed well in both *Arabidopsis* tissues (Figure 3.5). This result suggests that the -853 bp promoter truncation may expose a phloem repressor element either by removing the binding site of a competitive activator (Seo et al. 2013), changing the conformation of the DNA allowing a repressor to bind (Raiber et al. 2012) or simply removing a phloem activator (Lu et al. 2008) which lead to a decrease of GUS expression in the phloem tissues. This could explain the visible decrease in GUS expression in the leaf tissues, as veins are delicately balanced with a small amount of phloem and xylem in order to transport nutrients and water to and from the plant's extremities. *Arabidopsis* hypocotyls, however, are similar in structure to woody stems (Lev-Yadun 1997) which have a higher xylem to phloem ratio and are actively depositing a great deal of cellulose (Plomion et al. 2001). This may explain the constant GUS expression observed for the -853 bp promoter construct in the hypocotyls and supports the findings of Lu et al. (2008) which identified the phloem repressor element between -850 and -700 bp (Figure 3.1).

Lu et al. (2008) focused their study on elements involved in the formation of tension wood and identified a region (-588 bp to -566 bp) of the *EgrCesA8* promoter that appeared to up regulate reporter gene expression during tension wood formation (Figure 3.1). However, this element was not identified in this study because the constructs were not evaluated during tension wood induction. However, this finding may explain some of the results observed for the -588 bp construct where GUS was observed in three of the five lines, but some individual plants within these lines did not show GUS expression (Supplementary 3.1 and 3.2) even though PCR tests confirmed the presence of the -588 bp construct in these plants (Figure 3.4).

Ko et al. (2004) found that plant body weight affects SCW formation and listed a number of genes including transcription factors that are expressed in response to a change in weight on the stem. These finding suggest the possible existence of weight response elements in the promoters of secondary cell associated genes. Since a number of plants were grown in a single tray in this study, it is possible that some of the plant stems were supported by neighbouring plants and so no GUS expression was observed in these plants due to a possible response element in the -588 promoter region only being active in the plants under stress. However, a separate study will have to be conducted to ascertain if this and other constructs harbour elements involved in various stress responses.

Taken together these results suggest that there is interplay between activators/enhancers and repressors along the length of the *EgrCesA8* promoter which are spatio-temporally guided by tissue-specific response elements. Cellulose deposition in the plant cell wall is an important developmental process and secondary cell walls are only a feature of a few specialized plant cells, including fibers and vessels (Plomion et al. 2001). The secondary cell walls are highly enriched for cellulose which is deposited into the cell wall at a specific time point, after cell elongation has occurred (Gorshkova et al. 2012). Three *CesA* genes have been associated with SCW formation by a number different expression studies in many different plant species including *Arabidopsis*, *Populus* and *Eucalyptus* (Hamann et al. 2004; Samuga and Joshi 2004; Ranik and Myburg 2006). *EgrCesA8* is a SCW-associated *CesA* and like its fellow SCW-*CesA* genes (*EgrCesA4* and *7*) should only be activated once the cells have completed elongation and then its activity will cease as the cell dies (Schrader et al. 2004). The processes of SCW deposition and programmed cell death are irreversible and therefore need to be prevented in cells which require only a primary cell wall. This suggests that the SCW-related *CesA* genes will require both repressors and activators in order to function effectively.

3.5.2 Identification of a putative secondary cell wall CRM harbouring known xylogenic-related cis-elements

Upon comparing the *cis*-element map, the truncation map (Figure 3.1) and the GUS expression data from tobacco (Lu et al. 2008), *Arabidopsis* and *Eucalyptus* (Figure 3.8) a number of interesting observations can be made. Firstly, only a few known *cis*-element occurrences mapped to the promoter regions that activated (-1036 to -853 and -2012 to -1865) or repressed (-1411 bp to -1266 bp) GUS expression (Figure 3.8). Recent studies have suggested from yeast-1-hybrid data that over 30 transcription factors can interact with the *Arabidopsis CesA8* promoter and so many *cis*-elements would be expected (Hazen 2012). It is known that *cis*-elements are often found clustered together in *cis*-regulatory modules (CRMs) within promoters, suggesting that the low abundance of *cis*-elements in these functionally active regions are due to the lack of *cis*-element information for *Eucalyptus CesA* promoters. Currently only our previous studies have specifically investigated the *cis*-elements of the *CesA* promoters in *Eucalyptus* (Creux et al. 2008, Chapter 2) and thus most of the *cis*-elements represented in Table 3.1 are from *Arabidopsis* and *Populus* studies which may be under represented in the *Eucalyptus CesA8* promoter.

Secondly, both this study and that of Lu et al. (2008) identified a region within the *EgrCesA8* promoter (region -853 to -699) that played a role in tissue specificity. This functional part of the promoter also overlaps with a region in the *cis*-element map where a number of known elements clustered together (Figure 3.8). This tight cluster of *cis*-element occurrences in the phloem repressor region identified by Lu et al. (2008) consists of a number of *cis*-element motifs (M46RE, ACTYP, DOF, CRPE31, CTRMCAMV35S and BPC1), some of which have previously been associated with SCW formation. The agreement of the GUS expression data and the *cis*-element mapping data suggests the presence of a putative

SCW-related CRM within the *EgrCesA8* promoter. The CRPE31 (Figure 3.1 light pink) and CTRMCAMV35S (Figure 3.1 dark turquoise) elements, identified in this putative CRM, are a general feature of *Eucalyptus CesA* promoters associated with both primary and secondary cell wall formation and may play a general role in *Eucalyptus CesA* gene regulation. Several of the other *cis*-elements (M46RE, ACTYP and SNBE) identified in this cluster have been directly associated with well-known transcription factors in the SCW regulatory network such as master regulators *At*SND1 and *At*MYB46 (Winzell et al. 2010; Zhong et al. 2010b; McCarthy et al. 2011; Kim et al. 2012).

The ACTYP element (Figure 3.1 red: CCACCAAC) and the MBSIIG/ M46RE/ SMRE element (Figure 3.1 medium green: RKTWGGTR) are in perfect reverse compliment of each other and were mapped to the exact same place in the putative CRM on opposite strands (red and green element; Figure 3.8). Winzell et al. (2010) identified the ACTYP element in promoters of hemicellulose genes from *Populus*, *Arabidopsis* and *Eucalyptus* which were highly expressed in xylem. Similarly, the MBSIIG/ M46RE/ SMRE element has been associated with SCW formation in both *Arabidopsis* and *Eucalyptus* (Goicoechea et al. 2005; Kim et al. 2012; Zhong and Ye 2012). In *Eucalyptus* it was hypothesized that this element is involved in the regulation of lignin deposition. This element binds *Eg*MYB2 and when over expressed produced highly thickened cell walls, leading the authors to speculate that it may also be involved in the regulation of cellulose and hemicelluloses genes (Goicoechea et al. 2005). Later it was found that *Eg*MYB2 is an ortholog of *At*MYB83, which is a functional ortholog of *At*MYB46 (McCarthy et al. 2009; Cassan-Wang et al. 2013). The *Arabidopsis* study found this motif in the *AtCesA4*, *7* and *8* promoters by transactivation analysis and yeast-1-hybrid showed direct interaction of *AtMYB46* with the *AtCesA4* promoter (Kim et al. 2012; Zhong and Ye 2012). These studies support the

occurrence of this motif in the tissue-specific region of the *EgrCesA8* promoter and could be important for the regulation of this gene during phloem and xylem formation.

Two other *cis*-elements were identified in this putative CRM, which have yet to be associated with cellulose deposition. They were identified after a scan of the *EgrCesA8* promoter sequence on the TRANSFAC database, which houses all experimentally verified *cis*-elements for all eukaryotic species (www.gene-regulation.com; Wingender et al. 2000; Wingender 2008). The first element was P$DOF_Q2 (NNAAAAAGNAN; Figure 3.8 light turquoise), which has been identified to bind Dof transcription factors (Yanagisawa and Sheen 1998). The plant-specific Dof proteins have a single zinc finger and all appear to bind the same conserved motif, but due to the nature of zinc fingers they have also been found to readily form hetero- and homo-dimers, which may provide the binding specificity observed in most DNA-protein interactions (Yanagisawa 2004). Recently a number of Dof proteins have been implicated in gene regulation in the vascular cambium and xylogenic tissues of plants (Guo et al. 2009; Gardiner et al. 2010; Gerhardt et al. 2011) and therefore it is possible that these proteins could bind the *EgrCesA8* promoter and mediate its tissue-specific expression.

The second identified element is P$BPC1_Q2 (Figure 3.1: yellow and Table 3.1), which is a GA-rich element (RGARAGRRA) found in the promoters and 5'UTRs of different plant genes. These GA-rich elements are bound by the Basic Pentacysteine (BPC) proteins and this interaction changes the conformation of the DNA which in turn represses or activates gene expression (Meister et al. 2004; Kooiker et al. 2005). BPC proteins interact with important homeotic proteins such as the MADS genes and plays a role in gene regulation in a number of different plant developmental processes including leaf, flower, seed and meristem development (Monfared et al. 2011; Simonini et al. 2012). Whether or not these two elements

and the proteins they bind are directly involved in the regulation of cellulose deposition through the interaction with the *CesA* promoters remains to be seen.

3.5.3 TC-repeats at the TSS are central to the regulation and initiation of EgrCesA gene expression in Eucalyptus

Multiple occurrences of the CTRMCAMV35S element (Table 3.1: TCTCTCTCT) occurred at the transcriptional start site (TSS) of the *EgrCesA8* promoter in position -244 to -222 (Figure 3.8; dark turquoise). Previous studies have shown that many of the *Eucalyptus CesA* promoters (primary and secondary cell wall related) do not have a TATA-box, but most have a CRPE31 and CTRMCAMV35S element in the region of the TSS (Creux et al. 2008). A nucleotide diversity study also revealed that this region over the TSS, harbouring the CTRMCAMV35S elements, was highly conserved across 13 different *Eucalyptus* species which points to functional constraint in the region (Chapter 2). The CTRMCAMV35S element appears to be a feature of the core *Eucalyptus CesA* promoters regardless of primary or secondary cell wall associations and it is not as well represented in the *Populus* and *Arabidopsis* promoters analyzed (Chapter 2). TATA-less promoters have been well documented in the past for many different organisms including *Saccharomyces*, *Drosophila* and *Arabidopsis* (Kutach and Kadonaga 2000; Yang et al. 2007; Bernard et al. 2010). These TATA-less promoters do not have the well-known TATA-box consensus sequence and instead rely on other core initiator elements to start gene expression (Smale 1997; Yang et al. 2007; Zuo and Li 2011). Molina and Grotewold (2005) identified a similar repeat element at the TSS of some *Arabidopsis* promoters and a comparative genome wide study of rice and *Arabidopsis* promoters also identified a CT-rich (Y-patch) region close to the TSS which was absent in mammalian genomes (Yamamoto et al. 2007). It is possible that the

CTRMCAMV35S element at the TSS of the *EgrCesA8* promoter represents one of these previously identified initiator elements.

To test the functionality of the multiple CTRMCAMV35S repeat (TC-repeat element) region at the TSS of this promoter the GUS expression modulated by the 5'UTR with or without the repeats was analyzed in *Arabidopsis* (Figure 3.5 C, D, H and I) and *Eucalyptus* (Figure 3.7 A). In *Arabidopsis* the UTR alone was not sufficient to express GUS, however when the TC-repeat element was present, the expression pattern was more similar to the constitutive CAMV35S promoter than to the tissue-specific pattern observed for other *EgrCesA8* promoter constructs (Figure 3.5). A similar loss of tissue specificity was observed in the *Eucalyptus* stem tissue where the UTR-only construct showed X1 expression and the TC-UTR construct expressed GUS in both phloem and developing xylem similar to the 35S control (Figure 3.7 A and 3.8). A *Eucalyptus*-specific genome-wide *cis*-element search identified a TC-repeat element at the TSS of a large number of promoters that were TATA-less and this group of promoters was enriched for promoters of genes involved in cellulose biosynthesis (van Jaarsveld *personal communication*). These results suggest that the inclusion of the TC-repeat element with no supporting regulatory elements drives non-specific gene expression in *Eucalyptus*. A recent review of eukaryotic promoters has described a type of TCT promoter which regulates highly expressed genes in metazoans indicating that CT-rich regions appear to be important for gene regulation in many different species, even across great evolutionary distances (Lenhard et al. 2012).

Closer inspection of the TC-repeat element at the *EgrCesA8* TSS revealed that these occurrences are part of a $CT_{(11)}$-microsatellite that begins at the point where most of the transcriptome reads for the gene began (Mizrachi et al. 2012, BOGAS; http://bioinformatics.psb.ugent.be/webtools/bogas). Recently, Bernard et al. (2010) found a plant-specific element in the TATA-box position (-50 bp) that is TC-rich and they proposed

that this element often occurs with a CT-microsatellite. Indeed, we observed a second TC-rich region at -50 from the TSS, suggesting that the *EgrCesA8* core promoter structure is similar to promoter structures previously observed for rice and *Arabidopsis* (Yamamoto et al. 2007).

Repeat/microsatellite regions in DNA have been shown to regulate transcription on many different levels by facilitating different DNA-protein interactions, binding transcription factors directly or by influencing secondary DNA structures (Zhang et al. 2006; Berger and Dubreucq 2012; Bochman et al. 2012). One hypothesis for the role of the $CT_{(11)}$-microsatellite in regulating *EgrCesA8* gene expression may be explained by a family of poorly characterized transcription factors known as the BBR/BPC (Barley B Recombinant/Basic Pentacysteine) protein family. The different members of this protein family can bind to GAGA and CTCT elements within the promoters on which they act (Meister et al. 2004; Kooiker et al. 2005; Berger and Dubreucq 2012). They appear to be functional orthologs to the GA Factors (GAFs) first identified in *Drosophila*. GAFs bind to GAGA elements and play many roles in animal development. One such role is the stabilization of the pre-initiation complex (PIC) of the polymerase by binding to GAGA elements closely associated with the TSS (Reviewed in Berger and Dubreucq 2012). It is possible that the $CT_{(11)}$-microsatellite in *Eucalyptus* plays a role in the formation or stabilization of the PIC and therefore may explain the strong non-specific expression of GUS observed in the presence of this element. However, further detailed analyses such as yeast-1-hybrid, gel-shift assays and Chip-seq are required to understand the role of the BBR/BPC proteins and the $CT_{(11)}$-microsatellite in the regulation of cellulose biosynthesis in *Eucalyptus*.

GAFs can also change the DNA conformation when they bind to the GAGA element and in this way regulate transcription (Kooiker et al. 2005). CT- and GA repeats have been

implicated in DNA conformational changes even in the absence of DNA-binding proteins such as GAFs. Zhang et al. (2006) found that CT/GA microsatellites were over represented in the 5' non-coding regions of plants with respect to the rest of the genome and that they were well conserved across different plant species. For a long time CT/GA repeats have been linked to alternate DNA conformations such as cruciform, Z-DNA, triplex and G-quadruplex (Sinden 1994). In particular, CT/GA repeats have been implicated in the formation of DNA triplexs and G-quadruplexes under various *in vitro* conditions (Sinden 1994). During *in vivo* cellular processes such as DNA replication, and transcription the unwinding of the DNA can cause DNA supercoiling, which is known to stabilize these alternate DNA conformations (Bochman et al. 2012). It has been hypothesized that the change in conformation of the DNA in a promoter region could initiate or repress transcription and with the new evidence for the role of long non-coding RNAs in gene regulation they suggests that these RNA molecules may bind into a DNA helix to form a triple helix and regulate gene expression (Lee 2012). There is little known about the role of DNA conformation in gene regulation and is a field which is gaining popularity as technology, *in silico* tools and sequencing techniques are advancing (Buske et al. 2011; Cer et al. 2011).

3.5.4 EgrCesA8 promoter regions display similar expression patterns between distantly related species

The *EgrCesA8* promoter constructs were tested directly in the woody stems of *Eucalyptus* and *Populus* to assess how the *EgrCesA8* promoter and the various truncations thereof affect the GUS expression in tissues that are actively depositing SCW. All constructs (except the TC-UTR) expressed GUS in the *Eucalyptus* immature xylem (X1) zone where the cells differentiate into fiber and xylem vessels and deposit large amounts of cellulose into the cell

wall during SCW formation (Plomion et al. 2001; Gorshkova et al. 2012). It was unexpected that deletions of large parts of the *EgrCesA8* promoter left the GUS expression patterns largely unchanged in *Eucalyptus* stem tissue and these results appeared not to agree with the GUS expression patterns observed in the *Arabidopsis* tissues. Upon closer inspection and comparison of the *Arabidopsis* hypocotyls GUS expression patterns it was found that more similarities in expression occur. Specifically, the full-length (-2018) *EgrCesA8* promoter and the -1036, -835 and -588 promoter truncations all showed GUS expression in the *Arabidopsis* hypocotyls (Figure 3.5, 3.8 and Supplementary 3.1 and 3.2), a stem region which contains xylem vessels and fibers similar to the immature xylem of woody stems (Lev-Yadun 1997).

Little or no GUS expression was observed in *Arabidopsis* for the -1411 promoter construct and in *Eucalyptus* only four sectors were observed for this construct, which is significantly lower than for the other constructs that were tested in *Eucalyptus* (*n*=number of sectors; Figure 3.7). The low sector count for this construct in *Eucalyptus* may indicate failure of transformation, although, this is unlikely because all of the promoters were used for transformation at the same time in the same clones in replicate and none of the other promoter constructs displayed such low sector numbers (four sectors). It is more likely that the low sector counts are an indication of a decreased GUS expression driven by this promoter construct and that the four sectors that were produced were due to random insertions into the genome. The cambial cells that we transformed are continually developing into xylem and phloem cells and this suggests that many genes will be specifically expressed in the X1 region of the stem (Hertzberg et al. 2001; Schrader et al. 2004; Mizrachi et al. 2010) and therefore the possibility of randomly inserting GUS downstream of a strongly X1 active promoter is increased and could explain the X1 specificity of these four sectors observed for -1411. It is also possible that these four sectors are cases where the construct was inserted randomly at a strong promoter but in xylem mother cells which have already

differentiated and only produce xylem cells and therefore X1 sectors were produced. This would then suggest that the -1411 promoter construct is also repressed in *Eucalyptus* and concurs with the *Arabidopsis* and tobacco results (Figure 3.1 and Lu et al. 2008).

Another construct that showed similar results in *Arabidopsis* and *Eucalyptus* was the TC-UTR promoter construct ($CT_{(11)}$-microsatellite and UTR), in both species the TC-UTR promoter construct showed a loss of tissue specificity compared to the longer promoter fragments and produced expression profiles similar to those generated by the constitutive control promoter (2xCAMV35S), which supports the hypothesis that this element is a general enhancer/activator involved in initiation of transcription as discussed earlier (Figures 3.5, 3.7 and 3.8). When the $CT_{(11)}$-microsatellite was removed from the 5'UTR, the expression patterns changed and the 5'UTR alone was sufficient to drive tissues specific expression in the *Eucalyptus* stems, but not in the *Arabidopsis* tissues. This suggests there may be *Eucalyptus*-specific regulation occurring within the 5'UTR of the *EgrCesA8* promoter.

3.5.5 Two EgrCesA8 promoter regions display variable expression patterns between distantly related species

Two *EgrCesA8* promoter truncation constructs (-340 and UTR) showed different GUS expression patterns in *Arabidopsis* and *Eucalyptus*, where neither construct expressed GUS in *Arabidopsis* tissues, but both -340 and UTR expressed GUS in the *Eucalyptus* X1 tissue (Figures 3.5 and 3.7 A). Lu et al. (2008) had two fragments (Figure 3.1; A3Pmfe and A3Pacc) which are similar to the -340 fragment cloned in this study and they also observed little or no expression of GUS in tobacco under normal physiological conditions. These results can be explained in two ways; either -340 expresses low levels of GUS only in the immature xylem tissues of *Arabidopsis* and our staining method was not sensitive enough to

observe GUS in these tissues; or these promoter constructs harbour *Eucalyptus*-specific regulatory elements that cannot be recognized by the *Arabidopsis AtCesA8* regulatory machinery.

The ISSA results of the *EgrCesA8* promoter constructs in the *Populus* stem tissues compared to the results obtained for these constructs in *Arabidopsis* and *Eucalyptus* (Figures 3.5 and 3.7 A) were highly variable. This finding was similar to the ISSA expression patterns previously obtained for the six "full-length" *Eucalyptus CesA* promoters investigated in *Populus* and *Eucalyptus* woody tissues (Chapter 2). The sector frequencies observed for the constitutive 35S promoter were similar in *Eucalyptus* and *Populus* suggesting that the variation observed for the -340 fragment is likely due to the *Eucalyptus* origin of promoters. In a previous study, promoters of the *Populus PAL2* genes were isolated and transformed into *Populus* and tobacco with a GUS reporter gene. The heterologous *Populus* promoters expressed GUS differently in tobacco than in the *Populus* genetic background and the authors suggested that these promoters may function differently in the two species (Gray-Mitsumune et al. 1999). A genome-wide analysis of *cis*-elements across vertebrate evolution showed a number of linage-specific *cis*-element mutations (Yokoyama et al. 2011), suggesting that promoters from different genera can display variable reporter gene expression in transgenic model species such as tobacco and *Populus*.

Further support for the variable GUS expression patterns observed in this study are reflected in the analysis of the respective plant genomes. *Populus*, *Arabidopsis* and *Eucalyptus* have undergone a number of genome wide duplications some of which are not shared across these different species (Tuskan et al. 2006; Jiao et al. 2011). *Populus* has undergone a recent duplication and minimal gene loss, which together have led to a number of duplicated genes that are still expressed and this may affect the regulatory networks within this tree species (Hu et al. 2010). Particularly, if one looks at SCW formation and cellulose

deposition, it is evident that *Arabidopsis* and *Eucalyptus* have 10-11 expressed *CesA* genes but *Populus* has 18 expressed *CesA* genes (Richmond and Somerville 2000, Ranik et al. unpublished; Djerbi et al. 2005; Suzuki et al. 2006). Recently, it was shown that *CesA8A* (an *EgrCesA8* ortholog) in apple was located directly downstream of a *WDR53* gene and that these genes are bicistronic and that under stress response (such as pathogen infection) this bicistronic transcription leads to an alternate splice variant of *CesA8A* in apples. It was also shown that the *WDR53* and *CesA8A* genomic positioning was conserved across a number of different plant genera, except for *Arabidopsis* and *Eucalyptus* which both lack a *CesA8* duplicate. This suggests that the *CesA* genes may have different roles and therefore regulatory mechanisms in different plant species (Guerriero et al. 2012). Given the genomic differences in the three species analyzed in this study, the differences in GUS expression driven by the *EgrCesA8* promoter in these species suggests that some species may not be suitable for in depth heterologous promoter analysis and that a knowledge of the plant genomes involved can aid in the correct pairing of test and model organisms to obtain the most accurate results.

3.6 Conclusion

There have only been a handful of studies that have investigated the *cis*-element composition of the *CesA* promoters in plants (Creux et al. 2008; Lu et al. 2008; Ding et al. 2012 and Chapter 2), and only one study that has investigated the *Eucalyptus* SCW-associated *CesA8* promoter. In the latter they identified several general regions of activation/repression ranging in size from 600 bp to 135 bp (Lu et al. 2008). In this study we used the promoter deletions in the Lu et al. (2008) study and the *cis*-element data obtained from previous studies to further dissect the *EgrCesA8* promoter. The GUS results concurred with the results obtained by Lu et al. (2008) and the larger regulatory regions they had previously identified could be decreased to between 200 bp and 100 bp. In addition to confirming previous regions of repression and

activation several novel regions of activation were also identified in the *EgrCesA8* promoter. By comparing the *cis*-element data to the GUS expression data, a CRM that contained a cluster of known *cis*-elements that overlapped with active phloem-specific promoter region was identified (Figure 3.8). Previous *in silico* studies (Zhang et al. 2006; Yamamoto et al. 2007; Creux et al. 2008 and Chapter 2) identified a CT-microsatellite at the TSS of many plant genes and this was also identified in the *EgrCesA8* promoter. This is the first study to show that the presence of a CT-microsatellite strongly activated non-specific expression of GUS in both *Arabidopsis* and *Eucalyptus* (Figure 3.8). From these and previous results it is possible to put together a simple model of regulation for the *EgrCesA8* promoter where the *EgrCesA8* gene expression is modulated by a series of activators and repressors along the length of the promoter. The specificity is provided by a number of elements situated in the CRM found between -1036 and -700 bp upstream and the expression is initiated by a non-specific $CT_{(11)}$-microsatellite and not the well described TATA-box. From this point there are a number of interesting directions in which further studies can go. Firstly, with the current focus on next generation sequencing technologies it would be interesting to investigate the evolution of the CRM and $CT_{(11)}$-microsatellite in the *CesA8* promoter region across many different *Eucalyptus* species. Secondly, using the SCW regulatory network for direction it would be interesting to see which of the known transcription factors bind to the putative CRM region and other activator and repressor regions of the *EgrCesA8* promoter. This current study is important because it brings together *in silico* and functional analyses of the regulatory regions involved in cellulose deposition of an economically important plantation tree and begins to unravel the regulation of a key biological process conserved in all plant species.

3.7 Acknowledgements

The authors would like to acknowledge Mr Marius Laubscher and Mr Andrew de Santos for assisting with harvesting, screening and growing *Arabidopsis* plants. This work was supported with funding provided by Mondi and Sappi, through the Forest Molecular Genetics (FMG) Programme, the Technology and Human Resources for Industry Programme (THRIP) and the National Research Foundation of South Africa (NRF).

Table 3.1 Previously published *cis*-regulatory elements associated with secondary cell wall formation and cellulose deposition in plants.

***Cis*-element name/ID**[a]	***Cis*-element accession number**[b]	***Cis*-element consensus sequence**[a]	**Plant species**[c]	**Reference**
XYLAT	S000510	ACAAAGAA	At	Ko et al. (2006)
TERE	-	CTTNAAAGCNA	At	Pyo et al. (2007)
SNBE	-	WNNYBTNNNNNNNAMGNHW	At	Zhong et al. (2010b)
ACTYP	-	CCACCAAC	Pt	Winzell et al. (2010)
MBSIIG/ M46RE/ SMRE	-	RKTWGGTR	At, Egr	Goicoechea et al. (2005), Kim et al. (2012), Zhong et al. (2012)
CRPE31	-	GNGNAGNG	At, Pt, Egr	Creux et al. (2008)
CRPE28	-	NNGCATGC	At, Pt, Egr	Creux et al. (2008)
CRPE26	-	TCCTGCYG	At, Pt, Egr	Creux et al. (2008)
CRPE25	-	RCYSTGCCC	At, Pt, Egr	Creux et al. (2008)
CTRMCAMV35S	S000460	TCTCTCTCT	At, Pt, Egr	Pauli et al. (2004), Creux et al. (2008), Ding et al. (2012)
REALPHALGLHCB21	S000362	AACCAA	At, Pt, Egr	Degenhardt and Tobin (1996) , Creux et al. (2008)
PYRIMIDINEBOXOSRAMY1A	S000259	CCTTTT	At, Pt, Egr	Morita et al. (1998), Mena et al. (2002), Creux et al. (2008)
NODCON1GM	S000461, S000467	AAAGAT	At, Pt, Egr	Stougaard et al. (1990), Vieweg et al. (2004), Creux et al. (2008)
MYB1AT	S000408	WAACCA	At, Pt, Egr	Abe et al. (2003) Creux et al. (2008)
P$CBNAC_01 - CBNAC	M01188	TTGCTT	At	Kim et al. (2007)
P$HBPA1_Q6_01 - HBP-1a	M00936	GHCACGTCAC	Gh	Niu et al. (1996)
P$DREB1B_01 - DREB1B	M01700	CCGAC	Os	Kim et al. (2009)
P$BPC1_Q2 - BPC1	M01126	AGAAAR	At	Meister et al. (2004)
P$DOF_Q2 – Dof	M01136	NNAAAAAGNAN	Zm	Yanagisawa et al. (1998)
P$MADSA_Q2 - MADS-A	M00408	ADWCCAAAAATGGAAA	At	Riechmann et al. (1996)
P$C1_Q2 - C1	M00439, S000179	YNAACYAYCNS	Zm	Grotewold et al. (1994)
P$PEND_02 - PEND class	M02233, MA0127	ANTTCTTATK	At	Mertin et al. (1999)
P$PBF_01 - PBF PBF (MPBF)	M00355	NWNWAAGNGN	Zm	Yanagisawa et al. (2002)
P$O2_Q2 - Opaque-2	M00443	CAKWYSACGTVRW	Zm	Aukerman et al. (1991)
P$O2_03 - Opaque-2	M00374	GATGAYATGG	Zm	Izawa et al. (1993)
-300MOTIFZMZEIN	S000002	RTGAGTCAT	At, Pt	Thomas and Flavell (1990), Ding et al. (2012)
ARE1	S000022	RGTGACNNNGC	At, Pt	Rushmore et al. (1991), Ding et al. (2012)
CEREGLUBOX1PSLEGA	S000032	TGTTAAAGT	At, Pt	Shirsat et al. (1989), Ding et al. (2012)
CEREGLUBOX3PSLEGA	S000034	TGTAAAAGT	At, Pt	Shirsat et al. (1989), Ding et al. (2012)
MRNA3ENDTAH3	S000069	AATGGAAATG	At, Pt	Ohtsubo and Iwabuchi (1994), Ding et al. (2012)
INTRONUPPER	S000085	MAGGTAAGT	At, Pt	Brown (1986), Ding et al. (2012)
TATABOX1	S000108	CTATAAATAC	At, Pt	Grace et al. (2004),

				Ding et al. (2012)
TOPOISOM	S000112	GTNWAYATTNATNNG	At, Pt	Sander and Hsieh (1985), Ding et al. (2012)
LBOXLERBCS	S000126	AAATTAACCAA	At, Pt	Giuliano et al. (1988), Ding et al. (2012)
SPHCOREZMC1	S000154	TCCATGCAT	At, Pt	Suzuki et al. (1997), Ding et al. (2012)
TCA1MOTIF	S000159	TCATCTTCTT	At, Pt	Goldsbrough et al. (1993), Ding et al. (2012)
S2FSORPL21	S000166	CCATACATT	At, Pt	Lagrange et al. (1997), Ding et al. (2012)
RGATAOS	S000191	CAGAAGATA	At, Pt	Yin et al. (1997), Ding et al. (2012)
ACIIIPVPAL2	S000194	GTTAGGTTC	At, Pt	Hatton et al. (1995), Ding et al. (2012)
BOX1PVCHS15	S000208	TAAAAGTTAAAAAC	At, Pt	Lawton et al. (1991), Ding et al. (2012)
AUXRETGA2GMGH3	S000235	TGACGTGGC	At, Pt	Liu et al. (1994), Ding et al. (2012)
PROXBBNNAPA	S000263	CAAACACC	At, Pt	Busk and Pages (1998), Ding et al. (2012)
TELOBOXATEEF1AA1	S000308	AAACCCTAA	At, Pt	Tremousayque et al. (2002), Ding et al. (2012)
D1GMAUX28	S000328	ACAGTTACTA	At, Pt	Nagao et al. (1993), Ding et al. (2012)
AAGACGTAGATACL12	S000344	AAGACGTAG	At, Pt	Baerson et al. (1994), Ding et al. (2012)
GMHDLGMVSPB	S000372	CATTAATTAG	At, Pt	Tang et al. (2001), Ding et al. (2012)
E2FAT	S000417	TYTCCCGCC	At, Pt	Ramirez-Parra et al. (2003), Ding et al. (2012)
UPRMOTIFIAT	S000425	CCACGTCA	At, Pt	Oh et al. (2003), Ding et al. (2012)
UPRMOTIFIIAT	S000426	CCNNNNNNNNNNNNCCACG	At, Pt	Oh et al. (2003), Ding et al. (2012)
ANAERO4CONSENSUS	S000480	GTTTHGCAA	At, Pt	Mohanty et al. (2005), Ding et al. (2012)
ANAERO5CONSENSUS	S000481	TTCCCTGTT	At, Pt	Mohanty et al. (2005), Ding et al. (2012)
SORLIP3AT	S000484	CTCAAGTGA	At, Pt	Hudson and Quail (2003), Ding et al. (2012)
SORLIP4AT	S000485	GTATGATGG	At, Pt	Hudson and Quail (2003), Ding et al. (2012)
SORLREP2AT	S000487	ATAAAACGT	At, Pt	Hudson and Quail (2003), Ding et al. (2012)
SORLREP4AT	S000489	CTCCTAATT	At, Pt	Hudson and Quail (2003), Ding et al. (2012)

[a]*Cis*-element names and consensus sequences as documented in literature or in the PLACE and TRANSFAC databases

[b]Accession numbers as provided on PLACE (S-numbers) and TRANSFAC (M-numbers). Dashes represent *cis*-elements that have been reported in literature, but are not listed on either database and thus have no accession number.

[c]Plant species in which the *cis*-elements were identified (At-*Arabidopsis thaliana*, Pt-*Populus trichocarpa*, Egr-*Eucalyptus grandis*, Gh-*Gossypium hirsutum*, Os-*Oryza sativa*, Zm-*Zea mays*)

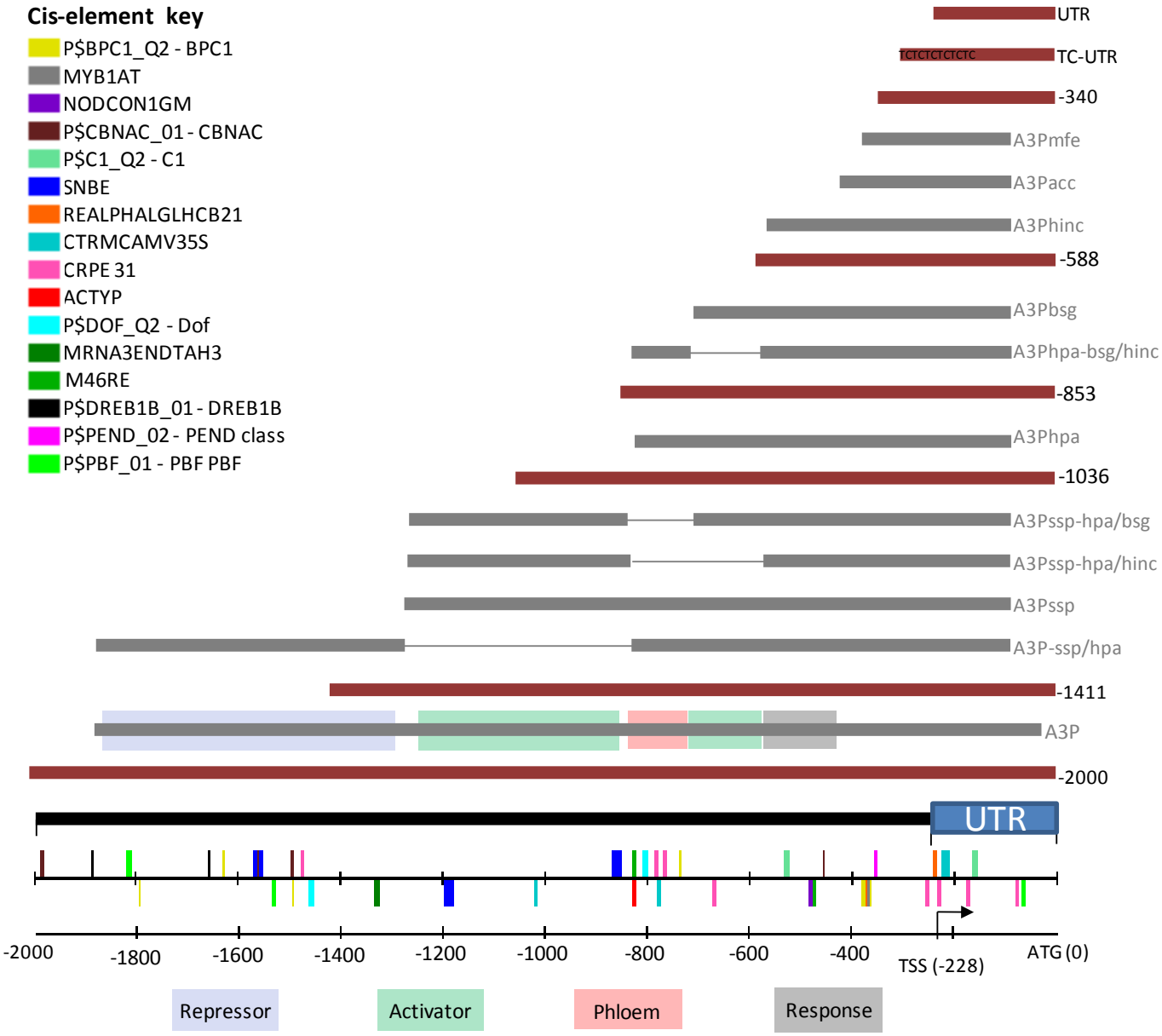
Table 3.2 Primer list for the generation of truncated promoter regions and for transgenic *Arabidopsis* screening by PCR amplification.

Primer name	Primer sequence(5'-3')	Position in *EgrCesA8* promoter
Promoter truncation primers:		
*EgrCesA8P*_ATG_reverse [a b]	CAATCTTCTCGCGACCCAAT	-7 to -28 (F_R)
*EgrCesA8P*_UTR_forward [a b]	CTGCTTCAACACAATGACAC	- 209 to -227 (C_F)
*EgrCesA8P*_UTR-TC_forward [a b]	CATGTCATCTCCCTCCTCTG	-269 to -288 (G_F)
*EgrCesA8P*_340_forward [a b]	AAGATGTAGCCAGCCCAACA	-321 to -340 (D_F)
*EgrCesA8P*_588_forward [a]	GGAGGGAATTGAGGTTGACA	-571 to -588
*EgrCesA8P*_853_forward [a]	GGTAGTTGGTGGGTTAACCT	-834 to -853
*EgrCesA8P*_1036_forward [a]	AGAGGGAGGGTAAGTTCACT	-1017 to -1036
*EgrCesA8P*_1411_forward [a]	TTGAGGTTGGATCGTGCTTCTG	-1390 to -1411
*EgrCesA8P*_2018_forward [a]	CCTTGCACATCCAATTGC	-2018 to -2018
***Arabidopsis* transgenic screening primers:**		
35s F primer [b]	GGTCAACATGGTGGAGCACGACACA	Figure 3.3 schematic (E_F)
GUS specific R primer [b]	TCCGGTTCGTTGGCAATACT	Figure 3.3 schematic (B_R)
819B03 Forward [b]	GTCAATCTTGGCCATGGACT	Figure 3.3 schematic (A_F)
819B03 Reverse [b]	GGTGGTCCTTGGTGTTGTTC	Figure 3.3 schematic(A_R)

[a] Primers that were used for *EgrCesA8* promoter truncation

[b] The letters (A-G) with subscript F (forward) or R (reverse) indicate the primers used for PCR confirmation of *Arabidopsis* transgenics as indicated in Figure 3.3

Figure 3.1 Comparison of the promoter truncates and known *cis*-element map of the *EgrCesA8* promoter. The promoter truncates produced in this study are indicated by dark red lines, while the grey lines indicate constructs developed by Lu et al. (2008). The thin grey line depict promoter deletions within the Lu et al. (2008) promoter constructs, which were created with various restriction enzyme digests as indicated in the names (*Ssp*I, *Hpa*I, *Hinc*I, *Bsg*I, *Acc*I and *Mfe*I). The A3P construct of the Lu et al. (2008) study represents their longest full promoter fragment. The pale coloured rectangles along A3P represent the repressor (blue), activator (green) and phloem-specific (pink) modules identified by Lu et al. (2008). The promoter constructs in this study were labelled according to the position upstream of the start codon where they were truncated (-2018, -1411, -1036, -853, -588, -340). Two constructs containing just the 5' UTR and the 5' UTR with a $CT_{(11)}$-microsattelite respectively were also included. The thin black line towards the bottom of the image represents the 2 kb *EgrCesA8* promoter sequence with the small coloured blocks representing positions of mapped *cis*-element occurrences along the length of the promoter. Coloured blocks above the line indicate *cis*-elements on the sense strand, while those below the line are on the complimentary strand. A colour key is provided where the *cis*-element ID or reference for each element is listed. The bottom line is a position maker broken into 200 bp units across the full 2 kb region. The position of the transcriptional start site is indicated in relation to the ATG (position 0).



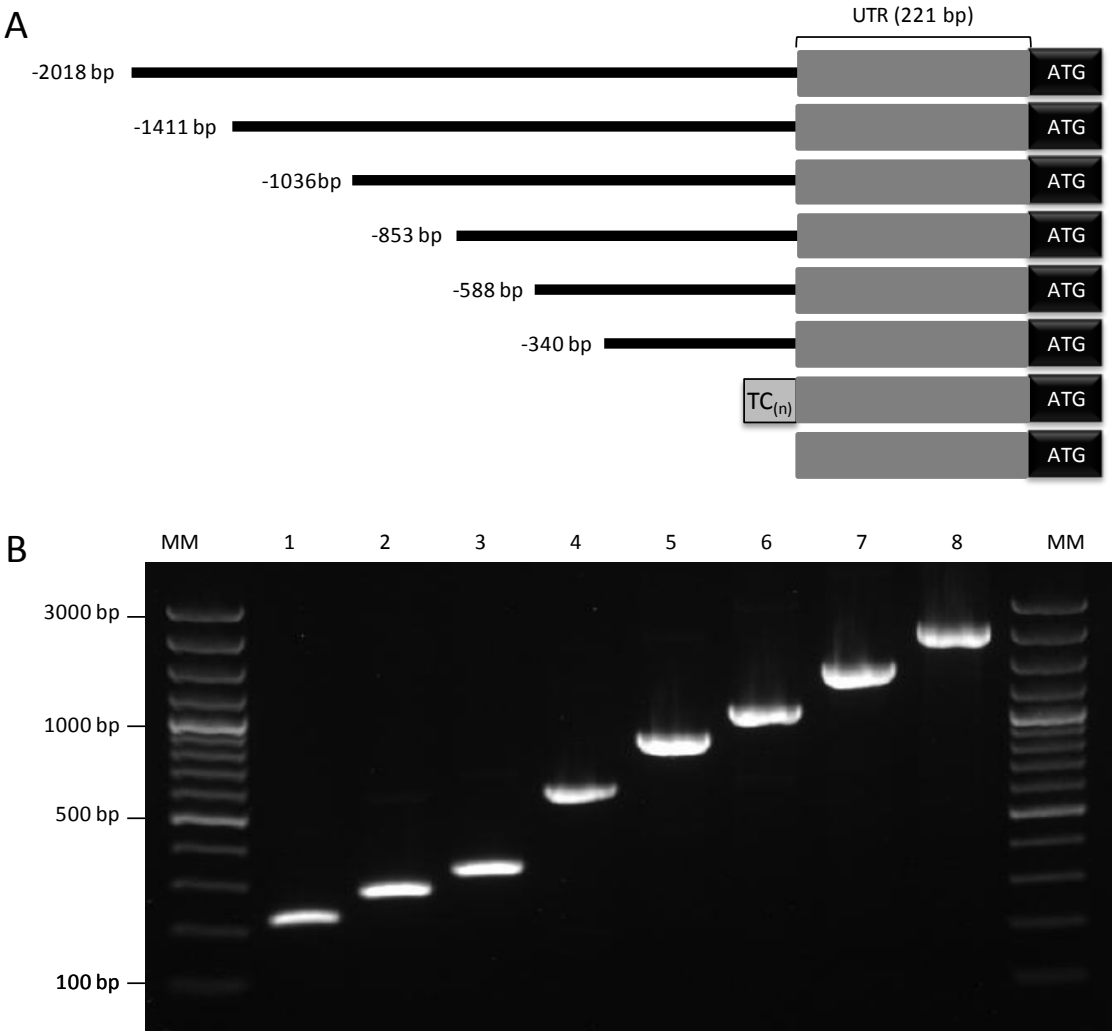


Figure 3.2 Serial truncation of *EgrCesA8* promoter for vector construction. (A) Schematic representing the step-wise PCR truncation of the *EgrCesA8* promoter. The -2018 bp fragment represents the full-length promoter and beneath it the progressively shorter promoter fragments are depicted. The start codon (ATG) is shown on the right by the black blocks. Grey blocks represent the 5'UTR and the solid black lines represent the fragment of *EgrCesA8* promoter included in each case. The size of each fragment from the ATG is indicated to the left and the length of the UTR is indicated above. The light grey block indicates the TC-repeat element/$CT_{(11)}$-microsatellite which is positioned at the transcriptional start site (Chapter 2). (B) Amplification of the *EgrCesA8* promoter and truncations. The reverse primer (Table 3.2) was the same in each case and located at the ATG, while the forward primers (Table 3.2) where distributed along the length of the promoter (Figure 3.2A). MM indicates the size standard (100 bp ladder) and the sizes are indicated on the left. Lanes 1 – 8 contain the promoter fragments and these fragments correspond with the sizes shown in Figure 3.4A.

Figure 3.3 Amplicon map for PCR confirmation of transgenic *Arabidopsis* lines containing the different *EgrCesA8* promoter constructs. Black lines and boxes represent the different regions amplified by the different primers sets (A_F and A_R; C_F and B_R; D and B_R; E_F and B_R). The primer names and sequences are listed in Table 3.2 and subscript F and R indicate forward and reverse primers respectively. The grey bars indicate regions not amplified during PCR screening. The wild type Col-0 plants will only produce one amplicon of 450 bp (primers A_F and A_R) as no transgenic inserts are present and indicates PCR amplification was successful. All plants screened will show the 450 bp amplicon. All transgenic plants will present a second amplicon during multiplex PCR. Transgenic *Arabidopsis* harboring *EgrCesA8* promoter constructs -2018 to -340 will produce a 2300 bp fragment when the GUS gene (black block) specific primer (B_R) and the promoter specific primer (D_F) are used for amplification. The positive constitutive CAMV 35S construct was amplified from the plants using the GUS-specific primer and a CAMV 35S specific primer (E_F). The *EgrCesA8* UTR promoter truncate was amplified with B_R and a UTR specific forward primer (C_F) resulting in a 2280 bp fragment. The *EgrCesA8* CT-UTR truncate (grey block) was amplified with 5'UTR specific forward and reverse primers (G_F and F_R) because G_F was not compatible with B_R (Schematic not to scale).

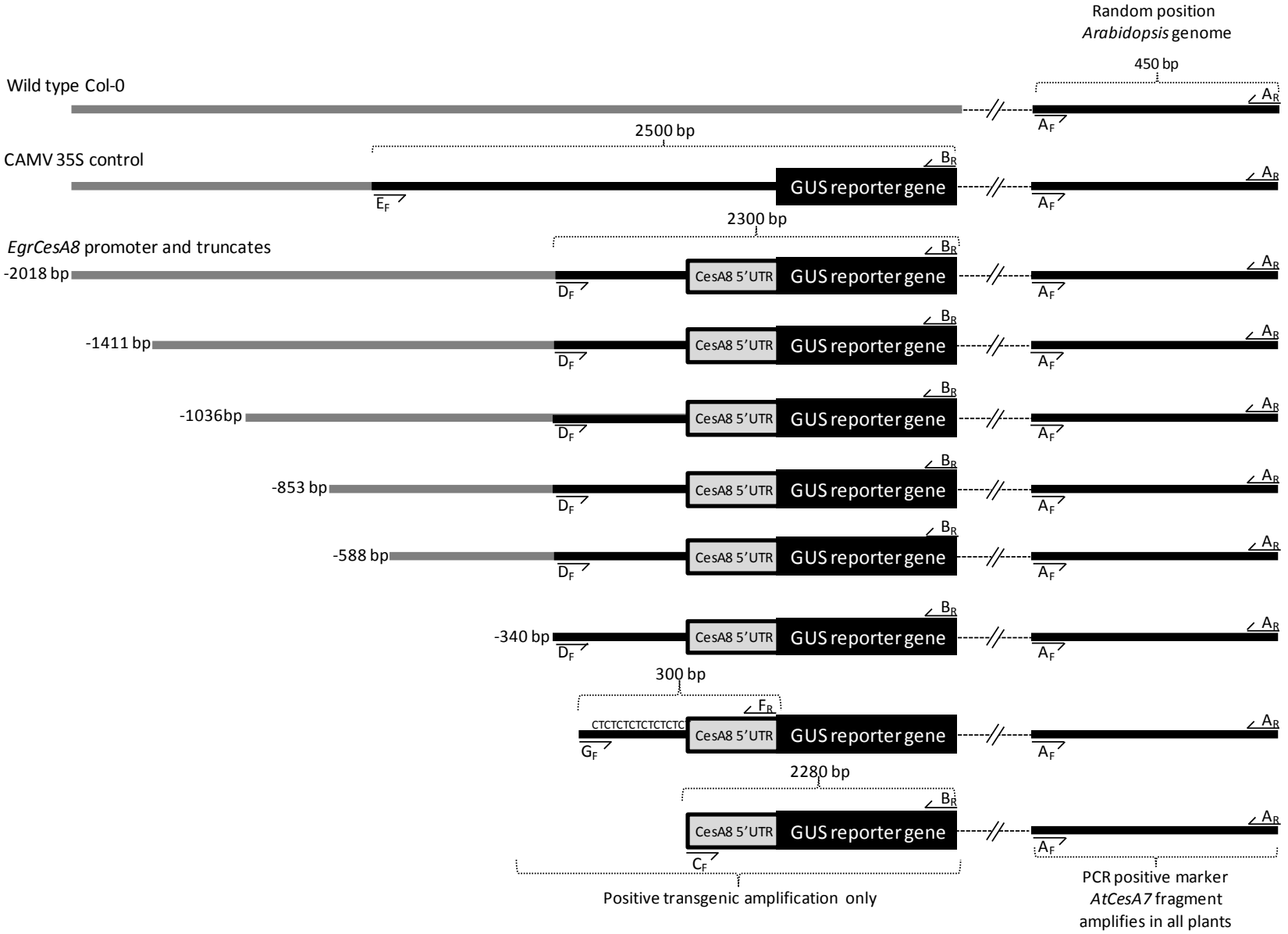
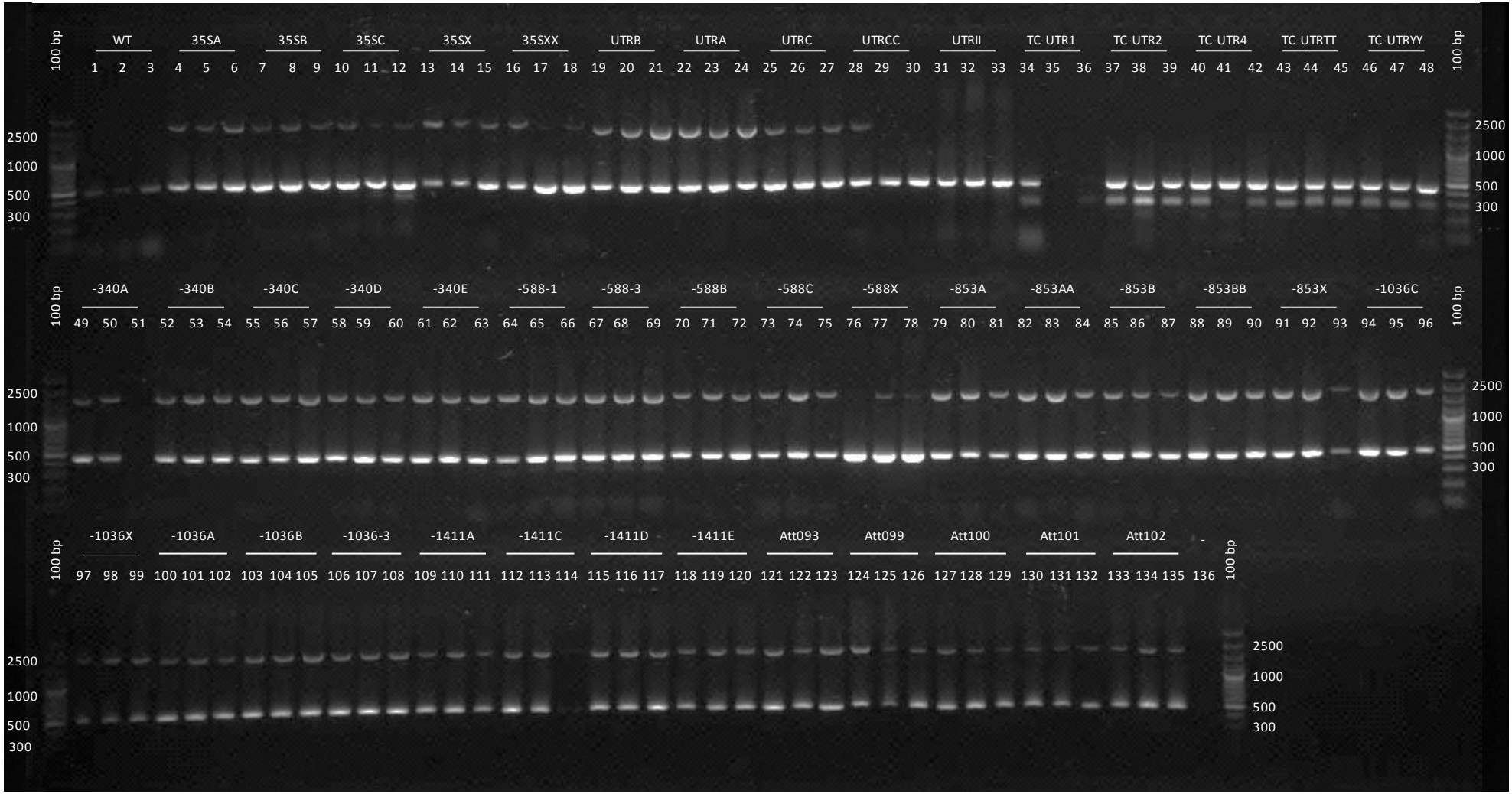


Figure 3.4 PCR confirmation of the *Arabidopsis* transgenic lines with the constitutive 35S promoter, *EgrCesA8* full-length promoter (-2018) and various truncations (UTR, TC-UTR, -340, -588, -1036, -1411) thereof, upstream of the GUS reporter gene. The image shows three representative amplifications of part of the *EgrCesA8* promoter and the GUS reporter gene for three individual plants from 4-5 independent lines for the full-length *EgrCesA8* promoter (Lines Att093, Att099, Att100, Att101, Att102; Lanes 121-135) and each of the various truncations (UTRA/B/C/CC/II, 340A/B/C/D/E, 5881/3/B/C/X, 853A/AA/B/BB/X, 1036A/B/C/X/3 and 1411A/C/D/E) in lanes 20-33 and 49-120, respectively. The double bands represent positive transgenic plants and the 2300 bp fragment represents the presence of the promoter or truncation and the GUS reporter gene in the correct orientation. The lower band (450 bp) in all samples (Lanes 1-155) is the positive PCR control indicating successful amplification. The wild type plants (Lanes 1-3) only contain the 450 bp band indicating they are not transgenic. Lanes 4 – 19 contain fragments amplified with a 35S promoter-specific and GUS-specific primer, which produce a 2500 bp fragment indicating the 35S promoter and GUS reporter gene are incorporated into the *Arabidopsis* genome in the correct orientation. The TC-UTR promoter truncation was amplified with promoter specific primers (Table 3.2) and produced a band of 300 bp in the lines (TC-UTR1/2/4/YY/TT) transformed with this construct (Lanes 34-48). A 100 bp molecular weight standard was loaded at the beginning and end of each row of samples and the 2500 bp, 1000 bp, 500 bp and 300 bp bands are indicated in each case. Construct name and line designation is indicated above each set.



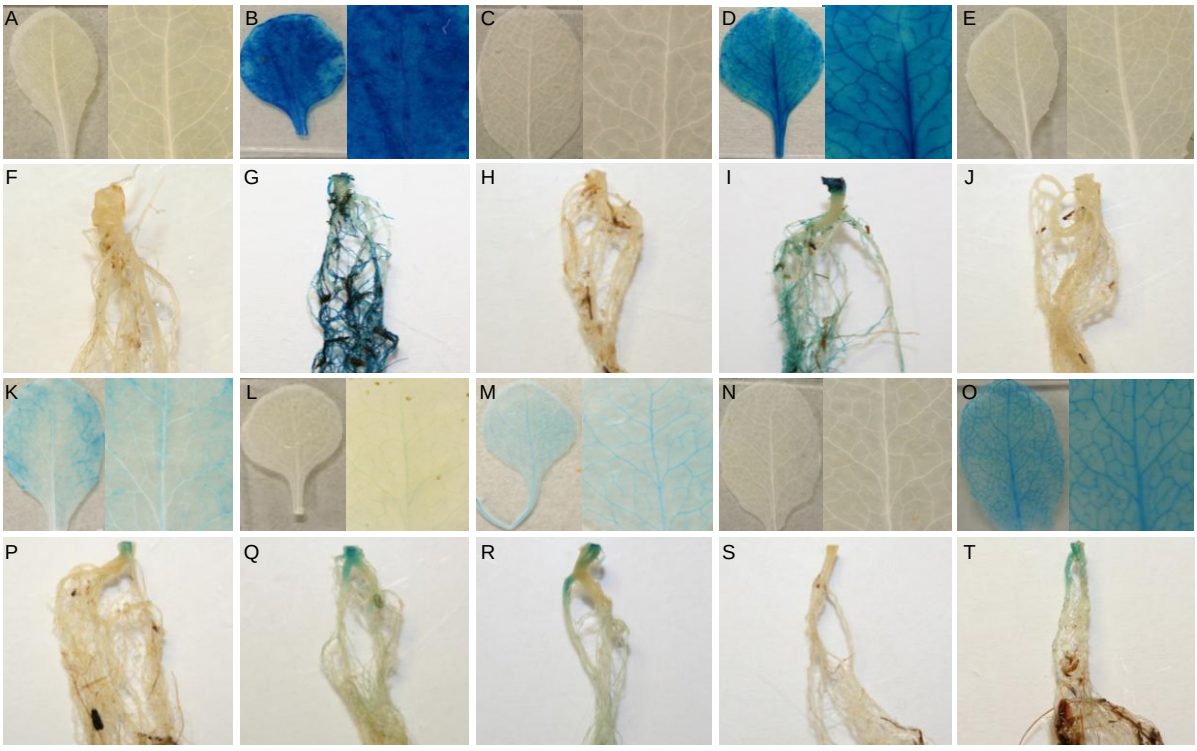


Figure 3.5 Histological GUS assays of the promoter truncation constructs in *Arabidopsis* leaves and hypocotyls. The leaf images with the hypocotyls image directly below each indicates the leaf and hypocotyl from the same plant of a specific construct. A and F show no GUS expression in the control wild type plants with no construct, while B and G represent the GUS expression pattern of the 35S constitutive promoter positive control. After the positive control the sets of images represent the typical expression patterns of each construct from the shortest (UTR; C and H) to the longest (full-length; -2018; O and T). C and H show that no GUS expression was produced by the UTR construct in any of the tissues investigated. A similar result was observed for the -340 (E and J) and -1411 (N and S) *EgrCesA8* truncations. The TC-UTR (D and I), -588 (K and P) and -853 (L and Q) showed variable expression in the roots and leaves when compared to the full-length promoter (O and T). The -1036 promoter truncation (M and R) showed a similar pattern GUS expression pattern to the full-length promoter. All images represent the consensus GUS expression pattern produced across three different plants in four/five independently transformed lines.

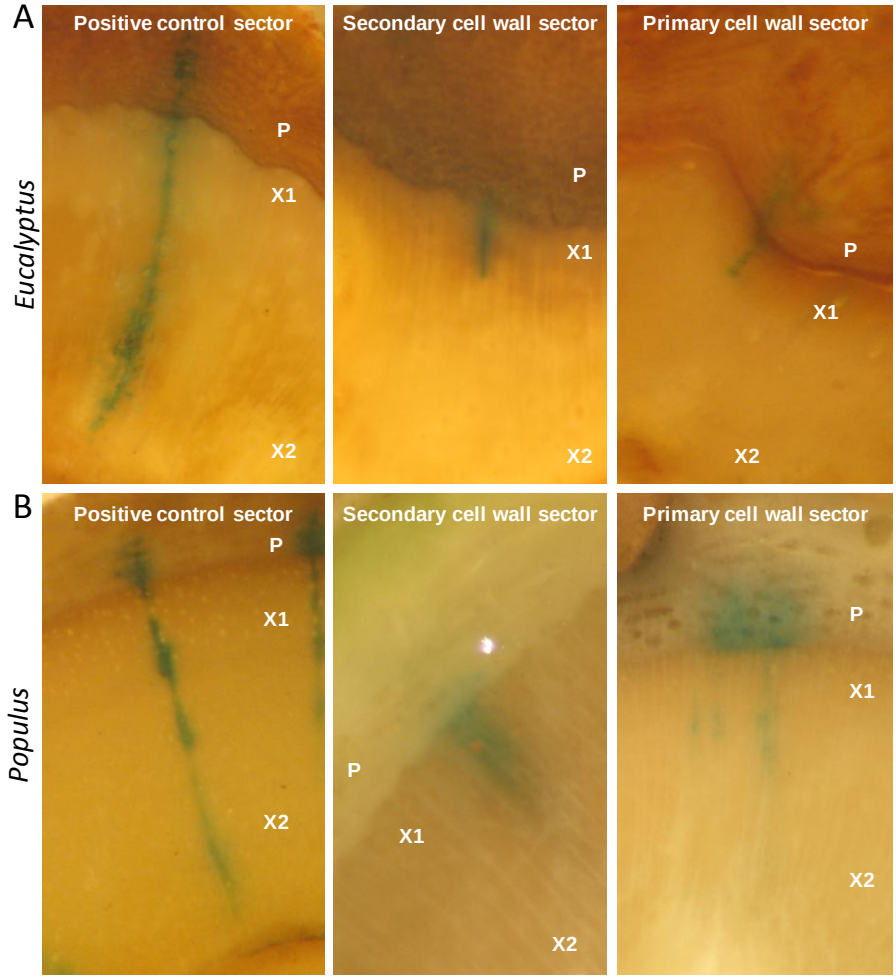


Figure 3.6 Sector types identified with Induced Somatic Sector Analysis (ISSA) of primary and secondary *CesA* and 35S promoter constructs in *Eucalyptus* and *Populus*. A and D represent typical GUS expression sectors produced under the control of a constitutive 35S promoter in *Eucalyptus* and *Populus*, respectively. GUS staining extends from the phloem (P) to the immature xylem (X1) and mature xylem (X2) tissues. B and E represent typical GUS sectors produced under the control of the *CesA* promoters associated with secondary cell wall deposition in *Eucalyptus* and *Populus*, respectively. These sectors display GUS expression predominantly in the X1 region of the stem. C and F give the typical GUS expression pattern observed for the *EgrCesA* promoters associated with primary cell wall formation in *Eucalyptus* and *Populus*, respectively. These sectors show staining in the P and X1 regions of the woody stem.

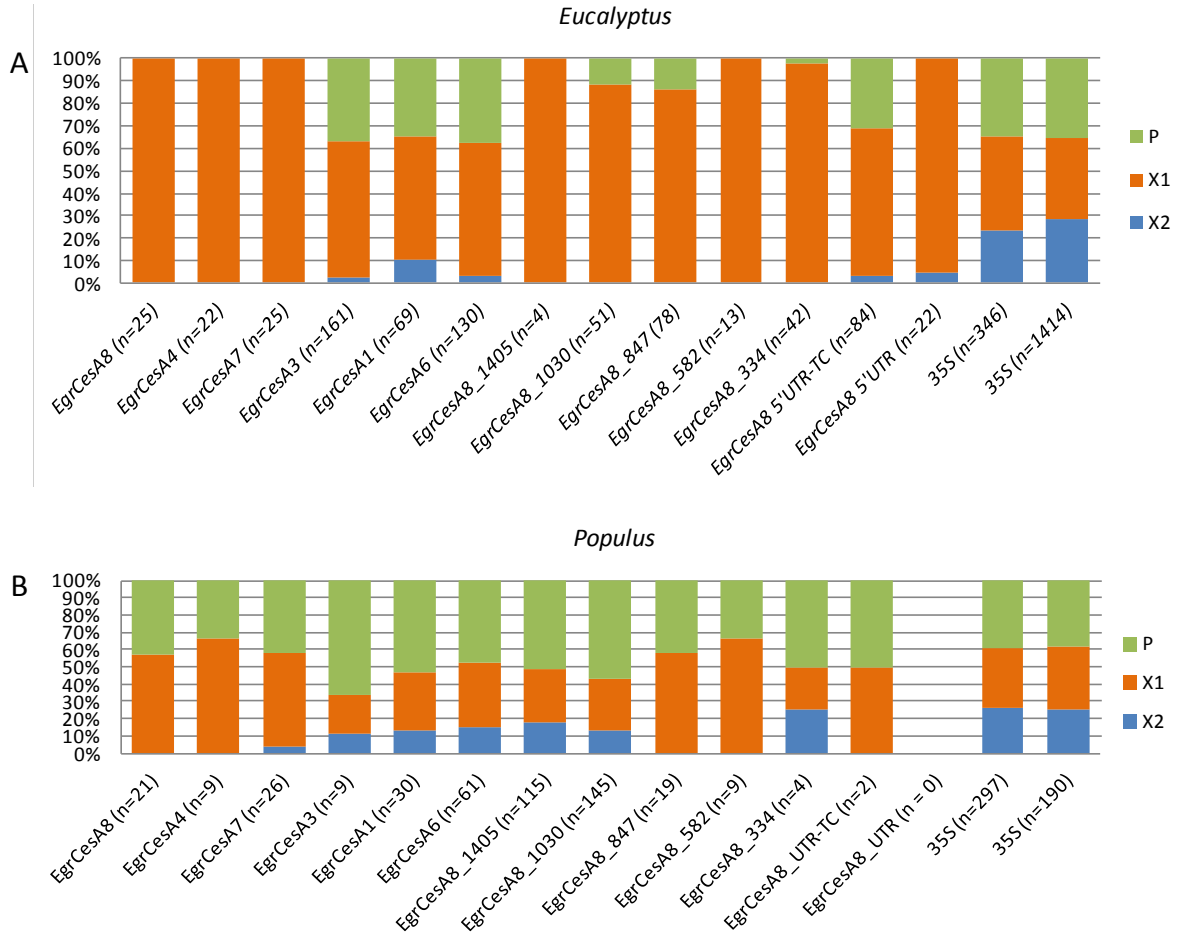
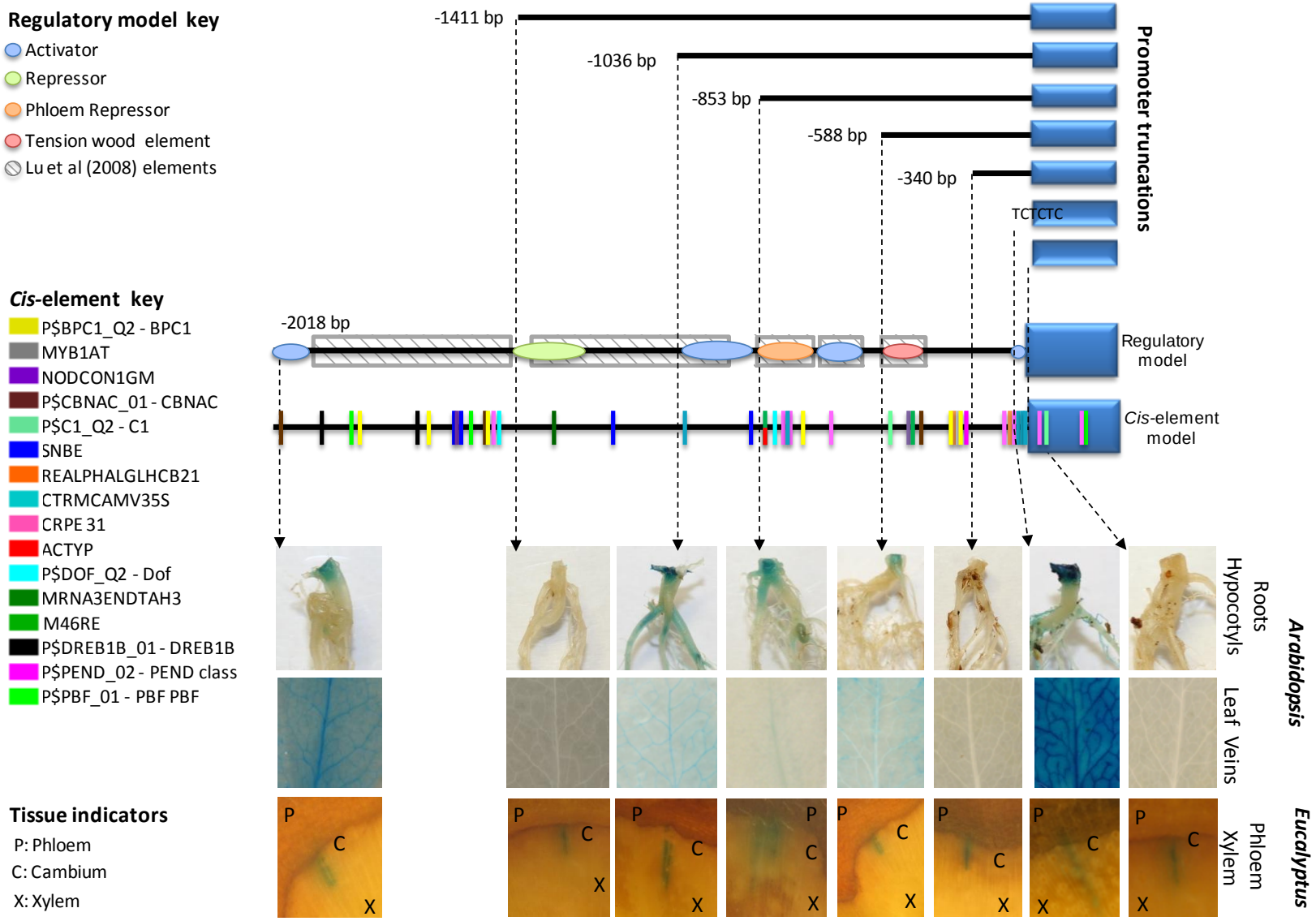


Figure 3.7 Induced Somatic Sector Analysis (ISSA) of the *EgrCesA8* promoter and truncations in *Eucalyptus* (A) and *Populus* (B). The bar graph indicates the frequencies of the phloem (P), immature xylem (X1) and mature xylem (X2) sector counts across all *Eucalyptus* full-length primary (*CesA1*, *3* and *6*) and SCW-associated promoters (*CesA4*, *7* and *8*), and the eight *EgrCesA8* promoter truncates. The total number of sectors counted for each promoter or truncate analyzed is represented by n.

Figure 3.8 Composite figure of *cis*-element, *Arabidopsis-GUS* and *Eucalyptus-GUS* data for the *EgrCesA8* promoter and assorted truncations. Black bars represent the full-length *EgrCesA8* promoter or promoter truncations and blue blocks correspond to the 5'UTR. The vertical dashed lines show which regions of the full-length promoter were removed during truncation. The colour lines on the full-length promoter represent *cis*-elements previously identified in this promoter sequence (Figure 3.1). The *cis*-element key is depicted on the left. First two rows of images below the full-length promoter show the average GUS staining produced by each promoter truncate (full-length promoter to UTR indicated from left to right) in *Arabidopsis*. Bottom images represent the GUS staining patterns produced by the promoter truncations in *Eucalyptus* stem tissues where phloem (P), cambium (C) and xylem (X) are represented**.**



3.10 References

Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) *Arabidopsis* *At*MYC2 (bHLH) and *At*MYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell 15: 63-78

Altobelli G (2012) Bioinformatics applied to gene transcription regulation. J. Mol. Endocrinol. 49: R51-R59

Alwen A, Moreno RMB, Vicente O, Heberle-Bors E (1992) Plant endogenous *β*-glucuronidase activity: how to avoid interference with the use of the *E. coli* *β*-glucuronidase as a reporter gene in transgenic plants. Transgenic Res. 1: 63-70

Aukerman MJ, Schmidt RJ, Burr B, Burr FA (1991) An arginine to lysine substitution in the bZIP domain of an opaque-2 mutant in maize abolishes specific DNA binding. Genes & Dev. 5: 310-320

Baerson SR, Heiden MG, Lamppa GK (1994) Identification of domains in an *Arabidopsis* acyl carrier protein gene promoter required for maximal organ-specific expression. Plant Mol. Biol. 26: 1947-1959

Berger N, Dubreucq B (2012) Evolution goes GAGA: GAGA binding proteins across kingdoms. Biochim. Biophys. Acta. 1819: 863-868

Bernard V, Brunaud V, Lecharny A (2010) TC-motifs at the TATA-box expected position in plant genes: a novel class of motifs involved in the transcription regulation. BMC Genomics 11: 166

Bochman ML, Paeschke K, Zakian VA (2012) DNA secondary structures: stability and function of G-quadruplex structures. Nature Rev. Genet. 13: 770-780

Brown JW (1986) A catalogue of splice junction and putative branch point sequences from plant introns. Nucleic Acids Res. 14: 9549-9559

Burton RA, Shirley NJ, King BJ, Harvey AJ, Fincher GB (2004) The *CesA* gene family of barley. Quantitative analysis of transcripts reveals two groups of co-expressed genes. Plant Physiol. 134: 224-236

Busk PK, Pages M (1998) Regulation of abscisic acid-induced transcription. Plant Mol. Biol. 37: 425-435

Buske FA, Mattick JS, Bailey TL (2011) Potential *in vivo* roles of nucleic acid triple-helices. RNA Biol. 8: 427-439

Cassan-Wang H, Goue N, Saidi MN, Legay S, Sivadon P, Goffner D, Grima-Pettenati J (2013) Identification of novel transcription factors regulating secondary cell wall formation in *Arabidopsis*. Front. Plant Sci. 4:189

Cer RZ, Bruce KH, Mudunuri US, Yi M, Volfovsky N, Luke BT, Bacolla A, Collins JR, Stephens RM (2011) Non-B DB: a database of predicted non-B DNA-forming motifs in mammalian genomes. Nucleic Acids Res. 39: D383-D391

Creux NM, Ranik M, Berger DK, Myburg AA (2008) Comparative analysis of orthologous cellulose synthase promoters from *Arabidopsis*, *Populus* and *Eucalyptus*: evidence of conserved regulatory elements in angiosperms. New Phytol. 179: 722-737

Das P, Joshi NC (2011) Minor modifications in obtainable *Arabidopsis* floral dip method enhances transformation efficiency and production of homozygous transgenic lines harboring a single copy of transgene. Advances Biosci. Biotechnol. 2: 59-67

Degenhardt J, Tobin EM (1996) A DNA binding activity for one of two closely defined phytochrome regulatory elements in an Lhcb promoter is more abundant in etiolated than in green plants. Plant Cell 8: 31-41

Demura T, Ye Z-H (2010) Regulation of plant biomass production. Curr. Opin. Plant Biol. 13: 298-303

Ding J, Hu H, Li X (2012) Thousands of *cis*-regulatory sequence combinations are shared by *Arabidopsis* and poplar. Plant Physiol. 158: 145-155

Djerbi S, Lindskog M, Arvestad L, Sterky F, Teeri TT (2005) The genome sequence of black cottonwood (*Populus trichocarpa*) reveals 18 conserved cellulose synthase (*CesA*) genes. Planta 221: 739-746

Gardiner J, Sherr I, Scarpella E (2010) Expression of DOF genes identifies early stages of vascular development in *Arabidopsis* leaves. Int. J. Dev. Biol. 54: 1389-1396

Gerhardt IR, Filippi SB, Okura V, Coutinho J, Rizzato AP, Gui K, Vessali N, Pontes JH, Cordeiro T, Silva SM (2011) Overexpression of walldof transcription factor increases secondary wall deposition and alters carbon partitioning in poplar. BMC Proceedings 5(Suppl 7): O35

Giuliano G, Pichersky E, Malik V, Timko M, Scolnik P, Cashmore A (1988) An evolutionarily conserved protein binding sequence upstream of a plant light-regulated gene. Proc. Natl. Acad. Sci. USA 85: 7089-7093

Goicoechea M, Lacombe E, Legay S, Mihaljevic S, Rech P, Jauneau A, Lapierre C, Pollet B, Verhaegen D, Chaubet-Gigot N, Grima-Pettenati J (2005) *Eg*MYB2, a new

transcriptional activator from *Eucalyptus* xylem, regulates secondary cell wall formation and lignin biosynthesis. Plant J. 43: 553-567

Goldsbrough AP, Albrecht H, Stratford R (1993) Salicylic acid-inducible binding of a tobacco nuclear protein to a 10 bp sequence which is highly conserved amongst stress-inducible genes. Plant J. 3: 563-571

Gorshkova T, Brutch N, Chabbert B, Deyholos M, Hayashi T, Lev-Yadun S, Mellerowicz EJ, Morvan C, Neutelings G, Pilate G (2012) Plant fiber formation: state of the art, recent and expected progress, and open questions. CRC Cr. Rev. Plant Sci. 31: 201-228

Grace ML, Chandrasekharan MB, Hall TC, Crowe AJ (2004) Sequence and spacing of TATA-box elements are critical for accurate initiation from the *β*-phaseolin promoter. J. Biol. Chem. 279: 8102-8110

Gray-Mitsumune M, Molitor EK, Cukovic D, Carlson JE, Douglas CJ (1999) Developmentally regulated patterns of expression directed by poplar PAL promoters in transgenic tobacco and poplar. Plant Mol. Biol. 39: 657-669

Grotewold E, Drummond BJ, Bowen B, Peterson T (1994) The myb-homologous P gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. Cell 76: 543-553

Guerriero G, Spadiut O, Kerschbamer C, Giorno F, Baric S, Ezcurra I (2012) Analysis of cellulose synthase genes from domesticated apple identifies collinear genes WDR53 and *CesA8A*: partial co-expression, bicistronic mRNA, and alternative splicing of *CESA8A*. J. Exp. Bot. 63: 6045-6056

Guo Y, Qin G, Gu H, Qu L-J (2009) Dof5.6/HCA2, a Dof transcription factor gene, regulates interfascicular cambium formation and vascular tissue development in *Arabidopsis*. Plant Cell 21: 3518-3534

Hamann T, Osborne E, Youngs HL, Misson J, Nussaume L, Somerville C (2004) Global expression analysis of *CESA* and *CSL* genes in *Arabidopsis*. Cellulose 11: 279-286

Hatton D, Sablowski R, Yung MH, Smith C, Schuch W, Bevan M (1995) Two classes of *cis* sequences contribute to tissue-specific expression of a PAL2 promoter in transgenic tobacco. Plant J. 7: 859-876

Hazen S (2012) Transcriptional regulation of lignin and cellulose biosynthesis in *Brachypodium distachyon*. Plant and Animal Genome XX Conference (January 14-18, 2012). Plant and Animal Genome

Hertzberg M, Aspeborg H, Schrader J, Andersson A, Erlandsson R, Blomqvist K, Bhalerao R, Uhlen M, Teeri TT, Lundeberg J, Sundberg B, Nilsson P, Sandberg G (2001) A

transcriptional roadmap to wood formation. Proc. Natl. Acad. Sci. USA. 98: 14732-14737

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

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

Hudson ME, Quail PH (2003) Identification of promoter motifs involved in the network of phytochrome A-regulated gene expression by combined analysis of genomic sequence and microarray data. Plant Physiol. 133: 1605-1616

Hughes TR (2011) A handbook of transcription factors. Springer, London

Hussey SG, Mizrachi E, Spokevicius AV, Bossinger G, Berger DK, Myburg AA (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 Biol. 11: 173

Izawa T, Foster R, Chua N-H (1993) Plant bZIP protein DNA binding specificity. J. Mol. Biol. 230: 1131-1144

Jefferson RA, Kavanagh TA, Bevan MW (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. 6: 3901-3907

Jiao Y, Lau OS, Deng XW (2007) Light-regulated transcriptional networks in higher plants. Nature Rev. Genet. 8: 217-230

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473: 97-100

Kaufmann K, Pajoro A, Angenent GC (2010) Regulation of transcription in plants: mechanisms controlling developmental switches. Nature Rev. Genet. 11: 830-842

Kim HS, Park BO, Yoo JH, Jung MS, Lee SM, Han HJ, Kim KE, Kim SH, Lim CO, Yun D-J (2007) Identification of a calmodulin-binding NAC protein as a transcriptional repressor in *Arabidopsis*. J. Biol. Chem. 282: 36292-36302

Kim M-H, Sasaki K, Imai R (2009) Cold shock domain protein 3 regulates freezing tolerance in *Arabidopsis thaliana*. J. Biol. Chem. 284: 23454-23460

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

Ko J-H, Beers E, Han K-H (2006) Global comparative transcriptome analysis identifies gene network regulating secondary xylem development in *Arabidopsis thaliana*. Mol. Gen. Genomics 276: 517-531

Ko JH, Han KH, Park S, Yang JM (2004) Plant body weight-induced secondary growth in *Arabidopsis* and its transcription phenotype revealed by whole-transcriptome profiling. Plant Physiol. 135: 1069-1083

Kooiker M, Airoldi CA, Losa A, Manzotti PS, Finzi L, Kater MM, Colombo L (2005) BASIC PENTACYSTEINE1, a GA binding protein that induces conformational changes in the regulatory region of the homeotic *Arabidopsis* gene SEEDSTICK. Plant Cell 17: 722-729

Kumar M, Thammannagowda S, Bulone V, Chiang V, Han K-H, Joshi CP, Mansfield SD, Mellerowicz E, Sundberg B, Teeri T, Ellis BE (2009) An update on the nomenclature for the cellulose synthase genes in *Populus*. Trends Plant Sci. 14: 248-254

Kutach AK, Kadonaga JT (2000) The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. Mol. Cell Biol. 20: 4754-4764

Lagrange T, Gauvin S, Yeo H-J, Mache R (1997) S2F, a leaf-specific trans-acting factor, binds to a novel *cis*-acting element and differentially activates the RPL21 gene. Plant Cell 9: 1469-1479

Lawton MA, Dean SM, Dron M, Kooter JM, Kragh KM, Harrison MJ, Yu L, Tanguay L, Dixon RA, Lamb CJ (1991) Silencer region of a chalcone synthase promoter contains multiple binding sites for a factor, SBF-1, closely related to GT-1. Plant Mol. Biol. 16: 235-249

Lazo G, Stein P, Ludwig R (1991) A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. Biotechnology 9: 963-967

Lee JT (2012) Epigenetic regulation by long noncoding RNAs. Science 338: 1435-1439

Lelli K, Slattery M, Mann R (2012) Disentangling the many layers of eukaryotic transcriptional regulation. Annu. Rev. Genet. 46: 43-64

Lenhard B, Sandelin A, Carninci P (2012) Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nature Rev. Genet. 13: 233-245

Lev-Yadun S (1997) Fibres and fibre-sclereids in wild-type *Arabidopsis thaliana*. Ann. Bot. 80: 125-129

Liu Z-B, Ulmasov T, Shi X, Hagen G, Guilfoyle TJ (1994) Soybean GH3 promoter contains multiple auxin-inducible elements. Plant Cell 6: 645-657

Lu S, Li L, Yi X, Joshi CP, Chiang VL (2008) Differential expression of three *Eucalyptus* secondary cell wall-related cellulose synthase genes in response to tension stress. J. Exp. Bot. 59: 681-695

McCarthy RL, Zhong R, Ye Z-H (2009) MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in Arabidopsis. Plant Cell Physiol. 50: 1950-1964

McCarthy RL, Zhong R, Ye Z-H (2011) Secondary wall NAC binding element (SNBE), a key *cis*-acting element required for target gene activation by secondary wall NAC master switches. Plant Signal. Behav. 6: 1282-1285

Meister RJ, Williams LA, Monfared MM, Gallagher TL, Kraft EA, Nelson CG, Gasser CS (2004) Definition and interactions of a positive regulatory element of the *Arabidopsis* INNER NO OUTER promoter. Plant J. 37: 426-438

Mena M, Cejudo FJ, Isabel-Lamoneda I, Carbonero P (2002) A role for the DOF transcription factor BPBF in the regulation of gibberellin-responsive genes in barley aleurone. Plant Physiol. 130: 111-119

Mertin S, McDowall SG, Harley VR (1999) The DNA-binding specificity of SOX9 and other SOX proteins. Nucleic Acids Res. 27: 1359-1364

Mizrachi E, Hefer CA, Ranik M, Joubert F, Myburg AA (2010) *De novo* assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by Illumina mRNA-Seq. BMC Genomics 11: 681

Mizrachi E, Mansfield SD, Myburg AA (2012) Cellulose factories: advancing bioenergy production from forest trees. New Phytol. 194: 54-62

Mohanty B, Krishnan S, Swarup S, Bajic VB (2005) Detection and preliminary analysis of motifs in promoters of anaerobically induced genes of different plant species. Ann. Bot. 96: 669-681

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

Monfared MM, Simon MK, Meister RJ, Roig-Villanova I, Kooiker M, Colombo L, Fletcher JC, Gasser CS (2011) Overlapping and antagonistic activities of BASIC PENTACYSTEINE genes affect a range of developmental processes in *Arabidopsis*. Plant J. 66: 1020-1031

Morita A, Umemura T-a, Kuroyanagi M, Futsuhara Y, Perata P, Yamaguchi J (1998) Functional dissection of a sugar-repressed α-amylase gene (RAmy1A) promoter in rice embryos. FEBS Lett. 423: 81-85

Nagao RT, Goekjian VH, Hong JC, Key JL (1993) Identification of protein-binding DNA sequences in an auxin-regulated gene of soybean. Plant Mol. Biol. 21: 1147-1162

Nakashima K, Ito Y, Yamaguchi-Shinozaki K (2009) Transcriptional regulatory networks in response to abiotic stresses in *Arabidopsis* and grasses. Plant Physiol. 149: 88-95

Niu X, Adams CC, Workman JL, Guiltinan MJ (1996) Binding of the wheat basic leucine zipper protein EmBP-1 to nucleosomal binding sites is modulated by nucleosome positioning. Plant Cell 8: 1569-1587

Oh D-H, Kwon C-S, Sano H, Chung W-I, Koizumi N (2003) Conservation between animals and plants of the *cis*-acting element involved in the unfolded protein response. Biochem. Bioph. Res. Co. 301: 225-230

Ohtsubo N, Iwabuchi M (1994) The conserved 3′-flanking sequence, AATGGAAATG, of the wheat histone H3 gene is necessary for the accurate 3′-end formation of mRNA. Nucleic Acids Res. 22: 1052-1058

Pauli S, Rothnie HM, Chen G, He X, Hohn T (2004) The cauliflower mosaic virus 35S promoter extends into the transcribed region. J. Virol. 78: 12120-12128

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

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

Raiber E-A, Kranaster R, Lam E, Nikan M, Balasubramanian S (2012) A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. Nucleic Acids Res. 40: 1499-1508

Ramirez-Parra E, Fründt C, Gutierrez C (2003) A genome-wide identification of E2F-regulated genes in *Arabidopsis*. Plant J. 33: 801-811

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

Richmond TA, Somerville CR (2000) The cellulose synthase superfamily. Plant Physiol. 124: 495-498

Riechmann JL, Wang M, Meyerowitz EM (1996) DNA-binding properties of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. Nucleic Acids Res. 24: 3134-3141

Roberts AW, Bushoven JT (2007) The cellulose synthase (CESA) gene superfamily of the moss *Physcomitrella patens*. Plant Mol. Bio. 63: 207 - 219

Rosellini D (2012) Selectable markers and reporter genes: a well furnished toolbox for plant science and genetic engineering. CRC Cr. Rev. Plant Sci. 31: 401-453

Rushmore TH, Morton MR, Pickett CB (1991) The antioxidant responsive element. Activation by oxidative stress and identification of the DNA consensus sequence required for functional activity. J. Biol. Chem. 266: 11632-11639

Sambrook J, Russell DW (2001) Molecular cloning: A laboratory manual. Cold Springs Harbour Press, New York

Samuga A, Joshi CP (2004) Differential expression patterns of two new primary cell wall-related cellulose synthase cDNAs, *PtrCesA6* and *PtrCesA7* from aspen trees. Gene 334: 73-82

Sander M, Hsieh T-s (1985) *Drosophila* topoisomerase II double-strand DNA cleavage: analysis of DNA sequence homology at the cleavage site. Nucleic Acids Res. 13: 1057-1072

Schrader J, Nilsson J, Mellerowicz E, Berglund A, Nilsson P, Hertzberg M, Sandberg G (2004) A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. Plant Cell 16: 2278-2292

Seo JS, Koo YJ, Jung C, Yeu SY, Song JT, Kim J-K, Choi Y, Lee JS, Do Choi Y (2013) Identification of a novel jasmonate-responsive element in the *AtJMT* promoter and its binding protein for *At*JMT repression. PloS ONE 8: e55482

Shirsat A, Wilford N, Croy R, Boulter D (1989) Sequences responsible for the tissue-specific promoter activity of a pea legumin gene in tobacco. Mol. Gen. Genet. 215: 326-331

Simonini S, Roig-Villanova I, Gregis V, Colombo B, Colombo L, Kater MM (2012) BASIC PENTACYSTEINE proteins mediate MADS domain complex binding to the DNA for tissue-specific expression of target genes in *Arabidopsis*. Plant Cell 24: 4163-4172

Sinden RR (1994) DNA Structure and function. Academic Press, Waltham

Smale ST (1997) Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. Biochim. Biophy. Acta. 1351: 73-88

Stormo GD, Zhao Y (2010) Determining the specificity of protein–DNA interactions. Nature Rev. Genet. 11: 751-760

Stougaard J, Jørgensen J-E, Christensen T, Kühle A, Marcker KA (1990) Interdependence and nodule specificity of *cis*-acting regulatory elements in the soybean leghemoglobin lbc3 and N23 gene promoters. Mol. Gen. Genet. 220: 353-360

Suzuki M, Kao CY, McCarty DR (1997) The conserved B3 domain of VIVIPAROUS1 has a cooperative DNA binding activity. Plant Cell 9: 799-807

Suzuki S, Li LG, Sun YH, Chiang VL (2006) The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthase-like genes in *Populus trichocarpa*. Plant Physiol. 142: 1233-1245

Tang Z, Sadka A, Morishige DT, Mullet JE (2001) Homeodomain leucine zipper proteins bind to the phosphate response domain of the soybean VspB tripartite promoter. Plant Physiol. 125: 797-809

Thomas-Chollier M, Sand O, Turatsinze J-V, Janky Rs, Defrance M, Vervisch E, Brohee S, van Helden J (2008) RSAT: regulatory sequence analysis tools. Nucleic Acids Res. 36: W119-W127

Thomas MS, Flavell RB (1990) Identification of an enhancer element for the endosperm-specific expression of high molecular weight glutenin. Plant Cell 2: 1171-1180

Tremousaygue D, Manevski A, Bardet C, Lescure N, Lescure B (2002) Plant interstitial telomere motifs participate in the control of gene expression in root meristems. Plant J. 20: 553-561

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, de Peer YV, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313: 1596-1604

Van Beveren K, Spokevicius AV, Tibbits J, Wang Q, Bossinger G (2006) Transformation of cambial tissue *in vivo* provides efficient means for Induced Somatic Sector Analysis (ISSA) and gene testing in stems of woody plants species. Funct. Plant Biol. 33: 629-638

Van Loo P, Marynen P (2009) Computational methods for the detection of *cis*-regulatory modules. Brief. Bioinform. 10: 509-524

Vieweg MF, Frühling M, Quandt H-J, Heim U, Bäumlein H, Pühler A, Küster H, Perlick AM (2004) The promoter of the *Vicia faba L.* leghemoglobin gene VfLb29 is specifically activated in the infected cells of root nodules and in the arbuscule-containing cells of mycorrhizal roots from different legume and nonlegume plants. Mol. Plant Microbe In. 17: 62-69

Weigel D, Glazebrook J (2002) *Arabidopsis*: A Laboratory Manual. Laboratory Press, Cold Spring Harbor

Wingender E (2008) The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. Brief Bioinform. 9: 326-332

Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Prüß M, Reuter I, Schacherer F (2000) TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Res. 28: 316-319

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

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

Yanagisawa S (2002) The Dof family of plant transcription factors. Trends Plant Sci. 7: 555-560

Yanagisawa S (2004) Dof domain proteins: plant-specific transcription factors associated with diverse phenomena unique to plants. Plant Cell Physiol. 45: 386-391

Yanagisawa S, Sheen J (1998) Involvement of maize Dof zinc finger proteins in tissue-specific and light-regulated gene expression. Plant Cell 10: 75-89

Yang C, Bolotin E, Jiang T, Sladek FM, Martinez E (2007) Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene 389: 52-65

Yin Y, Chen L, Beachy R (1997) Promoter elements required for phloem-specific gene expression from the RTBV promoter in rice. Plant J. 12: 1179-1188

Yin Y, Huang J, Xu Y (2009) The cellulose synthase superfamily in fully sequenced plants and algae. BMC Plant Biol. 9: 99

Yokoyama KD, Thorne JL, Wray GA (2011) Co-ordinated genome-wide modifications within proximal promoter *cis*-regulatory elements during vertebrate evolution. Genome Biol. Evol. 3: 66-74

Zhang L, Zuo K, Zhang F, Cao Y, Wang J, Zhang Y, Sun X, Tang K (2006) Conservation of noncoding microsatellites in plants: implication for gene regulation. BMC Genomics 7: 323

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

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

Zhong R, McCarthy RL, Lee C, Ye Z-H (2011) Dissection of the transcriptional program regulating secondary wall biosynthesis during wood formation in poplar. Plant Physiol. 157: 1452-1468

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

Zuo Y-C, Li Q-Z (2011) Identification of TATA and TATA-less promoters in plant genomes by integrating diversity measure, GC-Skew and DNA geometric flexibility. Genomics 97: 112-120

Chapter 4

Identification of transcription factors interacting with regulatory modules in the *Eucalyptus CesA8* promoter

N.M. Creux[1,2], C. Martiz-Olivier[1], A.A. Myburg[1,2]

Department of Genetics[1], Faculty of Natural and Agricultural Sciences, Forestry Argicultural Biotechnology Institute (FABI)[2], University of Pretoria, Pretoria, 2000, South Africa

This chapter was prepared in the format of a manuscript for a research journal (e.g. *Planta*). I performed all of the bait vector constructions, *cis*-element analysis, yeast transformations and yeast-1-hybrid screening. Prof. Zander Myburg and Dr. Christine Maritz-Olivier provided advice, direction and supervision in planning of the project. They also provided direction in the interpretation of the results and provided critical revision of the Chapter. Mr Jonathan Botha identified and isolated *Eucalyptus* transcription factors which were used as preys during the Y1H screens for this chapter as part of his MSc study.

4.1 Abstract

Eucalyptus is one of the major crops for cellulose production, which is used by a number of industries such as pulp, paper, textiles, biofuel and food. Currently, there is little information on how cellulose biosynthesis is regulated in this important plantation species. Cellulose is deposited in the plant cell wall by a large membrane bound cellulose synthase (CesA) complex which consists of different CesA subunits. The CesA subunits are coded for by different *CesA* genes which have highly specific expression patterns. In general, *CesA* expression can be divided into two main expression groups, where three *CesA* genes are expressed during secondary cell wall (SCW) formation and a different set are expressed during primary cell wall formation. *CesA* gene expression can also vary depending on tissue type, developmental stage or environmental factors. In Chapter 2 several regions of the SCW-related *EgrCesA8* promoter where found to contain conserved *cis*-elements and could modulate reporter gene expression. We used these regions of the *EgrCesA8* promoter to screen a panel of 15 different *Eucalyptus* SCW-related transcription factors for interaction in a yeast-1-hybrid assay. We identified five SCW-related transcription factors which interacted with a previously identified *EgrCesA8 cis*-regulatory module (CRM). *Egr*MYB31 (homolog of *At*MYB46) and a homolog of *At*C3H14 interacted with the 5'UTR of *EgrCesA8.* These findings suggest that parts of the SCW regulatory network are conserved in *Eucalyptus* and that several of the previously identified regions of the *EgrCesA8* promoter, which harbour conserved *cis*-elements, interact with transcription factors. *EgrCesA8* was also identified as a novel target for the poorly characterized transcription factors *Egr*ZincFinger-A, *Egr*MYB80 and *Egr*MYB87 (*At*C3H14, *At*MYB52 and *At*MYB54 homologs).

Keywords: secondary cell wall (SCW), regulatory network, transcription factors, yeast-1-hybrid, cellulose synthase, *cis*-regulatory modules (CRM)

4.2 Introduction

Cellulose is a highly abundant biopolymer, which forms the basis for many different industrial applications. *Eucalyptus* is currently one of the major cellulose producing crops and it is important to understand cellulose deposition and regulation in this plantation species. Cellulose is deposited, in the plant cell wall, by a large membrane bound protein complex consisting of several different cellulose synthase (CesA) subuints (reviewed in Lei et al. 2012). *CesA* genes have been isolated from many different plant species including model species (*Arabidopsis* and *Populus*) and crop species such as *Eucalyptus*, barley, rice and maize (Richmond and Somerville 2000; Hazen et al. 2002; Appenzeller et al. 2004; Burton et al. 2004; Djerbi et al. 2005; Ranik and Myburg 2006). Most plant genomes have multiple *CesA* genes, ranging from 10 in *Arabidopsis* to 18 in *Populus* (Richmond and Somerville 2000; Djerbi et al. 2005). Many studies have focused on identifying the genes involved in cellulose deposition and assess their expression under different conditions, stages or tissue types (reviewed in Taylor 2008; Mizrachi et al. 2012). Recently, studies have begun to identify and classify the transcription factors which regulate the biosynthetic genes involved in primary and secondary cell wall (SCW) formation such as the *CesA* genes (reviewed in Demura and Ye 2010; Zhong et al. 2010a). However, there is little information on the promoter regions of the *CesA* genes and how they interact with the various SCW-associated transcription factors.

Previous studies have revealed that the *CesA* genes have very intricate expression patterns which can vary depending on the developmental stage, tissue type and environment. Initial studies of *Arabidopsis* stem development, revealed highly specific expression patterns for different cellulose synthase genes, where three genes (*AtCesA4*, *AtCesA7* and *AtCesA8*) are associated with SCW formation and others are expressed during primary cell wall formation (Hamann et al. 2004). The distinct primary and SCW expression patterns of the

different *CesA* genes is conserved across most plant taxa and plays an important role during development (Appenzeller et al. 2004; Burton et al. 2004; Suzuki et al. 2006; Obembe et al. 2009). The combination of *CesA* genes being expressed in the cell wall is also tissue-type dependant (Betancur et al. 2010; Stork et al. 2010; Harpaz-Saad et al. 2011). In stems *CesA4*, *7* and *8* are involved in SCW formation, but in seed coats *CesA5* appears to play an important role in cell wall thickening (Harpaz-Saad et al. 2011). Generally *CesA1*, *3* and *6* are known as primary cell wall *CesAs*, however, *CesA6* expression can be substituted by other *CesAs* such as *CesA9* which is active in mature roots (Desprez et al. 2007). Aside from the developmental and tissue-specificity observed for different *CesA* genes, it has also been noted that *CesA* gene expression can vary depending on biotic and abiotic stresses. Mechanical stress on a plant stem or branch can result in changes in cellulose deposition and several *CesA* genes have modified expression under this condition (Bhandari et al. 2006; Lu et al. 2008). *CesA* genes are also down-regulated in drought tolerant rice cultivars when compared to drought sensitive cultivars suggesting some *CesA* regulation is involved during drought response (Cal et al. 2013). Hernandez-Blanco et al. (2007) also found that *Arabidopsis CesA* knockout mutants showed enhanced disease resistance and this could indicate re-organisation of cell wall during pathogen attack. The highly co-ordinated and specific expression patterns of the different *CesA* genes suggests there are many regulatory mechanisms modulating this expression, however little is known about the regulatory elements that maybe involved in these different processes.

Thus far there have only been a few studies on the regulatory regions of the *CesA* genes (Creux et al. 2008; Lu et al. 2008; Wu et al. 2009; Chapter 3 and Supplementary file 3.2). Different studies have shown that transgenic *CesA* promoters from *Eucalyptus* or cotton can produce tissue-specific expression patterns in model species such as *Arabidopsis* and tobacco, which suggests there is some conservation of the *cis*-regulatory modules regulating

CesA gene expression (Creux et al. 2008; Lu et al. 2008; Wu et al. 2009). A number of putative *cis*-regulatory element sequences were conserved in the *CesA* promoters of *Arabidopsis*, *Populus* and *Eucalyptus* and could play an important regulatory role during *CesA* gene expression (Creux et al. 2008). *Cis*-elements often cluster together in a promoter region and facilitate the interaction with transcription factor complexes and these clusters are often referred to as *cis*-regulatory modules (CRMs; Levine and Davidson 2005). Using this feature of *cis*-elements, Ding et al. (2012) performed an extensive *in silico cis*-element analysis in *Arabidopsis* and *Populus* promoters and could identify *cis*-element groupings that were enriched in *CesA* promoters and conserved across both species. Many putative *cis*-elements have been identified by various *in silico* methods and several of these are conserved across large evolutionary distances, but very few of these regions have been experimentally verified or have been linked to the transcription factors regulating SCW formation.

Experimentally tested *cis*-elements are scarce throughout the SCW regulatory network with only 11 unique sequences reported in literature thus far (Patzlaff et al. 2003; Goicoechea et al. 2005; Shen et al. 2009; Ohashi-Ito et al. 2010; Rahantamalala et al. 2010; Winzell et al. 2010; Zhong et al. 2010b; Wang et al. 2011; Xie et al. 2011; Zhong and Ye 2012; Kim et al. 2013). In *Arabidopsis* Kim et al. (2013) identified the M46RE binding site (RKTWGGTR) in all of the SCW-associated *CesA* promoters, and that each promoter sequence directly interacted with the SCW master regulator, *At*MYB*46*. Another master regulator, *AtVND7*, was also shown to interact directly with the SCW-associated *CesA* promoters but the *cis*-element facilitating this interaction is still disputed (Zhong et al. 2010b; Yamaguchi et al. 2011). Zhong et al. (2010b) identified a semi-palindromic *cis*-element (SNBE) in the promoters of several structural genes involved in SCW deposition, which were bound by several of the NAC-transcription factor family master regulators such as *At*SND1, *At*VND6, *At*VND7, *At*NST1 and *At*NST2. M46RE and SNBE are the only *cis*-elements in the

SCW regulatory network with extensive experimental data (Zhong et al. 2010b; McCarthy et al. 2011; Zhong and Ye 2012; Kim et al. 2013) most likely due to the low-throughput nature of *cis*-element identification and verification techniques. Hazen (2012) estimated that at least 30 different transcription factors can bind to a single *CesA* promoter in *Brachypodium* and this suggests that in the network one would expect a plethora of different binding sites.

Transactivation studies, while not necessarily indicating direct DNA-protein interactions, have identified several transcription factors in the SCW regulatory network which activate *CesA* gene expression (reviewed in Chapter 1, Figure 1.1). The NAC transcription factor, *At*SND2 has been linked to *AtCesA8* activation in both transactivation and microarray studies (Zhong et al. 2008; Hussey et al. 2011). Similarly, a number of other transcription factors, including *At*SND3, *At*MYB103, *At*MYB52, *At*MYB54, *At*C3H14 and *At*KNAT7 act as transactivators of the SCW-associated *CesA* promoters (Zhong et al. 2008; Ko et al. 2009). *At*KNAT7 mutants have a unique *irx* (*irregular xylem*) mutant phenotype with the vessels having the typical "wrinkled" appearance, while the fiber cells show thickened cell walls suggesting that *At*KNAT7 is a negative regulator of SCW formation in fiber cells (Li et al. 2012). *At*KNAT7 is known to be involved in protein-protein interactions with a number of different partners including the ovate family proteins (OFP), knotted-like *Arabidopsis thaliana* (*At*KNAT) proteins and the STM (Shoot meristemless) protein (Hackbusch et al. 2005). *At*KNAT7 complexes with either *At*OFP4 or *At*MYB75 and leads to strong repression of the genes they regulate, suggesting that *At*KNAT7 is often part of repressor complexes (Bhargava et al. 2010; Li et al. 2011). Considering the extent to which transcription factors act together with co-factors (Ramsay and Glover 2005; Immink et al. 2009; Kaufmann et al. 2010), it is highly likely that there are many other interactions and co-factors still to be added to our understanding of the SCW regulatory network.

The plant cell wall is the initial point of contact between a plant and its environment and has to respond to different external cues including biotic and abiotic stress (reviewed in Seifert and Blaukopf 2010). The main cell wall components including lignin, cellulose and hemicellulose can have altered structure or content during response to different factors such as pathogen attack or environmental perturbations (Hammerschmidt et al. 1984; Hernández-Blanco et al. 2007; Tronchet et al. 2010; Cal et al. 2013). Recently, reports have linked several of the SCW transcription factors to stress responses. Ramirez et al. (2011a; 2011b) found that *At*MYB46 knockout mutants showed strong resistance to *Botrytis cinerea*. The authors propose that *At*MYB46 is a conditional repressor, which under normal conditions activates cell wall biosynthesis and represses the cell wall bound peroxidase III class resistance genes. Upon infection with *B. cinerea*, *At*MYB46 switches to an activator of resistance genes possibly by changing the interacting cofactors and this allows for slight remodelling of the cell wall (Ramírez et al. 2011a). *At*MYB83 is a functional ortholog of *At*MYB46 (Zhong and Ye 2012) and it is possible that during infections *At*MYB46 acts as an activator of the resistance genes and *At*MYB83 assumes the role as a SCW master regulator, although this remains to be investigated.

An example of response by the SCW regulatory network to abiotic stress was identified during the over-expression of *At*MYB52. *At*MYB52 is activated by several of the SCW master regulators including *At*MYB46 (Zhong and Ye 2012). However, no transactivation of the cell wall biosynthetic genes was observed for *At*MYB52 and it was suggested that it either required certain cofactors to function or may have a repressor function (Ko et al. 2009). Park et al. (2011) found that the over-expression of *At*MYB52 produced plants which were drought resistant but also showed a down-regulation of the cell wall biosynthetic genes, confirming the repressor role. These findings suggest that the SCW regulatory network may overlap with other key regulatory networks involved in biotic and

abiotic stress and further studies are required to dissect the possible pleiotropic effects of each gene within the network.

Eucalyptus is a valuable forestry species being one of the major plantation crops in the southern hemisphere. However, there is very little information available on the regulation of *Eucalyptus* cellulose synthase genes in terms of protein-DNA interactions and the extent to which the SCW regulatory network is conserved in this economically important crop (Patzlaff et al. 2003; Goicoechea et al. 2005; Legay et al. 2007; Creux et al. 2008; Lu et al. 2008; Legay et al. 2010; Zhong et al. 2010a). The current chapter aimed to identify which *Eucalyptus* transcription factors, interacted with the different *EgrCesA8* promoter regions identified by truncation analyses (Chapter 3). The main objectives were firstly, to identify which putative *Eucalyptus* homologs of the SCW-associated transcription factors interact with the *EgrCesA8* CRM and other regulatory regions identified in Chapter 3. Secondly, we investigated whether conserved SCW-related *cis*-elements could be identified in the transcription factor binding regions of the promoters from 13 different *Eucalyptus* species. The final objective was to test the effect of the $CT_{(11)}$-microsatellite (Chapter 3) on interactions between the transcription factors and the 5'UTR. Using yeast-1-hybrid (Y1H) screening we identified six putative *Eucalyptus* homologs of key SCW-associated transcription factors that interacted with different regions in the *EgrCesA8* promoter. Combining different lines of evidence such as GUS assays, conserved *cis*-element analysis and Y1H data allowed us to construct a model for the transcriptional regulation of *EgrCesA8.* We show that the $CT_{(11)}$-microsatellite, 5'UTR and putative CRM identified as key *CesA* regulatory regions in Chapter 3 bind transcription factors of the SCW regulatory network and may regulate SCW *CesA* expression in *Eucalyptus*.

4.3 Materials and Methods

4.3.1 Plant Material

Immature xylem tissue was harvested in the field from a single ramet of a pure species *Eucalyptus grandis* clone from the Mondi Tree Improvement trial at Pietermaritzberg, KZN, South Africa. The tissue was collected by peeling back the bark, which separates from the exposed stem at the cambial layer and then lightly scraping the stem. The immature xylem tissue was flash frozen on site with liquid nitrogen immediately after harvest and stored at -80°C until required.

4.3.2 Bait selection and construction

From the promoter truncation study in Chapter 3 we could identify regions that putatively repress/activate/mediate the specificity of *EgrCesA8* gene expression, eight (A-H) of which were used as bait fragments for Y1H analysis (Figure 4.1). The bait fragments completely covered the promoter from the start codon (+1) to – 1411 bp up stream. The region from -1411 bp to -1850 bp was not used for bait design as there was little reporter gene information for this region in both the Lu et al. (2008) study and our analysis (Chapter 3). A small fragment from -1850 bp to -2018 showed activation of the promoter in Chapter 3 and was also included in the bait design (Figure 4.1). A two-step directional cloning approach was used to clone each fragment into the multiple cloning site of the pHIS2.1 vector (Clontech, CA, USA). The pHIS2.1 vector was specifically designed for yeast-1-hybrid assays and can be used for very short target sequences as short as 50 bp. Forward and reverse primer (Table 4.1) was designed for each of the baits, where the forward primer contained a *Mlu*I cut site and the reverse primer contained a *Spe*I cut site. *Mlu*I and *Spe*I were selected for cloning because neither of these sites was present in the "entry vector" pTZ57R/T (Thermo Scientific,

MA, USA). The bait fragments were amplified from a plasmid containing the previously published full length -2018 bp promoter fragment (Creux et al. 2008). The amplified fragments were gel extracted and purified with the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, Duren, Germany) according to manufacturer's instructions. Purified PCR products were T/A cloned in to the pTZ57R/T vector with the InsTAclone PCR Cloning Kit (Thermo Scientific, MA, USA).

Five positive colonies were selected for each bait-pTZ57R/T construct after blue/white screening and confirmed with colony PCR. A small part of the colony was picked and eluted in 15 µl of SABAX water and heated to 95°C for 5 min. The heated colony suspension was briefly centrifuged and 5 µl was used as template for a standard 20 µl colony PCR reaction with a final concentration of 0.2 mM for each dNTP (Thermo Scientific, MA, USA), 0.4 µM of each primer (Inqaba biotech, PTA, South Africa), 0.8 U of ExSel taq polymerase and 1x ExSel buffer (Jain Biologicals, Haryana, India). A standard PCR protocol was used with an initial denaturation step of 95°C for 2 min followed by 30 cycles of: 95°C for 20 sec, 56°C for 30 sec and an extension step of 72°C for 90 sec, followed by a final extension step of 72°C for 10 min, run on a T100 Thermal Cycler (BioRad, CA, USA). Amplified DNA fragments were visualised on a 1 % agarose gel containing ethidium bromide (Sigma, MO, USA). Amplification was performed using M13 forward and reverse primers (Table 4.1). All colonies with bands of the correct size were inoculated into LB broth and grown at 37°C overnight for plasmid extraction with the GenJET Plasmid Miniprep Kit (Thermo Scientific MO, USA). Positive plasmid extracts were submitted for sequencing (Macrogen, Seoul Korea) with the M13 forward and reverse primers (Table 4.1) and the sequences were analysed using Vector NTI (Invitrogen, CA, USA). Clones with the identical sequence to the reference sequence (*EgrCesA8* sequence from Creux et al. 2008) were selected for further experiments.

Sequence verified pTZ57R/T plasmids containing bait fragments were extracted and double digested with *Mlu*I, *Spe*I and NEBuffer 2 (New England BioLabs, MA, USA) in a 20 µl reaction with a final concentration of 0.2-0.5 µg of plasmid DNA, 1x NEBuffer 2 and 10 U of each restriction enzyme incubated at 37°C overnight. The pHIS2.1 plasmid (Clontech, CA, USA) was also digested in a similar reaction. Digest products were visualised on a 1 % agarose gel and the smaller excised fragments were gel extracted and purified with the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, Duren, Germany). Inserts and linearized pHIS2.1 vector (Clontech, CA, USA) were used in a standard T4 ligase (NEB) reaction and incubated overnight at 22°C. The ligation reaction was used to transform DH5*α* *E.coli* cells. The ligation mix was added to 50 µl of thawed DH5*α* competent cells and incubated on ice for 10 min. Cells were heat shocked for 45-50 sec in a waterbath at 42°C and immediately placed on ice for 2 min after heat shock step. SOC media (900 µl; recipe in *E.coli* competent cells protocols; Promega, WI, USA) was added to the cells and incubated at 37°C with shaking for 60 min. Cells were then plated on standard LB plates with 100 µg/ml Kanamycin and incubated at 37°C overnight. Five colonies per promoter bait were selected for colony PCR screening and sequence confirmation (Macrogen, Seoul, Korea). Colony PCR screening was performed with pHIS2.1 vector specific primers (Table 4.1) as described above and colonies with the correct size insert were inoculated in to LB broth with 1 mg/ml of Kanamycin. Plasmid DNA was extracted and sent for sequencing (Macrogen). Sequences were analysed using vector NTI (Invirtogen). All eight bait fragments (A-H) were successfully cloned into pHIS2.1 and confirmed by sequencing.

4.3.3 Production of Yeast-1-hybrid bait lines

Saccharomyces cerevisiae strain Y187 (Clontech, CA, USA) was transformed with each of the *EgrCesA8* promoter bait constructs (A-H; Figure 4.1) and a pHIS2.1 vector containing the regulatory elements for the human P53 protein as a positive control, using the Frozen-EZ Yeast Transformation II kit (Zymo Research, CA, USA). Transformants were selected for each bait construct on single drop out media lacking tryptophan (SDO, -Trp; Clontech, CA, USA). The bait lines were maintained on -Trp SDO media and 25% glycerol stocks were also produced and stored at -80°C. Bait lines were stored on plates at 4°C and transferred to fresh –Trp SDO media every week and glycerol stocks were replaced every 2-3 months. A Y1H positive control line was produced, which contained the pHIS2.1 vector with a human P53 repeat element and the pGADT7-Rec2 vector with the P53 coding sequence obtained from the Matchmaker One-Hybrid Library Construction and Screening Kit (Clontech, CA, USA) and this was maintained on double drop out media lacking both tryptophan and leucine (DDO; -Trp -Leu).

4.3.4 Prey selection and construction

A number of transcription factors (Table 4.2) from the *Arabidopsis* SCW regulatory network were selected and full-length *Eucalyptus* cDNAs were amplified, cloned and sequenced (Botha et al. –unpublished). Total RNA was isolated from frozen *E. grandis* immature xylem samples using a CTAB method adapted from Chang et al. (1993). Genomic DNA contamination was removed by DNAse treatment in a 40 µl reaction with 20 µl of RNA (>>200 ng/µl), 5 U of DNaseI (Roche, Basel, Switzerland) and 10 U of Rnasin (Promega, WI, USA). The DNaseI reaction buffer was added to a final concentration of 1x and 1 µl of $MgCl_2$ (25mM) solution was added. Reactions were incubated for 30 min at 37°C. To purify

the DNase treated RNA, the volume was adjusted to 100 µl using RNase free water then 350 µl of RLT buffer (Qiagen, Hinden, Germay) and 250 µl of absolute ethanol were added and the Qiagen Rneasy Mini Spin Kit was used as per manual instructions (continuing with step 6). Total RNA quality was measured with the Experion (Biorad) and had an RQI score of 9.5. Total RNA was aliquoted and stored at -80°C until further use.

The DNaseI treated purified total RNA was used as template in a PCR amplification (with intron-spanning *EgrCesA8* primers) to test for any residual DNA contamination. To isolate mRNA from the total RNA, the Dynabeads mRNA Purification Kit (Invitrogen, MO, USA) was used as described in the manual. First strand cDNA was synthesised using the SMART cDNA synthesis technology (Clontech, CA, USA). cDNA was amplified in a 100 µl reaction containing 2 U Phusion Taq (Finnzymes; Espoo, Finland), 8 µl of 2.5 mM dNTPs, 2 µl of cDNA HF Phusion buffer and 12 µM of each primer (PCR primer IIA and Fusion SMATer PCR primer; Clontech, CA, USA). The PCR was performed on a T100 thermocycler (Biorad) with an initial denaturation step at 98°C for 1 min, and then 20 cycles of 98°C for 15 sec, 65°C for 30 sec and 68°C for 6 min. The amplified cDNA was used as template to amplify full-length cDNA for each of the transcription factors selected (Table 4.2; Botha et al. - unpublished). The full-length transcription factor clones where verified by sequencing various clones and comparing this to the *Eucalyptus grandis* genome sequence. Clones that showed sequence variation where translated and the base pair changes were checked for silent mutations. If a base pair change caused a change in amino acid sequence the DNA sequence was checked against the transcriptome data of three *Eucalyptus grandis* individuals to look for possible allelic variation (Botha et al. - unpublished). Only sequences that were a match to the reference sequences and individual transcriptome data were used for prey vector construction in this study.

Full-length cDNAs were amplified directly from the cDNA for 13 of the selected transcription factors (Table 4.2) and were T/A cloned into the entry vector using the pCR8/GW/TOPO TA Cloning kit (Invitrogen, MO, USA). The sequence verified plasmids were used in a homologous recombination reaction (pCR8/GW/TOPO TA Cloning kit; Invitrogen, MO, USA) with the pDEST-GADT7 vector (TAIR Accession-*vector:1010229695*; ABRC) designed by Rossignol et al. (2007) and the sequence and orientation were again verified by sequencing (Botha et al. personal communication). Full-length versions of two transcription factors (*Egr*KNAT7 and *Egr*MYB31) could not be amplified from xylem cDNA and were synthesised (Genscript, NJ, USA) using the *E. grandis* genome references (Eucgr.D01935 and Eucgr.B03684, respectively) which were provided in the pUC57 plasmid. The synthesised gene sequences were verified using the extensive *Eucalyptus* transcriptome and genome sequence data (http://bioinformatics.psb.ugent.be/webtools/bogas, http://www.phytozome.org, and http://EucGenIE.html). The synthesized plasmid was used as a template to amplify the full-length *Egr*KNAT7 and *Egr*MYB31 cDNA sequences and they were cloned into the pDEST-GADT7 vector as described above using the pCR8/GW/TOPO TA Cloning Kit (Invitrogen, MO, USA).

4.3.5 Yeast transformations and dilutions for Y1H screening

The yeast bait lines (A - H) were transferred to fresh -Trp SDO media plates and grown for three days at 30°C. Patches of yeast cells were scraped off the agar and re-suspended in 1 ml of SABAX water, and 100 µl was spotted onto fresh -Trp SDO plates and incubated for three days at 30°C. Two spots of cells were scraped for each mini yeast transformation reaction (Maritz-Olivier 2005) and re-suspended in 1 ml of SABAX water. The cell suspension was

centrifuged at 210 g for 2 min to pellet cells. The supernatant was discarded and the cells were re-suspended in 100 mM LiAc and incubated for 5 min at 30°C with no shaking. Cell suspensions were again centrifuged and the supernatants discarded. Added to the pellet in the specific order listed here were 240 µl of 50 % w/v PEG 4000, 36 µl 1 M LiAc, 25 µl of 2mg/ml Herring testes carrier DNA and finally 500 ng of plasmid made up to 50 µl with SABAX water. The mixture was vortexed for at least 1 min and incubated at 42°C for 25 min. After incubation, the cell suspension was centrifuged at 210 g for 2.5 min to pellet cells and re-suspended in 200 µl of SABAX. Cells were plated out on DDO media using plating beads for even spread. Plates were incubated at 30°C for 5-7 days.

Positive colonies growing on DDO media containing both the pHIS2.1 and the pDEST-GAD vector and can synthesise tryptophan and leucine. Three to five positive colonies were picked for each of the different bait and prey combinations including the positive and negative controls and re-streaked on fresh DDO plates. The replica plates were grown for a further 3-5 days at 30°C and then the yeast was scraped off the plates and eluted in 1 ml SABAX water in 96 deep well storage blocks. From each colony, 100 µl of the cell suspension was transferred to a 96-well UV flat bottom microtiter plate (ThermoScientific, MA, USA) and OD readings were measured with a MultiSkanGO (ThermoScientific, MA, USA) plate reader at a wave length of 600 nm. All cell suspensions were diluted to the lowest OD on the plate by transferring the cell suspension to a new storage block and adding the correct amount of sterile water to each well. Diluting the cell suspension ensured that the same number of cells was plated out during screening (See Supplementary file 4.1 for actual cell counts for each plate). There would be plate to plate variation with this method, but experimental design ensured that all of the different *EgrCesA8* promoter baits were in a single storage block for each prey and we could compare interactions of single preys on different regions of the *EgrCesA8* promoter.

4.3.6 Y1H screening of single prey and bait

The eight bait lines were screened against 14 different *Eucalyptus* transcription factors (Table 4.2) and an empty pDEST-GAD vector as a negative control. In order to screen yeast colonies for Y1H interactions, 10 µl of the cell suspension (Supplementary file 4.1) was placed onto 145 mm petri dishes containing different drop out media. The media used for the initial screen was -Trp or -Leu SDO, -Trp-Leu DDO media and triple drop out media (TDO) lacking tryptophan, leucine and histidine (-Trp-Leu-His). The SDO media indicate that the prey and bait vector are present and the DDO ensures that both are present in the same cell. The TDO media screens for interaction between the transcription factor in the prey vector and the *EgrCesA8* promoter fragment in the bait vector which will activate the expression of the HIS gene on the pHIS2.1 plasmid (Clontech). In some cases, the baits (A, B, G and H) showed auto-activation and initiated HIS expression in the absence of a prey fragment. To observe interactions for these baits the media was supplemented with varying concentrations of 3-amino-1,2,4-triazole (3AT; Sigma). The same cell suspensions were used and 10 µl was spotted onto TDO plates with the different 3AT concentrations and incubated for 5 days at 30°C. Supplementary files 4.1 and 4.2 show the raw plate images from which figures were constructed.

4.3.7 Y1H screening of dual prey and single bait

Sixteen yeast lines containing a prey vector with *Egr*MYB80 or *Egr*MYB87 (homologs of *At*MYB52/*At*MYB54) and each of the bait vectors (A-H) were grown on DDO media, transformed with a second prey vector *Egr*KNAT7, and plated on TDO media. Colonies

growing on the TDO media were picked and streaked out on fresh TDO media. The colonies were diluted and screened (section 4.3.6) and the strength of the interaction was measured on TDO media with varying concentrations (1 mM, 2 mM, 3 mM, 5mM and 10 mM) of 3AT. PCR amplification with *Egr*KNAT7, *Egr*MYB80 or *Egr*MYB87 and pHIS2.1 specific primers (Table 4.2) were used to verify the presence of all three vectors in the colonies screened (Supplementary file 4.3). Part of each spot was picked and eluted in 20 µl of SABAX water. The colonies were heated to 98°C for 10 min, then immediately flash frozen with liquid nitrogen and returned to 98°C for a further 10 min to rupture the cells and release the plasmid. After brief centrifugation, 5 µl of supernatant was used as template for a standard colony PCR.

4.4 Results

4.4.1 Y1H screening of SCW transcription factors against promoter fragments of the EgrCesA8 gene

In Chapter 3 we identified several regions within the *EgrCesA8* promoter that were putatively involved in the regulation of gene expression (Figure 3.8). These promoter regions were the focus of the bait design for this Y1H assay. Six different bait constructs (C-H) were designed to overlap with previously identified regulatory regions of *EgrCesA8* promoter (Chapter 3). Bait sizes ranged from approximately 200 bp to 400 bp (Figure 4.1). Two more baits (A-B) were designed to correlate to the UTR (bait A) and TC-UTR (bait B) fragments used for the GUS (*β*-glucuronidase) assays in Chapter 3. *Arabidopsis* literature was used to identify a panel of 15 transcription factors for the Y1H assays (Table 4.2). The *Eucalyptus* SCW-associated transcription factors were identified by phylogenetic analysis which included *Populus* and *Arabidopsis* homologs (Botha et al.; Hussey et al.; and Solar et al.; unpublished).

The full-length *Eucalyptus* transcription factors were amplified from immature xylem cDNA, cloned and sequenced (Table 4.2; Botha et al. unpublished). Three of the transcription factors (*Egr*MYB31, *Egr*KNAT7 and *Egr*NAC75) were selected because their *Arabidopsis* homologs (Table 4.2) have previously been linked to *AtCesA8* gene regulation (Zhong et al. 2007; Bhargava et al. 2010; Zhong and Ye 2012; Kim et al. 2013). *At*MYB85 was implicated in lignin biosynthesis (Zhong et al. 2008) and was therefore not expected to activate the *EgrCesA8* promoter. Screening with *Egr*MYB61 (*At*MYB85 homolog) served as a second negative control. The remaining transcription factors were selected because little is known about their function in the secondary cell wall regulatory network with respect to cellulose deposition (Table 4.2).

A total of 120 pair-wise bait and prey combinations were transformed into *Saccharomyces cerevisea* (strain Y187). Three transcription factors (Table 4.2; *Egr*NAC75, *Egr*MYB2 and *Egr*NAC61) did not produce any yeast colonies even after multiple rounds of transformation suggesting that these proteins may be lethal to the yeast cells. Therefore, there were a total of 96 pair-wise combinations which could be screened for Y1H interactions (Table 4.4). To ensure accurate screening, the number of cells plated per yeast spot was calculated (Supplementary file 4.1). Plates with a high standard deviation between spots (e.g. *Egr*KNAT7, *EgrMYB87* and *Egr*NAC64) where manually inspected to ensure that the differences in cell number did not affect the yeast growth (Table 4.4 and Supplementary file 4.1). An example of this is the *Egr*KNAT7 screen where spot F2 had the highest cell count and A3 had the lowest cell count on the plate, however both grew equally well on the DDO media and the F2 spot did not grow more than the A3 spot on TDO media (Supplementary file 4.1). This suggests that the cell numbers plated were within an acceptable range and did not cause significant variation of the results.

An initial Y1H screen was performed where all yeast colonies that contained both a bait and a prey vector were grown on plates containing -Trp or -Leu SDO media, DDO media and TDO media to identify interactions (Table 4.4). The negative controls in the Y1H screen contained a bait vector and an empty prey vector which contained no transcription factor and only the binding domain was present. Positive interactions were identified when yeast containing a specific bait and prey combination could grow on the TDO media and the respective negative control could not. The initial screen revealed that *Egr*NAC170 (*At*SND2 homolog; Table 4.2) interacted with the *EgrCesA8* promoter bait F. The yeast cells harbouring this bait-prey combination could grow on TDO media, while the negative control could not (Figure 4.2 A). The strength of the interaction observed between *Egr*NAC170 and bait F was tested on the series of TDO plates containing increasing concentrations of 3-amino1,2,4-triazole (3AT). The *Egr*NAC170-bait F interaction was not a strong interaction and most of the growth dropped off immediately when 1 mM of 3AT was added (Figure 4.2B).

During this initial screen we found that baits A, B, G and H showed auto-activation and the negative empty vector controls for these baits all showed growth on the standard TDO media (Figure 4.2A and Supplementary file 4.1). In order to observe interactions with baits A, B, G and H they were also plated on a series of TDO plates with increasing concentration of 3AT. A low 3AT concentration (1 mM) was sufficient to abolish any leaky HIS expression for baits A, B and G (Figure 4.2C; negative control and Supplementary file 4.2). Using the TDO 3AT plate series, two more interactions were observed for bait A (Figure 4.2C). An interaction was observed between the *Egr*ZincFinger-A transcription factor (homolog of *At*C3H14; Table 4.2) with bait A, where the negative control could not grow on TDO with 1 mM of 3AT, but the positive colonies continued to grow on media containing up to 2 mM 3AT (Figure 4.2C: *Egr*ZincFinger-A). *Egr*MYB31 showed a stronger interaction

with bait A and the positive colonies grew on media containing up to 5 mM 3AT (Figure 4.2C: *Egr*MYB31). Bait H with the empty pDEST-GAD vector showed strong Y1H interaction growing on TDO media containing up to 10 mM 3AT (Supplementary file 4.2). This suggests that there is activation of the reporter gene by endogenous yeast factors. Using bait H in a *cis*-element search in TRANSFAC (Wingender et al. 2000) revealed four yeast specific *cis*-elements that were bound by yeast transcription factors (YLR278C, YBR239C, RSC3 and RSC30). These *cis*-elements were not identified in any of the other downstream regions of the promoter and may explain why only bait H showed strong auto-activation of HIS expression.

4.4.2 A dual Y1H screen of EgrKNAT7 and EgrMYB80/87 against the regulatory modules of the EgrCesA8 promoter

Transcription factors seldom interact with DNA in isolation (reviewed in Lelli et al. 2012), and for this reason we considered a dual Y1H approach where two prey vectors containing different transcription factors are co-transformed into the yeast containing single promoter baits. In the SCW regulatory network, *At*KNAT7 has previously been shown to form hetero-dimers with other members of the network such as *AtOFP4* and *At*MYB75 (Bhargava et al. 2010; Li et al. 2011; Li et al. 2012; Bhargava et al. 2013). *At*MYB75 contains the R3 domain that interacts with the KNOX2 domain in *At*KNAT7 (Bhargava et al. 2013) and many of the other MYBs in the SCW regulatory network also contain the R3 domain. In this study we cloned the full-length *Eucalyptus grandis* homologs of *At*KNAT7, *At*MYB52 and *At*MYB54. Individually none of these transcription factors showed interaction with any of the *EgrCesA8* baits (Figure 4.3; Supplementary files 4.1 and 4.2). Given the previous protein-protein interaction data for *At*KNAT7, it is possible that DNA-protein interactions may only occur

when hetero-dimers are present, therefore we decided to test if co-transformation of *Egr*KNAT7 and *Egr*MYB80 (*At*MYB52 homolog) or *Egr*MYB87 (AtMYB54 homolog) would show interaction with any of the promoter baits in this study.

The *Egr*KNAT7::pDEST-GAD vector was transformed into the yeast strains already containing either *Egr*MYB80 or *Egr*MYB87 with each of the *EgrCesA8* promoter baits (A-H). The transformation mixture was plated directly on TDO media and positive colonies represent putative dual interactions. The yeast colonies should contain all three vectors, two of which had the same selectable marker (*leu*). Therefore we performed a colony PCR with *Egr*MYB80/87 and *Egr*KNAT7 specific primers to ensure both vectors were present in the cells (Supplementary file 4.3). Only the lines containing *Egr*MYB80-bait E or *Egr*MYB87-bait E produced colonies suggesting an interaction of bait E with either the *Egr*KNAT7-*Egr*MYB80 or *Egr*KNAT7-*Egr*MYB87 combinations. A full screen on the different drop out media revealed that yeast harbouring, *Egr*KNAT7-*EgrMYB80*-bait E or *Egr*KNAT7-*Egr*MYB87-bait E could produce HIS and grow on the TDO media (Table 4.4 and Figure 4.3A). To test the strength of the interaction the colonies were plated on TDO media with an increasing concentration of 3AT. Both of the interactions (*Egr*KNAT7-*EgrMYB80*-bait E and *Egr*KNAT7-*EgrMYB87*-bait E) were strong, compared to the other interactions observed in this study, and the cells grew on TDO media with up to 15 mM 3AT (Figure 4.3B and Supplementary file 4.3). The results suggest that these proteins form a regulatory complex which could bind co-operatively to the *EgrCesA8* promoter.

4.4.3 Combinatorial promoter analysis of EgrCesA8: including cis-element, sequence, expression and Y1H analyses

Comparing the different lines of evidence collected over the entire study (Chapters 1-3) allowed for the construction of an integrated model of *EgrCesA8* promoter regulation (Figure 4.4). The *cis*-element mapping and GUS reporter gene analysis revealed a putative CRM from position -1036 to -699 (summary Figure 4.4 and Chapter 3). This promoter region showed variable GUS expression depending on the region deleted and a number of SCW-associated *cis*-elements were mapped to the same region (Figure 4.4). The Y1H results support the hypothesis that this region is a key CRM in the regulation of *EgrCesA8* because at least four transcription factors (*Egr*NAC170, *Egr*MYB80, *Egr*MYB87 and *Egr*KNAT7) were shown to interact with this region in the Y1H screens (Table 4.4). The design of the baits and *cis*-element maps also allowed us to approximate the position of the protein binding sites in the CRM. None of the proteins investigated bound to bait D suggesting that the proteins screened are only involved with the upstream part (-1036 to -853) of the CRM (Figure 4.4). *Egr*MYB80 or *Egr*MYB87 in conjunction with *Egr*KNAT7 showed interaction with bait E but not bait F or D suggesting that the hetero-dimer binding site may straddle the junction between bait F and D at position -853 (Figure 4.4). An analysis of all the SCW MYB binding sites identified to date (Goicoechea et al. 2005; Winzell et al. 2010; Zhong and Ye 2012; Kim et al. 2013) revealed that they shared a conserved core region (TWGGT:ACCWA) which we termed the secondary cell wall MYB (SCWM) binding site (Figure 4.5A). This site was identified in bait E (-1036 to -966) and was found to be conserved across the *CesA8* promoter of 13 different *Eucalyptus* species (Figure 4.5B and C), which lends support for the interaction of *Egr*MYB80 and *Egr*MYB87 in this region.

*Egr*NAC170 (*At*SND2 homolog; Table 4.2) interacts with bait F which also overlaps with the CRM identified in the *EgrCesA8* promoter (Figure 4.4). Sequence analysis revealed a secondary wall NAC binding element (SNBE; Zhong et al. 2010b) and two SND1-like elements (Wang et al. 2011) in the bait F sequence (Figure 4.5A). All three occurrences conform to the imperfect palindrome pattern for NAC binding elements (Welner et al. 2012). Previouse studies have shown that *At*SND1 binds directly to the SNBE element in *Arabidopsis* (Zhong et al 2010b). *Egr*NAC61 (*At*SND1 homolog) may interact with this region of the *EgrCesA8* promoter, but the interaction could not be identified because *Egr*NAC61 was lethal to the yeast (Table 4.4). It is also possible that *Egr*NAC*170* could bind to these SNBE and SNBE-like sites between positions -900 and -853 (Figure 4.4 and 4.5B), which might be due to partial sequence similarity between the SNBE and *Egr*NAC170 binding sites. There may also be an *Egr*NAC170-specific site located somewhere within the bait F region which has yet to be identified. Sequence analysis of these occurrences across 13 different *Eucalyptus* species indicated that the SNBE occurrence was the most highly conserved while the two SND1-like sites showed more variation across the species (Figure 4.5C).

Bait A and B represent two constructs that overlap with the 5'UTR of *EgrCesA8* and differ by the presence of a $CT_{(11)}$-microsatellite at the 5' end of bait B. The Y1H analysis revealed that the *Egr*MYB31 and *Egr*ZincFinger-A (*At*MYB46 and *At*C3H14 homologs; Table 4.2) both interacted with bait A but not bait B suggesting the $CT_{(11)}$-microsatellite hinders protein interaction in some way and from the GUS assays we also noted that the microsatellite influences reporter gene expression (Figure 4.4). Sequence analysis revealed a SCWM binding site in the 5'UTR which was perfectly conserved across the *EgrCesA8* 5'UTRs of the 13 different *Eucalyptus* species (Figure 4.5). Two SCWM-like binding sites were identified that did not match the SCWM consensus sequence (ACCWA) but did share

the same sequence variant (ACCWT). This sequence change was conserved in all 13 *Eucalyptus* species suggesting there are multiple MYB binding sites in the *EgrCesA8* 5'UTR where *Egr*MYB31 may interact (Figure 4.5C).

4.5 Discussion

4.5.1 The $CT_{(11)}$-microsatellite plays a key role in the regulation of EgrCesA8

A number of different studies have shown that the 5'UTR can be integral to the regulation of plant genes (Bolle et al. 1994; David-Assael et al. 2005; Castillejo and Pelaz 2008; Liao et al. 2013). A prominent feature of the *EgrCesA8* 5'UTR is the $CT_{(11)}$-microsatellite positioned at the transcriptional start site (Chapter 2 and 3) The $CT_{(11)}$-microsatellite accompanied by the *EgrCesA8* 5'UTR produced strong, deregulated, constitutive GUS expression in all *Arabidopsis* and *Eucalyptus* tissues assayed (Figure 4.4 and Chapter 3). The $CT_{(11)}$-microsatellite also appear to affect the Y1H protein-DNA interactions where bait A (5'UTR) did not interact with any of the screened TFs, but bait B ($CT_{(11)}$-5'UTR) interacted with both *Egr*MYB31 and *Egr*ZincFinger-A. These results indicate that the $CT_{(11)}$-microsatellite could play an important role in the assembly of the basal transcriptional machinery. Without the upstream promoter regions and transcription factors to direct the recruitment of the basal transcriptional machinery it may override or hinder the subsequent recruitment of downstream factors such as *Egr*MYB31 and *Egr*ZincFinger-A (Figure 4.6). An alternative possibility is that removal of the upstream promoter sequence could influence the DNA structure and the transcription factors such as *Egr*MYB31 and *Egr*ZingFinger-A may not be recruited and this could alter GUS expression and lead to the interactions observed for the Y1H assays. GATA-like proteins are known to bind to CT-repeat elements and thus change the conformation of the promoter DNA (reveiwed in Berger and Dubreucq 2012). Changing

the surrounding DNA conformation of these elements may affect the recruitment of the transcription factors and the expression of the gene. Additional GUS analyses, Y1H assays and EMSAs will be required to fully elucidate the role of the $CT_{(11)}$-microsatellite in *EgrCesA8* gene regulation and initiation of transcription.

4.5.2 SCW-associated transcription factors interact with the EgrCesA8 5'UTR

*At*MYB46 (homolog of *Egr*MYB31) has been referred to as a master regulator of SCW formation as it can activate a number of transcription factors within the network including *At*MYB52, *At*MYB54 and *At*KNAT7 (Zhong et al. 2007). While most master regulators are situated in the top tiers of the hierarchical regulatory network this does not exclude them from directly activating structural genes such as the *CesAs* (Winzell et al. 2010; Yamaguchi et al. 2011). Recent work in *Arabidopsis* has revealed a strong interaction between the SCW-related *CesA* (*AtCesA4*, *AtCesA7* and *AtCesA8*) promoters and *At*MYB46 (Kim et al. 2013). Our results suggest direct interaction of *Egr*MYB31 and the *EgrCesA8* 5'UTR, which is supported by previous study by Kim et al. (2013) who found that *At*MYB46 interacted with the *AtCesA4* promoter which contained a large part of the 5'UTR. These findings suggest that there is some conservation of the SCW regulatory network in *Eucalyptus* and *Arabidopsis* initially proposed by Zhong et al. (2010a).

From the literature, a core secondary cell wall MYB (SCWM) binding element was identified and *cis*-element analysis revealed one fully conserved site and two SCWM-like sites in the *EgrCesA8* 5'UTR (Figure 4.5). These elements were well conserved across the *CesA8* 5'UTRs of different *Eucalyptus* species. Extending the sequence revealed good overlap with the SMRE/M46RE/ACTYP element (Figure 4.5A and C) identified previously to bind MYB46 orthologs in *Arabidopsis* and *Populus* (Winzell et al. 2010; Zhong and Ye

2012; Kim et al. 2013). The multiple copies of the SCWM and SCWM-like binding sites are also consistent with findings by Kim et al. (2013) who found at least two to three M46RE sites (RWTWGGTR:YACCWAWY) in all of the SCW-associated *CesA* promoters. Multiple copies of the SCWM and SCWM-like sites may also explain the relative strength of the interaction we observed. Further mutation and EMSA studies are required to confirm the SCWM and SCWM-like sites and their regulatory role in the *EgrCesA8* 5'UTR.

The two SCWM-like sites in the 5'UTR showed a single miss-match to the consensus sequence and in both cases the last adenine was substituted with a thymine (ACCWA→ACCWT). The mismatch in both occurrences was also conserved across all 13 *Eucalyptus* species and could be *Eucalyptus*-specific sites which still resemble the SCWM binding elements and may bind the same or different MYB proteins (Figure 4.5). The GUS analysis of Chapter 3 also noted that in *Eucalyptus* stems the GUS expression modulated by the 5'UTR fragment was confined to the immature xylem while in *Arabidopsis* no GUS expression was observed for the *Eucalyptus* 5'UTR fragment (Figure 4.4). The difference in GUS expression between *Arabidopsis* and *Eucalyptus* may be linked to the presence of the SCWM-like binding sites in the 5'UTR which might not be recognised by the *Arabidopsis* transcriptional machinery, although this will have to be further investigated before regulatory differences between *Eucalyptus* and *Arabidopsis* can be fully described.

A C3H14 zinc finger transcription factor (*Egr*ZincFinger-A) was also identified to interact with the 5'UTR of the *EgrCesA8* promoter (Figure 4.2B). Currently there is little information on the role of the *Arabidopsis* C3H14 (At1g66810) transcription factor in the SCW regulatory network. Previous studies have found that *At*C3H14 is a direct target of *At*MYB46 (Zhong et al. 2007; Kim et al. 2012; Kim et al. 2013). Ko et al. (2009) also found that *At*C3H14 could activate the expression of the cellulose synthase genes, which suggests the presence of a feed-forward loop activating *AtCesA8* gene expression. *At*MYB46 already

forms part of a separate feed forward loop at a higher level of the network where SCW master regulators activate *At*MYB46 expression and directly activate the cell wall biosynthetic genes which are activated by *At*MYB46 (Chapter 1; Figure 1.1). This is the first study to show direct DNA-protein interaction for a C3H14 protein in *Arabidopsis* or *Eucalyptus* and therefore no sequence motifs are currently implicated in this interaction and are still to be identified. The binding of multiple transcription factors to the *EgrCesA8* 5'UTR suggests the presence of a CRM in this region and further Y1H experiments, electrophoretic mobility shift assays (EMSAs) and protein pull-down assays could reveal the nature of this putative 5'UTR CRM and the *cis*-elements it contains.

4.5.3 Key transcription factors in the SCW regulatory network interact with the putative EgrCesA8 CRM

A CRM (*cis*-regulatory module) is a ≈ 300 bp site in a promoter where multiple transcription factors can interact with clustered *cis*-elements to modulate gene expression (Levine and Davidson 2005). In Chapter 3 a putative CRM was identified in the *EgrCesA8* promoter between -1036 and -699 bp (Figure 4.4). This CRM was used for Y1H bait design where bait F covered the 5' half of the CRM, bait D covered the 3' half of the CRM and bait E contained the entire CRM (Figure 4.1). During the Y1H screening we identified a number of the cloned *Eucalyptus* transcription factors (*Egr*NAC170, *Egr*KNAT7, *Egr*MYB80 and *Egr*MYB87) which interacted with the different CRM baits (Figure 4.2 and 4.3). *Egr*NAC170 interacted with bait F (Figure 4.2) and none of the other CRM baits, which is surprising as bait E overlaps bait F and the same interaction would be expected for both baits (Figure 4.1). The longer bait E has a larger distance between the *Egr*NAC170 binding region and the minimal promoter of the vector, which could weaken the interaction to a level that may not be

detected in the current analysis. The CRM region bound by *Egr*NAC170 was designated as an activator region from the GUS analysis (Figure 4.4) of the *EgrCesA8* promoter (Chapter 3). The direct interaction of *Egr*NAC170 and the activation region of the CRM support earlier suggestions that the *Arabidopsis* homolog (*At*SND2; Table 4.2) activates the expression of the SCW-associated *CesA* genes (Zhong et al. 2008; Hussey et al. 2011).

Two NAC-like *cis*-elements have previously been identified to bind NAC proteins regulating SCW deposition and these include the SND1 binding site (Wang et al. 2011) and the secondary wall NAC binding element (SNBE; Zhong et al. 2010b). Three of these NAC-like binding sites where identified in the *EgrCesA8* CRM region (bait E) binding *Egr*NAC170 (Figure 4.5A and B). Two of the NAC-like binding sites were similar to the *At*SND1 binding site (Wang et al. 2011), but they showed some sequence variation when compared to the consensus sequence (Figure 4.5C). These SND1-like elements still resemble an imperfect palindrome recognised by many different NAC transcription factors (Welner et al. 2012) and may bind *Egr*NAC170 or other key NAC proteins involved in *CesA* regulation at the CRM. A third NAC-like binding site, in the CRM region (bait E), was well matched to the SNBE consensus sequences and was highly conserved across the *EgrCesA8* promoters of 13 different *Eucalyptus* species, suggesting some functional constraint (Figure 4.5). The SNBE site interacts with different NAC domain transcription factors from the *Arabidopsis* SCW regulatory network including *At*SND1, *At*NST1, *At*VND6 and *At*VND7 (Zhong et al. 2010b; McCarthy et al. 2011). Most studies have focused on the SNBE element as an element for the master regulators, however, none of the other NAC transcription factors (e.g. SND2 and SND3) in the network have been tested against this element. The ambiguity of the SNBE consensus sequence (WNNYBTNNNNNNNAMGNHW) may be an indication that the SNBE sequence is a combination of different NAC binding elements and with further

investigation could be separated into more specific consensus elements for each of the NAC proteins in the SCW regulatory network.

In the initial Y1H screen, none of the other transcription factors in the panel interacted with the *EgrCesA8* CRM region. However, it is well known that during Y1H analysis, protein-DNA interactions may be missed because transcription factors often function as heterodimers (Lelli et al. 2012; Smaczniak et al. 2012; Schaart et al. 2013). In a dual Y1H screen we tested *Egr*KNAT7 with either *Egr*MYB80 or *Egr*MYB87 and observed strong dual Y1H interactions with the bait E region (Figure 4.3 and Supplementary file 4.3). The *Egr*KNAT7-*Egr*MYB80 and *Egr*KNAT7-*Egr*MYB87 putative complexes only interacted with bait E (not F or D) suggesting that the interaction occurred at the junction between bait F and bait D. *Egr*KNAT7 is a known repressor and at the junction between bait F and bait D there is a repression region identified by previous promoter truncation and deletion studies which supported the position of this interaction (Figure 4.4). We also identified two SCWM binding sites in bait E which were well conserved across 13 different *Eucalyptus* species (Figure 4.5) supporting the suggestion that a MYB protein interacts in this region of the *ErgCesA8* CRM. Currently there is no evidence of direct KNAT7-DNA interactions in the literature and therefore there is no information on the possible DNA sequences that mediate this interaction.

While *At*KNAT7 has been shown to interact with different transcription factors (*At*MYB75 and *At*OFP4) in order to suppress various genes within the SCW regulatory network (Bhargava et al. 2010; Li et al. 2011) there have been no direct links made between KNAT7 and MYB52 or MYB54 orthologs in any plant species to date. However, several indirect links can be identified in previous tansactivation and expression studies. *At*KNAT7 (*Egr*KNAT7 homolog), *At*MYB52 (*Egr*MYB80 homolog) and *At*MYB54 (*Egr*MYB87 homolog) are all transactivated by *At*MYB46 suggesting a form of co-regulation and co-

expression (Zhong et al. 2008). Ko et al. (2009) also noted that individually *At*KNAT7, *At*MYB52 and *At*MYB54 could not transactivate promoter::GUS constructs of key structural genes in the network suggesting they may work to repress parts of the network or require co-factors to operate. Co-regulation and co-expression of genes are often good indicators of interacting protein pairs (Bhardwaj and Lu 2005) and therefore these data support the dual Y1H interactions observed in this study. Elucidating all the interacting partners of *At*KNAT7 and the *cis*-elements involved will provide a more complete understanding of the SCW regulatory network.

There is very little information on the role of *At*MYB52 (*Egr*MYB80 homolog) and *At*MYB54 (*Egr*MYB87 homolog) in the SCW regulatory network. A study, focused on drought tolerance, found that the overexpression of *At*MYB52 induced drought tolerance and decreased the expression of key xyloglucan genes (Park et al. 2011). Cal et al. (2013) tested drought sensitive and drought tolerant rice cultivars and found that in the tolerant cultivar a host of SCW genes, including *OsCesA4* (*EgrCesA8* ortholog) and *OsCesA9* (*EgrCesA7* ortholog) where strongly down regulated. A number of studies showed that *At*KNAT7 is a key repressor protein in the SCW network which functions as a complex with proteins such as *At*MYB75 and *At*OFP4 to repress specific cell wall related genes (Bhargava et al. 2010; Li et al. 2011; Li et al. 2012; Bhargava et al. 2013). A recent study also identified *At*MYB52 as a repressor of the lignin biosynthetic pathway during SCW formation (Cassan-Wang et al. 2013). The current results suggest that *Egr*KNAT7 and *Egr*MYB80 and *Egr*MYB87 form a regulatory complex and these could be involved in the repression of SCW formation during normal plant development or in response to the environment. At least one of these interactions could play a role in repressing *EgrCesA8* expression during stress responses such as drought, although this would have to be confirmed in further studies. Our GUS assays of promoter truncation -1036, which contains the CRM region binding the *Egr*KNAT7-

*Egr*MYB80/*Egr*MYB87 complex and -853, produced an expression pattern similar to that observed for the full-length *EgrCesA8* promoter in both *Arabidopsis* and *Eucalyptus* (Figure 4.4). The GUS assays of promoter truncation -588 (that does not contain any part of the CRM) indicated expression in the *Arabidopsis* hypocotyls and *Eucalyptus* stem tissue, suggesting the *Egr*KNAT7-*Egr*MYB80 and *Egr*KNAT7-*Egr*MYB87 interaction complex does not influence normal stem specific-expression of the *EgrCesA8* gene during cell wall formation (figure 4.4). It is possible that no variation in expression was observed because the *Egr*KNAT7-*Egr*MYB80 and *Egr*KNAT7-*Egr*MYB87 complexes mediate response to biotic or abiotic stress (Bhandari et al. 2006, Ramirez et al. 2011, Park et al. 2011) and therefore would not show variation of GUS expression under normal growth conditions.

4.5.4 Conservation and additions to the SCW regulatory network revealed by Y1H assays

In terms of the current regulatory network (reviewed in Demura and Ye 2010; Zhong et al. 2010a; Handakumbura and Hazen 2012) our results confirmed that *Egr*MYB31 (homolog of *At*MYB46; Table 4.2) interacts with the *EgrCesA8* promoter, which indicates that key transcription factors in the network and their functions are conserved across distantly related plants with varying degrees of SCW formation. To date there have been no direct interactions shown for *At*SND2 and *At*C3H14, although both have been speculated to directly activate the SCW-associated *CesA* genes. Here we can add to the SCW network as we show direct Y1H interactions for *Egr*NAC170 (homolog of *At*SND2; Table 4.2) and *Egr*ZincFinger-A (homolog of *At*C3H14) with the *EgrCesA8* promoter (Figure 4.6). The interaction of both *Egr*MYB31 and *Egr*ZincFinger-A allowed us to identify an additional promising feed-forward loop within the SCW regulatory network where *Egr*MYB31 (*At*MYB46 homolog) can directly activate *CesA* and can activate the intermediate *Egr*ZincFinger-A, which in turn

activates the *CesA* promoters. In the SCW regulatory network there is little information on the protein-protein interactions which are crucial for transcription factor function. With our current study we have added two putative protein-protein interactions between *Egr*MYB80 or *Egr*MYB87 and *Egr*KNAT7 (Figure 4.6). There is little or no information of the role of *At*MYB52 or *At*MYB54 (homologs of *Egr*MYB80 and *Egr*MYB87; Table 4.2) during normal cell wall development and has only been linked to SCW formation because they are activated by the SCW master regulators (Zhong et al. 2010b). As discussed before *At*MYB52 is an important transcription factor in drought response during which cell wall biosynthetic genes are down-regulated (Park et al. 2011; Cal et al. 2013). This suggests that in addition to *AtKNAT7* playing a repressive role in the network to maintain homeostasis (Li et al. 2011; Li et al. 2012). It may also be a key transcription factor modulating the cell wall's response to various biotic and abiotic stresses. This is also the first report of *At*MYB54 or *At*MYB52 homologs (Table 4.2) binding to any SCW structural gene promoter and implicates them in regulatory protein complexes which may modulate *CesA* gene expression. Our results offer a good starting point for further investigation of these two poorly characterized transcription factors.

4.6 Conclusion

Our findings were similar to those of Zhong et al. (2010a) revealing functional conservation of the SCW regulatory network in *Arabidopsis* and *Eucalyptus*, where conserved roles for *Egr*NAC170 (*At*SND2) and *Egr*MYB31 (*At*MYB46) in cellulose synthase regulation where identified (Figure 4.6). This study also added to the current SCW regulatory network by revealing direct interactions with the *EgrCesA8* (ortholog of *AtCesA8*) promoter and several key SCW-associated transcription factors including *EgrNAC*170 and *Egr*ZincFinger-A. A

number of highly conserved MYB- and NAC-like binding sites were identified in the *CesA8* promoter of 13 different *Eucalyptus* species and are candidates for further studies to elucidate their role in the regulation of the promoter by the MYB and NAC interactions identified in this study. This is also the first study to show a link between *Egr*KNAT7 and *Egr*MYB80/*Egr*MYB87 by dual Y1H interaction and this putative regulatory complex may explain the down regulation of the SCW genes during drought response (Cal et al. 2013). This will have to be further explored before the exact mechanisms behind the regulation are uncovered. The identification of the putative *Egr*KNAT7-*Egr*MYB80/*Egr*MYB87 regulatory complex and previous findings that *At*MYB46 is involved in pathogen response suggest overlapping regulatory networks with many of the genes being involved in multiple processes. The identification of overlapping networks may be important for biotechnological applications where manipulating a gene in one network may have far reaching unintended effects on other important developmental or resistance processes. The current study begins to construct the *Eucalyptus* SCW regulatory network and these findings can be used to further investigate the different regulatory mechanisms affecting cellulose synthase regulation.

4.7 Acknowledgements

The authors would like to acknowledge Mr Jonathan Botha for identifying the *Eucalyptus* homologs of the SCW transcription factors and preparing several of the prey vectors used in this study. Acknowledgement also goes to Mr Steven Hussey and Dr Marcal Solar for providing the *Eucalyptus* gene names and identifiers used in this study from their respective manuscripts which are currently in preparation. This work was supported with funding provided by Mondi and Sappi, through the Forest Molecular Genetics (FMG) Programme, the

Technology and Human Resources for Industry Programme (THRIP) and the National Research Foundation of South Africa (NRF).

Table 4.1 Primer list for *EgrCesA8* Y1H bait construction and sequencing confirmation.

Primer Name	Primer sequence[a]	Use
*EgrCesA8*_5'UTR_F1-MluI	GCAT**ACGCGT**TGCACTGCTTCAACACAATGACAC	PCR Bait A[b]
*EgrCesA8*_2180B_R-SpeI	CTCT**ACTAGT**CTCGCAATCTTCTCGCGACCCAATGA	PCR Bait A and B[b]
*EgrCesA8*_5'UTR_F0-MluI	CTTG**ACGCGT**GTAGCATGTCATCTCCCTCCTCTG	PCR Bait B[b]
749-forward_MluI	AGAG**ACGCGT**AGAGTGGGAGAGCATTTCGATAA	PCR Bait C[b]
582-reverse_SpeI	TGCAGC**ACTAGT**GTAGCGGGTCTGACTAACCGATTGA	PCR Bait C[b]
A594_*EgrCesA8*_847_MluI	AGAG**ACGCGT**AGAGTGGTAGTTGGTGGGTTAACCT	PCR Bait D[b]
749-reverse_SpeI	TGCAGC**ACTAGT**GTAGCGGGTTCCCTAGATGGCCCTATTTC	PCR Bait D and E[b]
A597_*EgrCesA8*_1030_MluI	AGAG**ACGCGT**AGAGTAGAGGGAGGGTAAGTTCACT	PCR Bait E and F[b]
A199_*CesA8*_SW_222_SpeI	TGCAGC**ACTAGT**GTAGCGGGTTAACCCACCAACTACC	PCR Bait F[b]
1_1405_MluI	AGAG**ACGCGT**AGAGTTGAGGTTGGATCGTGCTTCTG	PCR Bait G[b]
1188-reverse_SpeI	TGCAGC**ACTAGT**GTAGCGGACTCTGCCTAGATAATGTTGTC	PCR Bait G[b]
A238_Egn*CesA8*_139F_MluI	AGAG**ACGCGT**AGAGTCCTTGCACATCCAATTGC	PCR Bait H[b]
1787-reverse_SpeI	TGCAGC**ACTAGT**GTAGCGGTCAGACTTGGGGACTGAT	PCR Bait H[b]
M13-Forward	GTTTTCCCAGTCACGAC	PCR/sequencing
M13-Reverse	CAGGAAACAGCTATGAC	PCR/sequencing
pHIS ForwardA	CTGCAAGGCGATTAAGTTGG	Verification PCR
pHIS ReverseB	GGTGTGATGGTCGTCTATGT	Verification PCR
pGAD ForwardA	AGTAGCAACGGTCCGAACCT	Verification PCR
pGAD ReverseB	GATGGTGCACGATGCACAGT	Verification PCR
*Egr*KNAT7_F	GACGCGCAGTAGACGAAGTA	Yeast validation
*Egr*KNAT7_R	GTTGCGAGACCAGGTCAATC	Yeast validation
Eucgr.K02297 (MYB52/54A)_F	ATGTGCACCAGAGGCCACTG	Yeast validation
Eucgr.K02297 (MYB52/54A)_R	CTAACAAGAGCTTCTGACCGATA	Yeast validation
Egcrg.F04277 (MYB52/54B)_F	ATGGACAATTCCAGACCTAACA	Yeast validation
Egcrg.F04277 (MYB52/54B)_	TCAAGATGTGATTCCTACCCCAAGAAA	Yeast validation

[a] Bold lettering indicates the *Mlu*I (A˅CGCGT) and *Spe*I (A˅CTAGT) restriction sites incorporated in the primers used to amplify the baits for directional cloning. All sequences listed from 5' to 3'.

[b]Indicates primers that were used to amplify baits (figure 4.1).

Table 4.2 Transcription factors selected for full-length cloning and Y1H screening of *EgrCesA8* baits.

Eucalyptus gene model	*Eucalyptus* homolog[b]	At number	*Arabidopsis* homolog	Gene family	References[c]
Eucgr.F02796	*EgrZincFinger* -A	At1g66810	*AtC3H14*	C3H14 zinc finger	Ko et al. 2009
Eucgr.B00047	*EgrZincFinger* -B	At1g72220	no name	C3H4 zinc finger	-[d]
Eucgr.D01935	*EgrKNAT7*	At1g62990	*AtKNAT7*	KNOX class II	Li et al. 2012
Eucgr.B03684	*EgrMYB31*	At5g12870	*AtMYB46*	R2R3 MYB	Zhong et al. 2007
[a]Eucgr.F02756	*EgrMYB80*	At1g17950/At1g73410	*AtMYB52*	R2R3 MYB	Zhong et al. 2008
[a]Eucgr.F04277.1	*EgrMYB87*	At1g17950/At1g73410	*AtMYB54*	R2R3 MYB	Zhong et al. 2008
Eucgr.G03385	*EgrMYB2*	At3g08500	*AtMYB83*	R2R3 MYB	McCarthy et al. 2009
Eucgr.D02014	*EgrMYB61*	At4g22680	*AtMYB85*	R2R3 MYB	Zhong et al. 2008
Eucgr.D01819	*EgrMYB60*	At1g63910	*AtMYB103*	R2R3 MYB	Zhong et al. 2008
Eucgr.E01053	*EgrNAC61*	At1g32770	*AtNST3/AtSND1*	NAC domain	Zhong et al. 2006, Mitsuda et al. 2008
Eucgr.D01671	*EgrNAC49*	At2g46770	*AtNST1*	NAC domain	Mitsuda et al. 2005
Eucgr.K01061	*EgrNAC170*	At4g28500	*AtSND2*	NAC domain	Zhong et al. 2008, Hussey et al. 2011
Eucgr.E03226.1	*EgrNAC64*	At1g28470	*AtSND3*	NAC domain	Zhong et al. 2008
Eucgr.A02887.1	*EgrNAC26*	At5g62380	*AtVND6*	NAC domain	Kubo et al. 2005
Eucgr.F02615	*EgrNAC75*	At1g71930	*AtVND7*	NAC domain	Kubo et al. 2005

[a]For *At*MYB52 and *At*MYB54 it was not possible to distinguish between the *Eucalyptus* homologs EucgrF02756 and EucgrF04277.1.Therefore *Egr*MYB80 and *Egr*MYB87 may be interchangeable in this study (Botha et al. *personal communication*).

[b]Naming conventions of the *Eucalyptus* NAC and MYB family proteins follow the naming conventions proposed in companion papers to the *Eucalyptus* genome paper which are currently in preparation (NAC; Hussey et al., MYB; Solar et al. *both in preparation*)

[c]References reporting the *Arabidopsis* transcription factors involved in secondary cell wall (SCW) formation.

[d]The involvement of this gene in the SCW regulatory network was not previously published but was implicated in the process during in house expression analysis of *Eucalyptus* transcriptomes (Mizrachi, *personal communication*).

Table 4.3 Yeast cell counts per well for Y1H screening.

Genes screened	[a]Average number of cells/spot (+/-SD)
*Egr*ZincFinger-A	2371.3 (+/- 578.6)
*Egr*ZincFinger –B	4266.8 (+/- 689.0)
*Egr*KNAT7	9665.1 (+/- 2406.1)
*Egr*MYB31	7762.1 (+/- 843.7)
*Egr*MYB80	8285.4 (+/- 1857.7)
*Egr*MYB87	8229.0 (+/- 2791.7)
*Egr*MYB61	5944.7 (+/- 1182.0)
*Egr*MYB60	3516.2 (+/- 1060.8)
*Egr*NAC49	2192.2 (+/- 433.2)
*Egr*NAC170	6045.9 (+/- 1173.7)
*Egr*NAC64	7250.0 (+/- 3295.4)
*Egr*NAC26	7155.3 (+/- 951.8)
[b]*Egr*MYB80:*Egr*KNAT7	8937.3 (+/- 1407.5)
[b]*Egr*MYB87:*Egr*KNAT7	8937.3 (+/- 1407.5)

[a]The average number of cells per 5 µl spot plated on each of the different drop out media and the overall standard deviation for each gene is represented in brackets.

[b]The dual yeast screens containing *Egr*KNAT7 and either *Erg*MYB80 or *Egr*MYB87 were performed as one screen with the same controls and therefore have identical values.

Table 4.4 Summary of the bait (columns A-H) and the prey vectors (rows) screened in Y1H and the single or dual interactions observed.

Name and gene model		Y1H interactions per bait line							
		A	B	C	D	E	F	G	H
*Egr*ZincFinger -A	Eucgr.F02796	[b]++	0	0	0	0	0	0	0
*Egr*ZincFinger -B	Eucgr.B00047	0	0	0	0	0	0	0	0
*Egr*KNAT7	Eucgr.D01935	0	0	0	[c]TTT	0	0	0	0
*Egr*MYB31	Eucgr.B03684	[b]+++	0	0	0	0	0	0	0
*Egr*MYB80	Eucgr.F02756	0	0	0	[c]TTT	0	0	0	0
*Egr*MYB87	Eucgr.F04277.1	0	0	0	[c]TTT	0	0	0	0
[a]*Egr*MYB2	Eucgr.G03385	-	-	-	-	-	-	-	-
*Egr*MYB61	Eucgr.D02014	0	0	0	0	0	0	0	0
*Egr*MYB60	Eucgr.D01819	0	0	0	0	0	0	0	0
[a]*Egr*NAC61	Eucgr.E01053	-	-	-	-	-	-	-	-
*Egr*NAC49	Eucgr.D01671	0	0	0	0	0	0	0	0
*Egr*NAC170	Eucgr.K01061	0	0	0	0	0	[b]+	0	0
*Egr*NAC64	Eucgr.E03226.1	0	0	0	0	0	0	0	0
*Egr*NAC26	Eucgr.A02887.1	0	0	0	0	0	0	0	0
[a]*Egr*NAC75	Eucgr.F02615	-	-	-	-	-	-	-	-

[a]Dashes indicate Bait (A-H) and Prey combinations that could not be screened due to lethality of the prey in yeast cells.

[b]Plus sign (+) indicates positive Y1H interactions observed with a single bait and prey combination observed on -Trp-Leu-His TDO media. (++) indicate interactions observed on TDO media supplemented with 1-5 mM 3AT and (+++) indicates interactions observed on TDO media supplemented with 5-10 mM 3AT

[c]TTT indicates strong positive interactions observed for dual prey–bait screens on TDO media supplemented with 5-10 mM 3AT, where both *Egr*KNAT7 and *Egr*MYB80 or *Egr*MYB87 were required to be present together for yeast-1-hybrid interaction to occur.

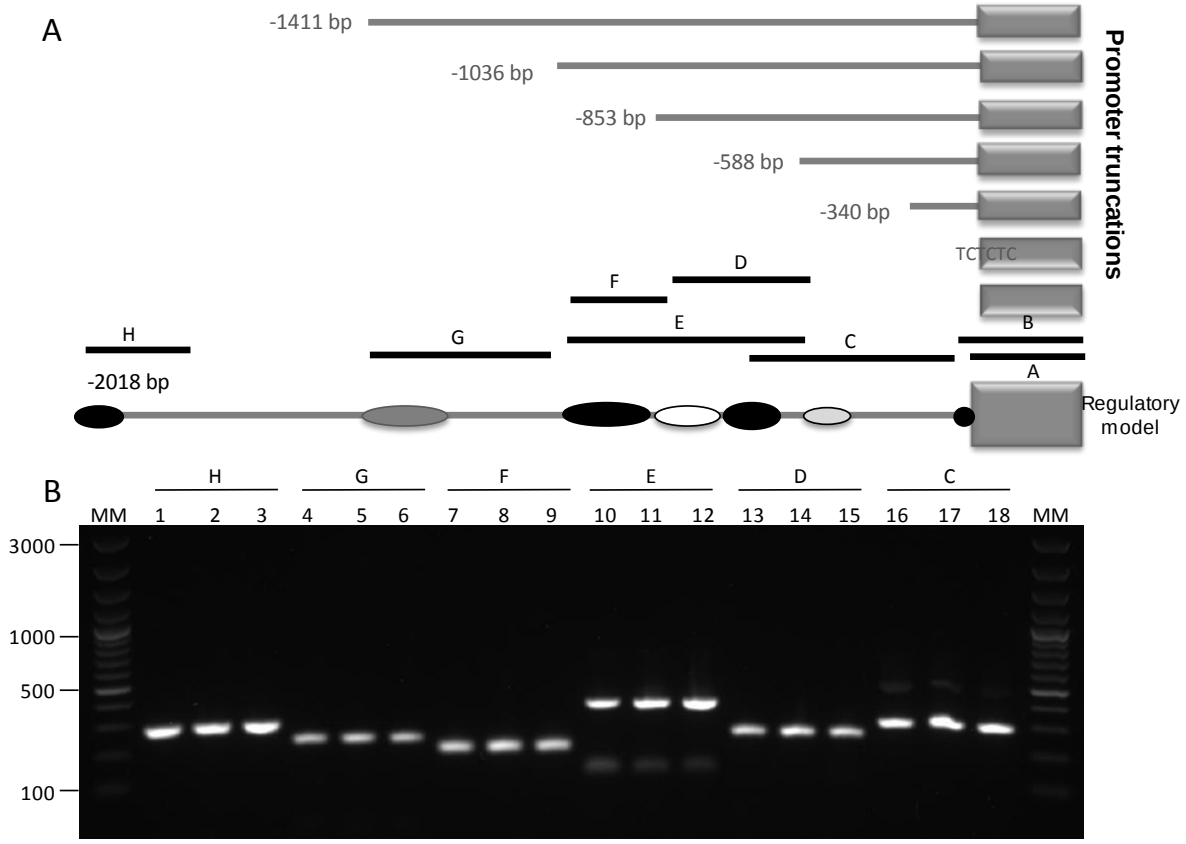


Figure 4.1 Yeast-1-hybrid (Y1H) bait positions along the length of the *EgrCesA8* promoter. (A) Adapted from Figure 3.7 indicating the *EgrCesA8* promoter truncations (grey lines) and position of the 5'UTR (grey block). The longest *EgrCesA8* promoter fragment at the bottom of the image shows the regions identified as activators (black oval), repressors (grey oval), phloem-specific repressor (white oval). The tension wood element identified by Lu et al. (2008) is indicated by the light grey oval edged in black. The black lines represent the different Y1H baits (A to H) designed to incorporate or remove various regulatory modules identified in Chapter 3. (B) Amplification of *EgrCesA8* promoter bait fragments for cloning and subsequent Y1H screening. Three replicates were amplified for each bait (A-H) as listed at the top of the gel image and fragments of the correct size was visualised in each lane 1-18. A 100 bp size standard was used for size determination (MM).

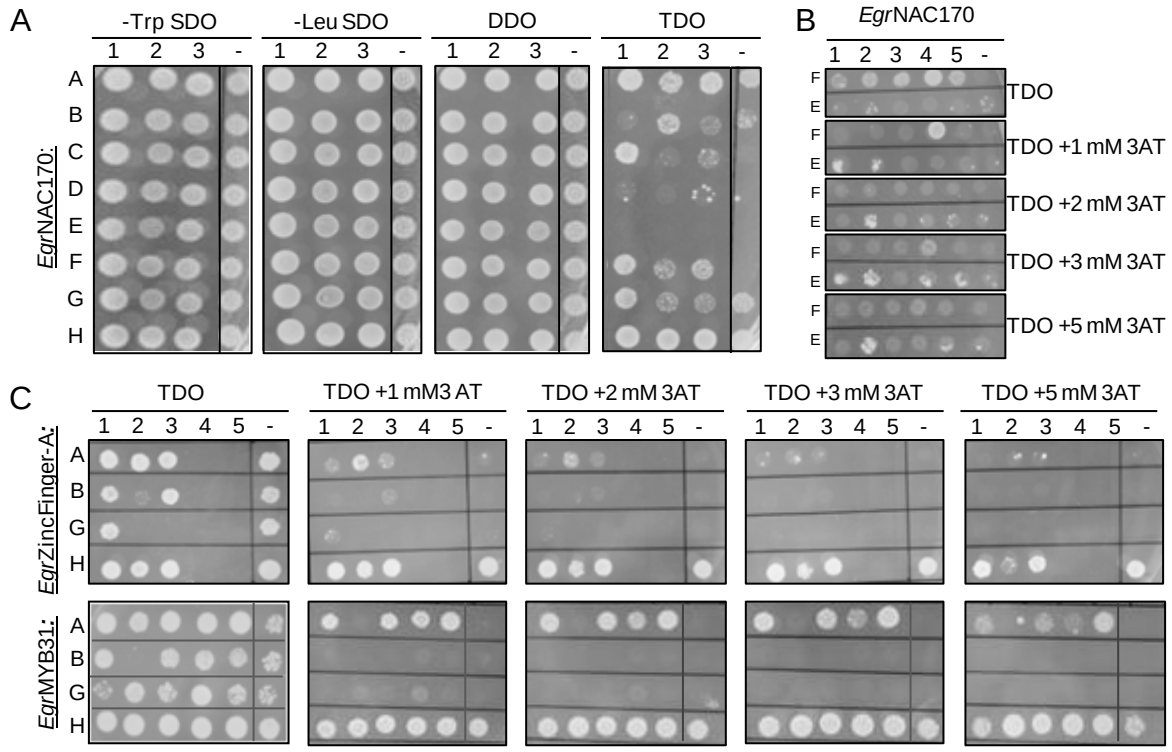


Figure 4.2 Yeast-1-hybrid (Y1H) screening of *EgrCesA8* baits (A-H) against three secondary cell wall (SCW)-related transcription factors. (**A**) Y1H screen with positive interaction on media lacking either tryptophan (-Trp; SDO) or leucine (-Leu; SDO) or both (-Trp-Leu; DDO). The interaction is evident on the triple drop out (TDO) plate lacking tryptophan, leucine and histidine (-Trp-Leu-His) row E where no growth is observed on the TDO. Growth is observed with bait F and *Egr*NAC*170* indicating interaction and no growth for the negative control (bait F and empty prey vector) confirms HIS expression. Auto-activation of some baits (A, B, G and H) was observed on TDO where the negative controls showed growth on the TDO plates. (**B**) Testing the strength of the *Egr*NAC*170*-bait F interaction. Five colonies were screened (1-5) and a negative control was included (-). Both bait F and E with *Egr*NAC*170* are shown. *Egr*NAC*170*-bait E show only some sparse background growth and *Egr*NAC*170*-bait F grows well on TDO, but growth drops off as 3AT concentration increases (right), indicating a weak interaction. (**C**) 3-amino-1,2,4-triazole (3AT) screen to reduce the effect of leaky HIS expression and observe interactions for the baits (A, B, G and H). Interactions between *EgrCesA8* promoter bait A and *Egr*ZincFinger-A or *Egr*MYB31 were observed when increasing concentrations of 3AT were added as indicated above each set of panels. Three to five independent colonies were screened for each prey and bait combination (Supplementary files 4.1 and 4.2).

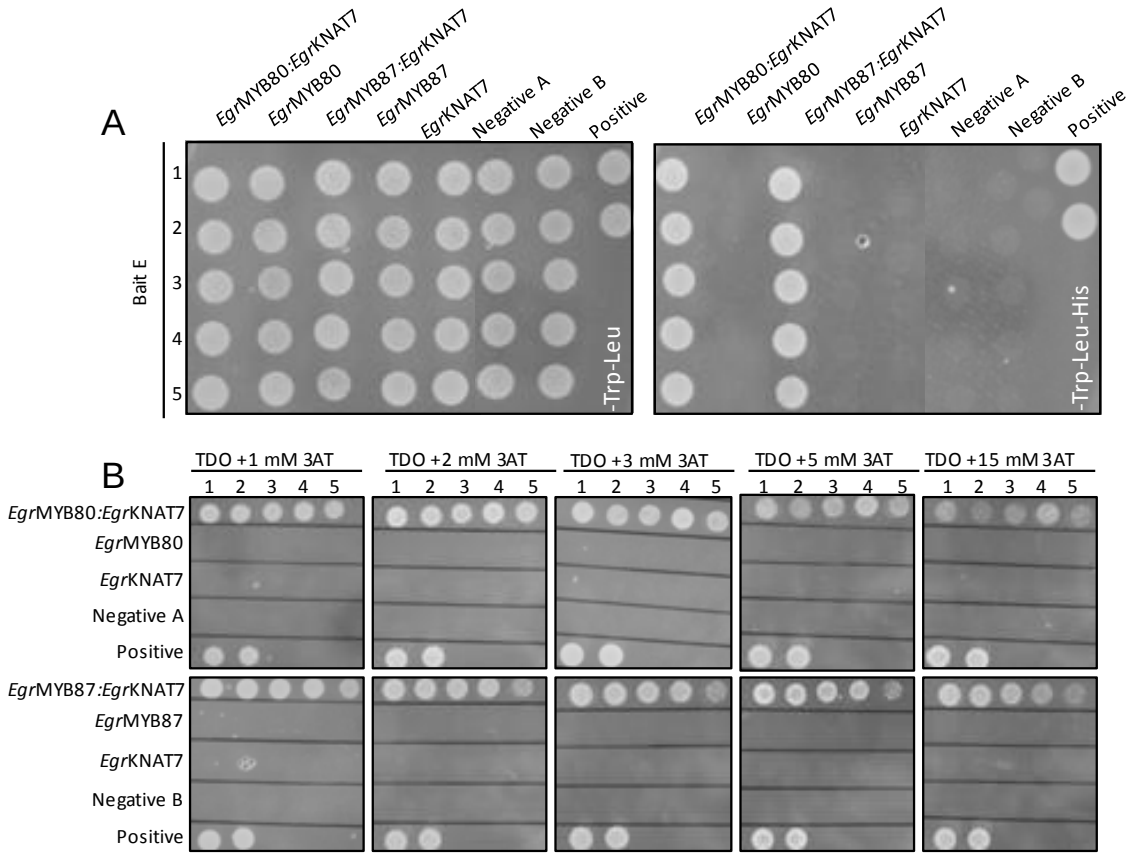
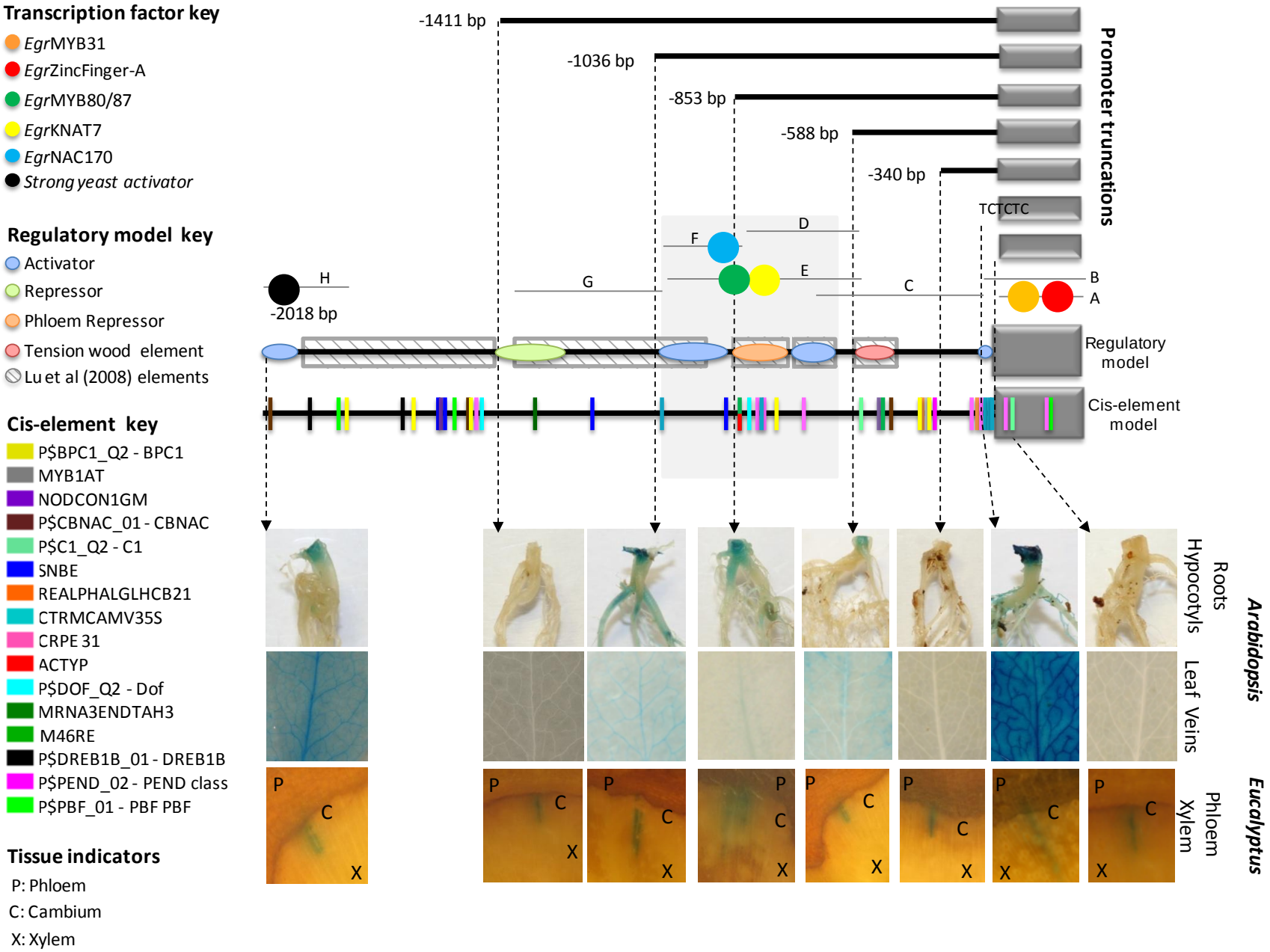


Figure 4.3 Dual interactions of *Egr*KNAT7 and *Egr*MYB80 or *Egr*MYB87 with *EgrCesA8* promoter fragment E. (A) All lines including the single lines with bait E and only *Egr*KNAT7 or *Egr*MYB80 or *Egr*MYB87 and the negative controls with only bait E grew well on the double drop out media (-Trp-Leu, DDO; indicated bottom left of each panel). Only the dual lines with both prey vectors (*Egr*MYB80 or *Egr*MYB87 and *Egr*KNAT7) and bait E showed growth on the triple dropout media (-Trp-Leu-His, TDO; bottom right panel). (B) Top row of panels show interaction with bait E of both *Egr*KNAT7 and *Egr*MYB80 on TDO media with varying concentrations (indicated above the panel) of 3-amino-1,2,4-triazole (3AT). Similarly, the second row of panels tests the interaction of bait E with *Egr*KNAT7 and *Egr*MYB87. In both (A) and (B) Five independent colonies were screened to replicate results (1-5 in each case). Two levels of negative control were used; firstly the strain containing just one of the transcription factors (*Egr*KNAT7 or *Egr*MYB80 or *Egr*MYB87) and bait E were plated out to show that individually these transcription factors do not interact with bait E. Secondly two independent transformations (Negative A and B) containing and empty pDEST-pGAD vector and bait E indicate no auto-activation. Two independent positive control for Y1H interaction lines containing the human P53 gene (pGADT7 vector) and matching *cis*-element (pHIS2.1 vector) shows a typical strong Y1H interaction.

Figure 4.4 Summary of the *cis*-element, yeast-1-hybrid (Y1H), *Arabidopsis* and *Eucalyptus GUS* data for the *EgrCesA8* promoter and its assorted truncations. Black bars represent the full-length *EgrCesA8* promoter or various promoter truncations and the blue blocks are the 5'UTRs. Dashed arrows show which regions of the full length promoter were removed during truncation. Colour lines on the full-length promoter represent different *cis*-elements previously identified in this promoter sequence (Figure 4.1) and the *cis*-element key is depicted on the bottom left. Light colour ovals on the full length promoter indicate the regulatory model predicted from GUS data and grey lined boxes indicate Lu et al. (2008) GUS results. Colour key for the regulatory model is indicated on the left. Different length dark grey lines above the full-length promoter indicate the different bait regions (A-H) screened with Y1H and the coloured circles show the transcription factors that interact with the different bait fragments in Y1H (Transcription factor key left). Grey shading indicates the *EgrCesA8* CRM. First row of images below the full length promoter show the average GUS staining produced by each promoter truncate (Full-length promoter to UTR indicated from left to right). Bottom images represent the different GUS staining patterns produced by the promoter truncations in *Eucalyptus* stem tissues where phloem (P), cambium (C) and xylem (X) are shown**.**



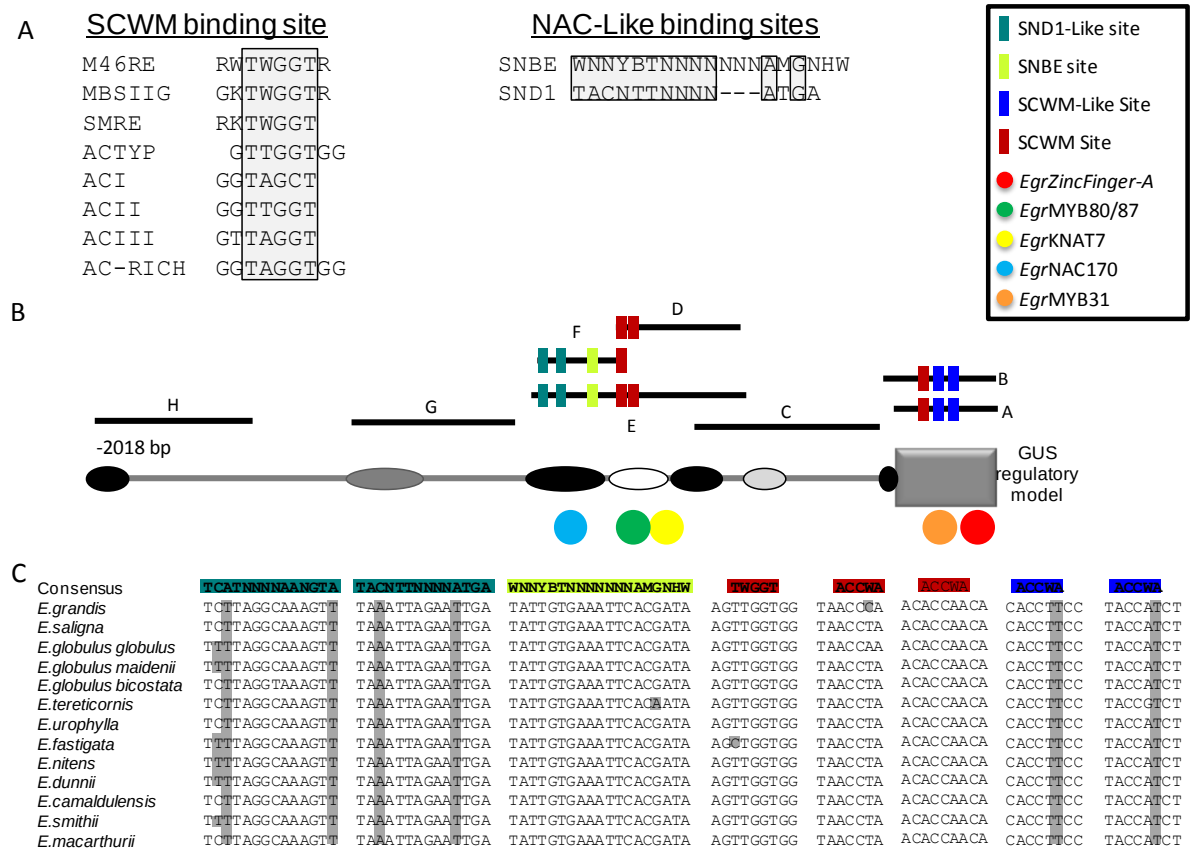


Figure 4.5 NAC and MYB binding sites mapped to specific *EgrCesA8* promoter bait regions showing interaction with NAC and MYB transcription factors. (A) The core secondary cell wall MYB (SCWM) binding site (TWGGT:ACCWA) identified by aligning all previously reported secondary cell wall (SCW) MYB binding sites (Kim et al. 2012, Zhong et al. 2012, Winzell et al. 2010, Hutton et al. 1995, Romero et al. 1998, Goicoechea et al. 2005, Raes et al. 2003). NAC-like binding sites indicate the conservation of previously identified SCW binding NAC elements (Zhong et al. 2010, Wang et al. 2011). (B) Adapted from the figure 3.7 (Chapter 3), indicating the *EgrCesA8* promoter truncations (grey lines) and position of the 5'UTR (grey block). The image also shows the regions identified as activators (black oval), repressors (grey oval) and phloem-specific repressor (white oval). Tension wood element identified by Lu et al. (2008) indicated by the light grey oval edged in black. Black lines represent the different yeast-1-hybrid (Y1H) baits (A to H) with four previously identified *cis*-elements (SNBE: light green, SND1-like: turquoise, SCWM: maroon and SCWM-like: blue) mapped to the different baits. Dots indicate the interacting transcription factors and the positions they interact at. (C) The conservation of the MYB and NAC binding site occurrences across the *CesA8* promoter of 13 different *Eucalyptus* species (listed left). The consensus sequence of each *cis*-element from A and B is highlighted according to *cis*-element key. The grey shading indicates where promoter sequences do not match the consensus sequence colour.

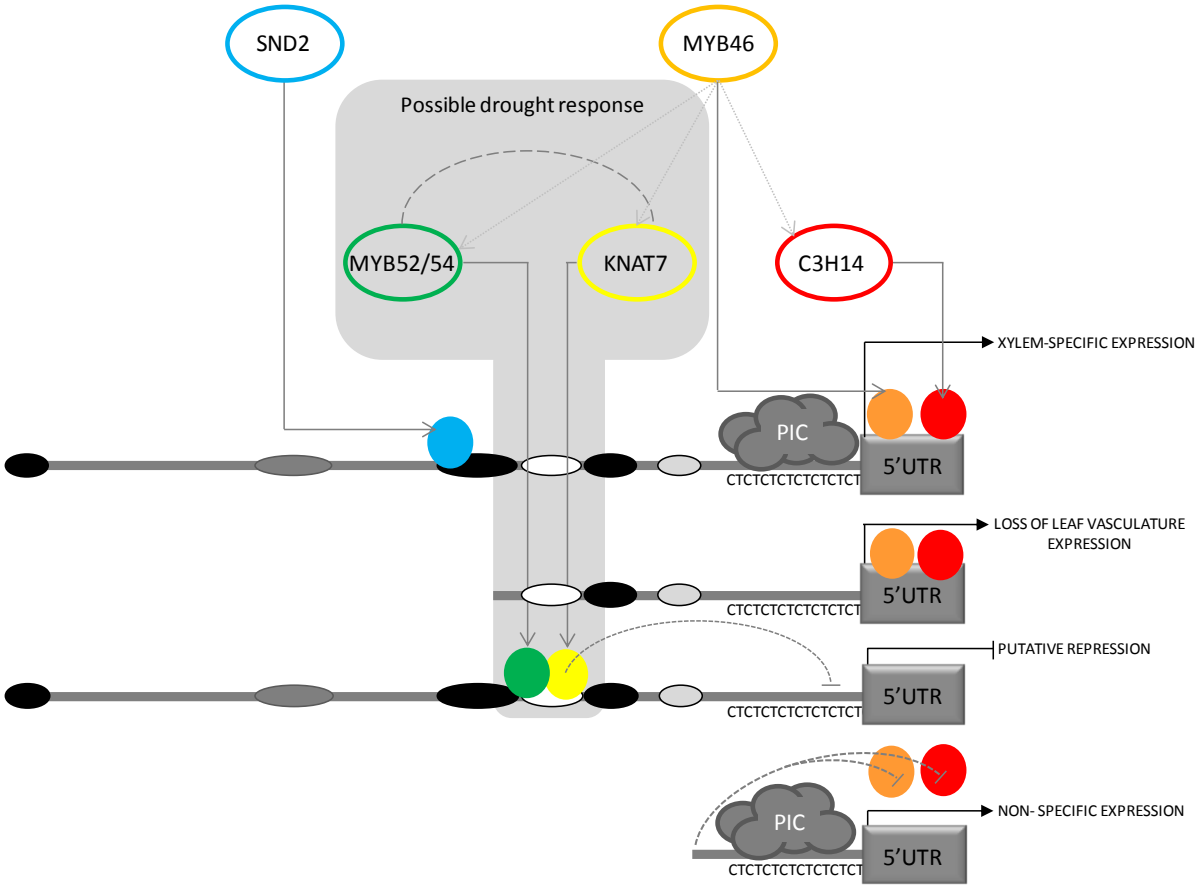


Figure 4.6 Model of *EgrCesA8* gene regulation and links to the current secondary cell wall regulatory network. The grey shaded region assigns a putative drought response described in literature (Cal et al. 2013, Park et al. 2011). The coloured ovals with gene names represent the proteins used in the yeast-1-hybrid (Y1H) study that interacted with the *EgrCesA8* promoter. The dashed line indicates a putative protein-protein interaction identified by a dual Y1H screen. The matching coloured circles and solid grey arrows indicate where in the promoter each protein interacts as revealed by this current study. The *EgrCesA8* promoter and truncations (grey lines), position of the 5'UTR (grey block) and the pre-initiation complex (PIC; Grey cloud) are indicated below the network. The image also shows the regions in the promoter identified as activators (black oval), repressors (grey oval), phloem-specific repressor (white oval) and the tension wood element identified by Lu et al. (2008) indicated by the light grey oval edged in black. The $CT_{(11)}$-microsatellite is indicated by a run of CT's before the transcriptional start site (black arrow). The wording on the right indicates the possible outcomes of various interactions and deletions to the promoter. Dotted grey lines indicate putative repression or hindrance activity in this regulatory model.

4.8 References

Appenzeller L, Doblin M, Barreiro R, Wang HY, Niu XM, Kollipara K, Carrigan L, Tomes D, Chapman M, Dhugga KS (2004) Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (*CesA*) gene family. Cellulose 11: 287-299

Berger N, Dubreucq B (2012) Evolution goes GAGA: GAGA binding proteins across kingdoms. Biochim. Biophys. Acta. 1819: 863-868

Betancur L, Singh B, Rapp RA, Wendel JF, Marks MD, Roberts AW, Haigler CH (2010) Phylogenetically distinct cellulose synthase genes support secondary wall thickening in *Arabidopsis* shoot trichomes and cotton fiber. J. Integr. Plant Biol. 52: 205-220

Bhandari S, Fujino T, Thammanagowda S, Zhang D, Xu F, Joshi CP (2006) Xylem-specific and tension stress-responsive coexpression of KORRIGAN endoglucanase and three secondary wall-associated cellulose synthase genes in aspen trees. Planta 224: 828-837

Bhardwaj N, Lu H (2005) Correlation between gene expression profiles and protein–protein interactions within and across genomes. Bioinformatics 21: 2730-2738

Bhargava A, Ahad A, Wang S, Mansfield SD, Haughn GW, Douglas CJ, Ellis BE (2013) The interacting MYB75 and KNAT7 transcription factors modulate secondary cell wall deposition both in stems and seed coat in *Arabidopsis*. Planta 247: 1199-1211

Bhargava A, Mansfield SD, Hall HC, Douglas CJ, Ellis BE (2010) MYB75 functions in regulation of secondary cell wall formation in the *Arabidopsis* inflorescence stem. Plant Physiol. 154: 1428-1438

Bolle C, Sopory S, Lübberstedt T, Herrmann R, Oelmüller R (1994) Segments encoding 5′-untranslated leaders of genes for thylakoid proteins contain *cis*-elements essential for transcription. Plant J. 6: 513-523

Burton RA, Shirley NJ, King BJ, Harvey AJ, Fincher GB (2004) The *CesA* gene family of barley. Quantitative analysis of transcripts reveals two groups of co-expressed genes. Plant Physiol. 134: 224-236

Cal AJ, Liu D, Mauleon R, Hsing Y-IC, Serraj R (2013) Transcriptome profiling of leaf elongation zone under drought in contrasting rice cultivars. PloS ONE 8: e54537

Cassan-Wang H, Goue N, Saidi MN, Legay S, Sivadon P, Goffner D, Grima-Pettenati J (2013) Identification of novel transcription factors regulating secondary cell wall formation in *Arabidopsis*. Front. Plant Sci. 4: 189

Castillejo C, Pelaz S (2008) The balance between CONSTANS and TEMPRANILLO activities determines FT expression to trigger flowering. Curr. Biol. 18: 1338-1343

Chang S, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. Plant Mol. Biol. Rep. 11: 113-116

Creux NM, Ranik M, Berger DK, Myburg AA (2008) Comparative analysis of orthologous cellulose synthase promoters from *Arabidopsis*, *Populus* and *Eucalyptus*: evidence of conserved regulatory elements in angiosperms. New Phytol. 179: 722-737

David-Assael O, Saul H, Saul V, Mizrachy-Dagri T, Berezin I, Brook E, Shaul O (2005) Expression of *At*MHX, an *Arabidopsis* vacuolar metal transporter, is repressed by the 5′ untranslated region of its gene. J. Exp. Bot. 56: 1039-1047

Demura T, Ye Z-H (2010) Regulation of plant biomass production. Curr. Opin. Plant Biol. 13: 298-303

Desprez T, Juraniec M, Crowell EF, Jouy H, Pochylova Z, Parcy F, Hofte H, Gonneau M, Vernhettes S (2007) Organization of cellulose synthase complexes involved in primary cell wall synthesis in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 104: 15572-15577

Ding J, Hu H, Li X (2012) Thousands of *cis*-regulatory sequence combinations are shared by *Arabidopsis* and poplar. Plant Physiol. 158: 145-155

Djerbi S, Lindskog M, Arvestad L, Sterky F, Teeri TT (2005) The genome sequence of black cottonwood (*Populus trichocarpa*) reveals 18 conserved cellulose synthase (*CesA*) genes. Planta 221: 739-746

Goicoechea M, Lacombe E, Legay S, Mihaljevic S, Rech P, Jauneau A, Lapierre C, Pollet B, Verhaegen D, Chaubet-Gigot N, Grima-Pettenati J (2005) *Eg*MYB2, a new transcriptional activator from *Eucalyptus* xylem, regulates secondary cell wall formation and lignin biosynthesis. Plant J. 43: 553-567

Hackbusch J, Richter K, Müller J, Salamini F, Uhrig JF (2005) A central role of *Arabidopsis thaliana* ovate family proteins in networking and subcellular localization of 3-aa loop extension homeodomain proteins. Proc. Natl. Acad. Sci. USA 102: 4908-4912

Hamann T, Osborne E, Youngs HL, Misson J, Nussaume L, Somerville C (2004) Global expression analysis of *CESA* and *CSL* genes in *Arabidopsis*. Cellulose 11: 279-286

Hammerschmidt R, Lamport D, Muldoon E (1984) Cell wall hydroxyproline enhancement and lignin deposition as an early event in the resistance of cucumber to *Cladosporium cucumerinum*. Physiol. Plant Pathol. 24: 43-47

Handakumbura PP, Hazen SP (2012) Transcriptional regulation of grass secondary cell wall biosynthesis: playing catch-up with *Arabidopsis thaliana*. Front. Plant Sci. 3:74

Harpaz-Saad S, McFarlane HE, Xu S, Divi UK, Forward B, Western TL, Kieber JJ (2011) Cellulose synthesis via the FEI2 RLK/SOS5 pathway and CELLULOSE SYNTHASE 5 is required for the structure of seed coat mucilage in *Arabidopsis*. Plant J. 68: 941-953

Hazen S (2012) Transcriptional regulation of lignin and cellulose biosynthesis in *Brachypodium distachyon*. Plant and Animal Genome XX Conference (January 14-18, 2012). Plant and Animal Genome

Hazen SP, Scott-Craig JS, Walton JD (2002) Cellulose synthase-like genes of rice. Plant Physiol. 128: 336-340

Hernández-Blanco C, Feng DX, Hu J, Sánchez-Vallet A, Deslandes L, Llorente F, Berrocal-Lobo M, Keller H, Barlet X, Sánchez-Rodríguez C (2007) Impairment of cellulose synthases required for *Arabidopsis* secondary cell wall formation enhances disease resistance. Plant Cell 19: 890-903

Hussey SG, Mizrachi E, Spokevicius AV, Bossinger G, Berger DK, Myburg AA (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 Biol. 11: 173

Immink RG, Tonaco IA, de Folter S, Shchennikova A, van Dijk AD, Busscher-Lange J, Borst JW, Angenent GC (2009) SEPALLATA3: the 'glue' for MADS box transcription factor complex formation. Genome Biol. 10: R24

Kaufmann K, Pajoro A, Angenent GC (2010) Regulation of transcription in plants: mechanisms controlling developmental switches. Nature Rev. Genet. 11: 830-842

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

Kim W-C, Ko J-H, Kim J-Y, Kim J, Bae H-J, Han K-H (2013) MYB46 directly regulates the gene expression of secondary wall-associated cellulose synthases in *Arabidopsis*. Plant J. 73: 26-36

Ko JH, Kim WC, Han KH (2009) Ectopic expression of MYB46 identifies transcriptional regulatory genes involved in secondary wall biosynthesis in *Arabidopsis*. Plant J. 60: 649-665

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

Legay S, Sivadon P, Blervacq AS, Pavy N, Baghdady A, Tremblay L, Levasseur C, Ladouce N, Lapierre C, Séguin A, Hawkins S, Mackay J, Grima-Pettenati J (2010) *Eg*MYB1, an R2R3 MYB transcription factor from *Eucalyptus* negatively regulates secondary cell wall formation in *Arabidopsis* and poplar. New Phytol. 188: 774-786

Lei L, Li S, Gu Y (2012) Cellulose synthase complexes: composition and regulation. Front. Plant Sci. 3:75

Lelli K, Slattery M, Mann R (2012) Disentangling the many layers of eukaryotic transcriptional regulation. Annu. Rev. Genet. 46: 43-64

Levine M, Davidson EH (2005) Gene regulatory networks for development. Proc. Natl. Acad Sci.USA 102: 4936-4942

Li E, Bhargava A, Qiang W, Friedmann MC, Forneris N, Savidge RA, Johnson LA, Mansfield SD, Ellis BE, Douglas CJ (2012) The Class II KNOX gene KNAT7 negatively regulates secondary wall formation in *Arabidopsis* and is functionally conserved in *Populus*. New Phytol. 194: 102-115

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

Liao L, Ning G, Liu C, Zhang W, Bao M (2013) The intron from the 5′-UTR of the *FBP11* gene in petunia displays promoter-and enhancer-like functions. Sci. Hortic. 154: 96-101

Lu S, Li L, Yi X, Joshi CP, Chiang VL (2008) Differential expression of three *Eucalyptus* secondary cell wall-related cellulose synthase genes in response to tension stress. J. Exp. Bot. 59: 681-695

Maritz-Olivier C (2005) The mechanisms regulating exocytosis of the salivary glands of the soft tick, *Ornithodoros savignyi*. University of Pretoria

McCarthy RL, Zhong R, Ye Z-H (2011) Secondary wall NAC binding element (SNBE), a key *cis*-acting element required for target gene activation by secondary wall NAC master switches. Plant Signal. Behav. 6: 1282-1285

Mizrachi E, Mansfield SD, Myburg AA (2012) Cellulose factories: advancing bioenergy production from forest trees. New Phytol. 194: 54-62

Obembe OO, Jacobsen E, Vincken J-P, Visser RG (2009) Differential expression of cellulose synthase (*CesA*) gene transcripts in potato as revealed by QRT-PCR. Plant Physiol. Bioch. 47: 1116-1118

Ohashi-Ito K, Oda Y, 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

Park MY, Kang J-y, Kim SY (2011) Overexpression of *At*MYB52 confers ABA hypersensitivity and drought tolerance. Mol. Cells 31: 447-454

Patzlaff A, Newman LJ, Dubos C, Whetten RW, Smith C, McInnis S, Bevan MW, Sederoff RR, Campbell MM (2003) Characterisation of *Pt*MYB1, an R2R3-MYB from pine xylem. Plant Mol. Biol. 53: 597-608

Rahantamalala A, Rech P, Martinez Y, Chaubet-Gigot N, Grima-Pettenati J, Pacquit V (2010) Coordinated transcriptional regulation of two key genes in the lignin branch pathway-CAD and CCR-is mediated through MYB-binding sites. BMC Plant Biol. 10: 130

Ramírez V, Agorio A, Coego A, García-Andrade J, Hernández MJ, Balaguer B, Ouwerkerk PB, Zarra I, Vera P (2011a) MYB46 modulates disease susceptibility to *Botrytis cinerea* in *Arabidopsis*. Plant Physiol. 155: 1920-1935

Ramírez V, García-Andrade J, Vera P (2011b) Enhanced disease resistance to *Botrytis cinerea* in MYB46 *Arabidopsis* plants is associated to an early down-regulation of *CesA* genes. Plant Signal. Behav. 6: 911-913

Ramsay NA, Glover BJ (2005) MYB–bHLH–WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci. 10: 63-70

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

Richmond TA, Somerville CR (2000) The cellulose synthase superfamily. Plant Physiol. 124: 495-498

Rossignol P, Collier S, Bush M, Shaw P, Doonan JH (2007) *Arabidopsis* POT1A interacts with TERT-V (I8), an N-terminal splicing variant of telomerase. J. Cell Sci. 120: 3678-3687

Schaart JG, Dubos C, Romero De La Fuente I, Houwelingen AM, Vos RC, Jonker HH, Xu W, Routaboul JM, Lepiniec L, Bovy AG (2013) Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (*Fragaria*× *ananassa*) fruits. New Phytol. 197: 454-467

Seifert GJ, Blaukopf C (2010) Irritable walls: the plant extracellular matrix and signaling. Plant Physiol. 153: 467-478

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

Smaczniak C, Immink RG, Muiño JM, Blanvillain R, Busscher M, Busscher-Lange J, Dinh QP, Liu S, Westphal AH, Boeren S (2012) Characterization of MADS-domain transcription factor complexes in *Arabidopsis* flower development. Proc. Natl. Acad. Sci. USA 109: 1560-1565

Stork J, Harris D, Griffiths J, Williams B, Beisson F, Li-Beisson Y, Mendu V, Haughn G, DeBolt S (2010) CELLULOSE SYNTHASE9 serves a nonredundant role in secondary cell wall synthesis in *Arabidopsis* epidermal testa cells. Plant Physiol. 153: 580-589

Suzuki S, Li LG, Sun YH, Chiang VL (2006) The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthase-like genes in *Populus trichocarpa*. Plant Physiol. 142: 1233-1245

Taylor NG (2008) Cellulose biosynthesis and deposition in higher plants. New Phytol. 178: 239-252

Tronchet M, BalaguÉ C, Kroj T, Jouanin L, Roby D (2010) Cinnamyl alcohol dehydrogenases-C and D, key enzymes in lignin biosynthesis, play an essential role in disease resistance in *Arabidopsis*. Mol. Plant Pathol. 11: 83-92

Wang H, Zhao Q, Chen F, Wang M, Dixon RA (2011) NAC domain function and transcriptional control of a secondary cell wall master switch. Plant J. 68: 1104-1114

Welner DH, Lindemose S, Mollegaard NE, Olsen AN, Helgstrand C, Skriver K, Lo Leggio L (2012) DNA binding by the plant-specific NAC transcription factors in crystal and solution: a firm link to WRKY and GCM transcription factors. Biochem. J. 444: 395-404

Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Prüß M, Reuter I, Schacherer F (2000) TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Res. 28: 316-319

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

Wu A-M, Hu JS, Liu J-Y (2009) Functional analysis of a cotton cellulose synthase A4 gene promoter in transgenic tobacco plants. Plant Cell Rep. 28: 1539-1548

Xie L, Yang C, Wang X (2011) Brassinosteroids can regulate cellulose biosynthesis by controlling the expression of *CESA* genes in *Arabidopsis*. J. Exp. Bot. 62: 4495-4506

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

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

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

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

Chapter 5

Concluding Remarks

5.1 Identification of core promoter elements in the *Eucalyptus CesA* promoters

An important feature of all eukaryotic promoters is the core promoter which harbours the elements required for recruitment of the pre-initiation complex (PIC) and initiation of transcription (Lenhard et al. 2012). Our study is the first study to investigate the core promoter elements of the cellulose synthase (*CesA*) genes in any plant species (Chapter 2). We found that the *Eucalyptus CesA* promoters are TATA-less, and may use alternative core promoter elements such as those described by Bernard et al. (2010). A number of putative core promoter elements were identified, such as CTRMCAM35S and CRPE31 which were often clustered together near the transcriptional start site (TSS) in regions of low nucleotide diversity (Chapter 2). CTRMCAM35S and CRPE31 were similar in sequence and position to alternative core promoter elements such as the plant-specific Y-patch and the initiator element (Yamamoto et al. 2007). These promoter elements could represent the core elements of the *CesA* promoters from which transcription is initiated.

The CTRMCAM35S element is composed of four CT dinucleotide repeats and when found in close succession produced a CT-microsatellite within the promoter regions of several of the *Eucalyptus CesA* promoters. CT-microsatellites were observed in close proximity to the TSS for several of the *Eucalyptus CesA* promoters analysed (Chapter 2). The secondary cell wall (SCW)-associated *CesA* promoter, *EgrCesA8*, was used as an example to test the conserved promoter elements identified in Chapter 2 such as the CTRMCAM35S, CRPE31 and the upstream regions of low nucleotide diversity. To investigate the role of the $CT_{(11)}$-microsatellite at the TSS of the *EgrCesA8* promoter, two GUS constructs were produced, one containing the $CT_{(11)}$-microsatellite appended to the 5'UTR and the second containing only the 5'UTR. The $CT_{(11)}$-microsatellite appended 5'UTR construct expressed GUS in a non-specific manner in *Arabidopsis* plants and *Eucalyptus* stems. However, the 5'UTR-only construct could not express GUS in *Arabidopsis* plants (Chapter 3). These two

constructs ($CT_{(11)}$-microsatellite appended 5'UTR or 5'UTR-only) were also used in a yeast-1-hybrid (Y1H) screen and two transcription factors, *Egr*MYB31 and *Egr*ZincFinger-A, interacted with the 5'UTR-only construct suggesting that the $CT_{(11)}$-microsatellite is hindering DNA-protein interactions in this system (Chapter 4). This result could explain the loss of tissue specificity observed in plant lines containing the $CT_{(11)}$-microsatellite appended 5'UTR-GUS construct. The lack of upstream sequence elements may impede the correct recruitment of the PIC and other regulatory proteins that may control tissue specificity. These findings suggest that the $CT_{(11)}$-microsatellite has a key role in the basal expression of *EgrCesA8* and should be further investigated to fully elucidate the regulatory mechanisms underpinning these results (Figure 4.6).

The interaction between *Egr*MYB31 and *EgrCesA8* promoter was expected since interactions of the respective *Arabidopsis* (AtMYB46) and *Populus* (PtMYB21) orthologs have been observed in previous studies (Winzell et al. 2010; Zhong and Ye 2012; Kim et al. 2013). Zhong et al. (2010) suggested that the master regulators *Egr*WND1, *Ptr*WND1 and *At*SND1 were functionally conserved across the different species and the interaction between *Egr*MYB31 (*At*MYB46 homolog) and the *EgrCesA8* promoter in this study confirms that at least parts of the secondary cell wall (SCW) regulatory network is conserved across distantly related plant species. We also identified several SCW MYB (SCWM) binding sites in the *EgrCesA8* 5'UTR which were highly conserved in the *CesA8* 5'UTRs of all 13 *Eucalyptus* species and could be the binding sites for *EgrMYB31* (Chapter 4). We identified a potentially novel interaction between the *Egr*ZincFinger-A transcription factor (*At*C3H14 homolog) and the *EgrCesA8* 5'UTR and this is the first direct target identified for this transcription factor or any of its homologs.

5.2 Identification of a *cis*-regulatory module (CRM) in the *EgrCesA8* promoter

Generally, transcription factors do not interact with a promoter in isolation, and normally function as part of a larger protein complex (Lelli et al. 2012). *Cis*-elements facilitate the interaction of large regulatory protein complexes with the promoter by clustering together in *cis*-regulatory modules (CRM; Levine and Davidson 2005). In Chapter 2 we found that many of the promoters had distinct regions of sequence conservation and in some cases these overlapped with the occurrence of putative *cis*-elements. In the *EgrCesA8* promoter, a strong decrease in species-level nucleotide diversity was observed in the broad region from -900 to -600 bp upstream of the translational start site (Figure 2.2), however, very few *cis*-elements mapped to this area. The lack of *cis*-element data in this region is likely due to the limited number of experimentally validated SCW-associated *cis*-elements at the time of the analysis. For Chapter 3 the *cis*-element catalogue was extended to include newly published and experimentally verified *cis*-elements. These were also mapped to the *EgrCesA8* promoter and a cluster of SCW-associated *cis*-elements was identified in the same -900 to -600 bp region suggesting this conserved region could be an important CRM for *EgrCesA8* regulation (Figure 3.1).

Qualitative GUS assays performed on various overlapping fragments of the *EgrCesA8* -900 to -600 putative CRM revealed that this region plays a role in tissue specificity. The presence of the CRM in a promoter fragment lead to an expression pattern similar to that of the full-length *EgrCesA8* promoter, suggesting the CRM forms part of the minimal promoter for this gene. Partial deletion of the CRM leads to loss of phloem-specific GUS expression (Lu et al. 2008) particularly in the leaf vasculature (Chapter 3). The -900 to -600 bp CRM was also found to be directly bound by several SCW-associated transcription factors including *Egr*NAC170, *Egr*KNAT7, *Egr*MYB80 and *Egr*MYB87 (Chapter 4). *Egr*NAC170 is a close homolog of *At*SND2 which has been implicated in *CesA* regulation by

transactivation and microarray studies (Zhong et al. 2008; Hussey et al. 2011). This is the first study to show a direct interaction for this transcription factor and any plant *CesA* promoter. We were also able to identify the position on the promoter where the interaction likely occurs and observed several conserved putative NAC-like *cis*-elements at this position that may be facilitating this interaction (Figure 4.5). The other transcription factors shown to interact with the *EgrCesA8* CRM (*Egr*KNAT7, *Egr*MYB80 and *Egr*MYB87) could not interact individually with any of the *EgrCesA8* bait constructs. However, in a dual Y1H screen we observed interaction with the bait fragment which contained the *EgrCesA8* CRM and the prey vector *Egr*KNAT7 together with either *Egr*MYB80 or *Egr*MYB87 (Chapter 4). This finding suggests that these proteins act as a complex to regulate the *EgrCesA8* gene and is one of the first putative protein complexes identified for the regulation of a SCW *CesA* gene. There is little or no information on the role of *Egr*MYB80 and *Egr*MYB87 or their homologs in SCW formation. One study suggested a repressive role for the *Arabidopsis* homolog (*At*MYB52) in drought response (Park et al. 2011) and it is possible that the *Egr*KNAT7-*Egr*MYB80/87 complex could play a similar role in *Eucalyptus*, but this remains to be investigated.

5.3 Future prospects and research topics

The research presented here can be extended into several different areas and topics. The most obvious extension of this research would be to confirm the protein-DNA interactions observed here with electrophoretic mobility shift assay (EMSA), protein pull-down assays or bimolecular fluorescence complementation (BiFC) and then to further truncate the promoter fragments until binding is lost to identify the exact sequence required for interaction. It would be particularly interesting to apply these techniques to the promoter regions bound by the

*Egr*ZincFinger-A and *Egr*KNAT7 TFs as there is currently no information on these binding sites in any plant species. Identifying these elements could lead to the detection of other target genes which could help elucidate the role of the poorly characterised transcription factors in the SCW regulatory network. Similarly, EMSA could be used to test for binding of the conserved putative SNBE and SCWM binding sites (identified in Chapter 4) by the different MYB and NAC proteins which were demonstrated to interact with the *EgrCesA8* promoter.

There were several regions of the *EgrCesA8* promoter which did not interact with any of the SCW-associated transcription factors tested in this study. This is most likely due to the limited number of transcription factors screened in the Y1H assay. One limitation of Y1H is that some transcription factors (e.g. VND7 and MYB83) may be lethal to yeast cells when expressed at high levels. It would be useful to extend our current transcription factor panel by cloning more of the SCW associated transcription factors from *Eucalyptus* and expanding our array of techniques for protein-DNA interaction analysis (e.g. ChIP, EMSA, DNA pull down or reporter assays). An alternative would be to screen a *Eucalyptus* or *Arabidopsis* cDNA expression library with the *EgrCesA8* promoter regions to identify novel transcription factors that may be involved in *CesA* regulation. This kind of strategy could help elucidate the function of the other novel GUS regulatory modules identified in our study.

Information on the protein-protein interactions of the SCW network is currently lagging behind that of transcription factor identification. Only the *At*KNAT7-*At*MYB75, *At*VND7-*At*VNI2 and now the putative *Egr*KNAT7-*Egr*MYB80/87 complexes have been identified in the SCW regulatory network (Bhargava et al. 2010; Yamaguchi et al. 2010; Chapter 4). Using the current *Eucalyptus* transcription factor panel in a yeast-2-hybrid (Y2H) screen could reveal other protein-protein interactions which have yet to be explored in the SCW regulatory network. A preliminary Y2H screen is currently underway to confirm the

putative *Egr*KNAT7-*Egr*MYB80/87 complex identified in our study. A Y2H screen of a cDNA expression library would also be effective in identifying other co-factors and transcription factor complexes which may be functioning in the SCW regulatory network.

The advancements in next generation sequencing have increased the power of *in silico* *cis*-element detection. The recent completion of the *Eucalyptus* genome and the wealth of transcriptome data make *Eucalyptus* a good candidate for genome-wide *cis*-element identification studies. Using our previous (Creux et al. 2008) and current (Chapter 2 and 3) studies as a guide, one could perform a genome-wide analysis of the over-represented sequence elements in co-regulated gene sets such as those involved in cellulose biosynthesis. Studies such as these will increase the number of *Eucalyptus*-specific promoter elements and could further elucidate the regulatory mechanisms underlying SCW biosynthesis. The *cis*-element analysis performed in Chapter 2 provided us with the first glimpse into the evolution of *Eucalyptus* promoters. The *Eucalyptus* promoter database compiled in Chapter 2 was a useful tool for identifying highly conserved sequence elements in promoter regions that modulated GUS expression and bound key SCW-associated transcription factors (Figure 4.6). The analysis in Chapter 2 should be extended to incorporate more alleles for each promoter from each species as this would give a more accurate representation of the *CesA* promoter diversity within these populations. As more *Eucalyptus* genomes become available it may also be feasible to investigate the evolution and architecture of other SCW-associate gene promoters in this important cellulose producing crop.

A number of quantitative trait loci (QTL) and expression QTL (eQTL) studies have been performed in different *Eucalyptus* populations for traits including growth, wood density and lignin content (Myburg et al. 2003; Thumma et al. 2010; Kullan et al. 2012). It would be interesting to investigate whether any of the non-coding sequence elements identified in this study are associated with QTLs or eQTLs linked to wood properties or cellulose biosynthesis.

Du et al. (2013) showed that several non-coding microsatellites in the *CesA* promoters of *Populus* were associated with traits such as holocellulose content and microfibril angle. This suggests that sequence variations in the *EgrCesA8* CRM or $CT_{(11)}$-microsatellite might have an effect on the cellulose content (by affecting gene regulation) which could be valuable for tree breeding programmes and other biotechnological applications.

Several plant species (including *Eucalyptus*) have proven recalcitrant to standard transformation techniques, and so testing promoters and monitoring reporter gene expression in these plants have been difficult. Forest trees have the added difficulties of long generation times and large space requirements, which makes transgenic studies in these organisms a complex issue. In Appendix A we present the optimization of induced somatic sector analysis (ISSA), for promoter::reporter gene analysis in woody stem tissues, using the six full-length *CesA* promoters as a case study. The ISSA results for the six full-length *Eucalyptus CesA* promoters and the promoter deletion constructs (Chapter 3) showed that there were large differences in GUS expression depending on whether *Populus* or *Eucalyptus* woody tissues were transformed. Similarly, the *EgrCesA8* 5'UTR construct could express GUS in *Eucalyptus* stems, but not in *Arabidopsis* plants. These findings suggest that, while the SCW regulatory network may be largely conserved, there could be some species-specific regulation of the *CesA* genes. These results could indicate that extrapolating results from model plant species to crop species may be problematic in some cases. ISSA could be a useful tool for fast confirmation of promoter function in woody species such as *Populus* or *Eucalyptus* before labour-intensive transformation procedures are initiated.

A second finding from our study which, has great implications for biotechnology, is the possible overlap of the SCW regulatory network with networks involved in stress response such as drought and pathogen response. It is important to identify which transcription factors or their paralogs are involved in the different processes so that one does

not unwittingly influence non-target pathways and negatively affect plant fitness. A first step could be to investigate the *EgrCesA8* promoter under different stress conditions by growing the *EgrCesA8* promoter truncate::GUS transgenic plants produced in this study under different stress conditions. Monitoring GUS expression of these plants would enable one to identify other regulatory modules involved in *CesA* regulation under different conditions. These GUS regulatory modules could then be further investigated to identify which transcription factors from the SCW regulatory network and the stress response networks bind to these regions. *Egr*MYB31 may be a candidate for these stress response studies because the *Arabidopsis* homolog (*At*MYB46), while described as a SCW master regulator, has also been implicated in response to pathogen infection (Ramírez et al. 2011). *Arabidopsis* possesses a functional ortholog of *At*MYB46 (*At*MYB83) and it is possible that under pathogen infection *At*MYB83 could play a central role in SCW regulation while AtMYB46 regulates pathogen response. These kinds of studies could help to tease apart the different regulatory networks and mechanisms governing *CesA* expression in response to the environment.

5.4 Final Remarks

This is one of the first studies to directly link regions of a *CesA* promoter to different transcription factors which modulate *CesA* expression. We showed that there is some conservation of the SCW regulatory network between *Eucalyptus* and *Arabidopsis*, and that *Egr*MYB31 (*At*MYB46 homolog) also binds the *CesA* promoter in *Eucalyptus*. This study has identified novel interactions between the *EgrCesA8* promoter and some poorly characterised transcription factors of the SCW regulatory network. For example we show a direct interaction between *Egr*ZincFinger-A (*At*C3H14 homolog) and the promoter of a cell wall biosynthetic gene. These results suggest direct interaction of *Egr*KNAT7 (*At*KNAT7

homolog), *Egr*MYB80 (*At*MYB52/54 homolog) and *Egr*MYB87 (*At*MYB52/54 homolog) with any of the SCW biosynthetic gene promoters. We have also discovered that these proteins potentially form a regulatory complex which interacts with the *CesA* promoter. The results of this study have added to the overall connectivity of the currently known SCW regulatory network and we have begun to reconstruct the *Eucalyptus* SCW regulatory network (Figure 5.1). Further studies on this important regulatory network in model and non-model species will greatly aid in elucidating the regulatory mechanisms behind secondary cell wall formation and cellulose deposition.

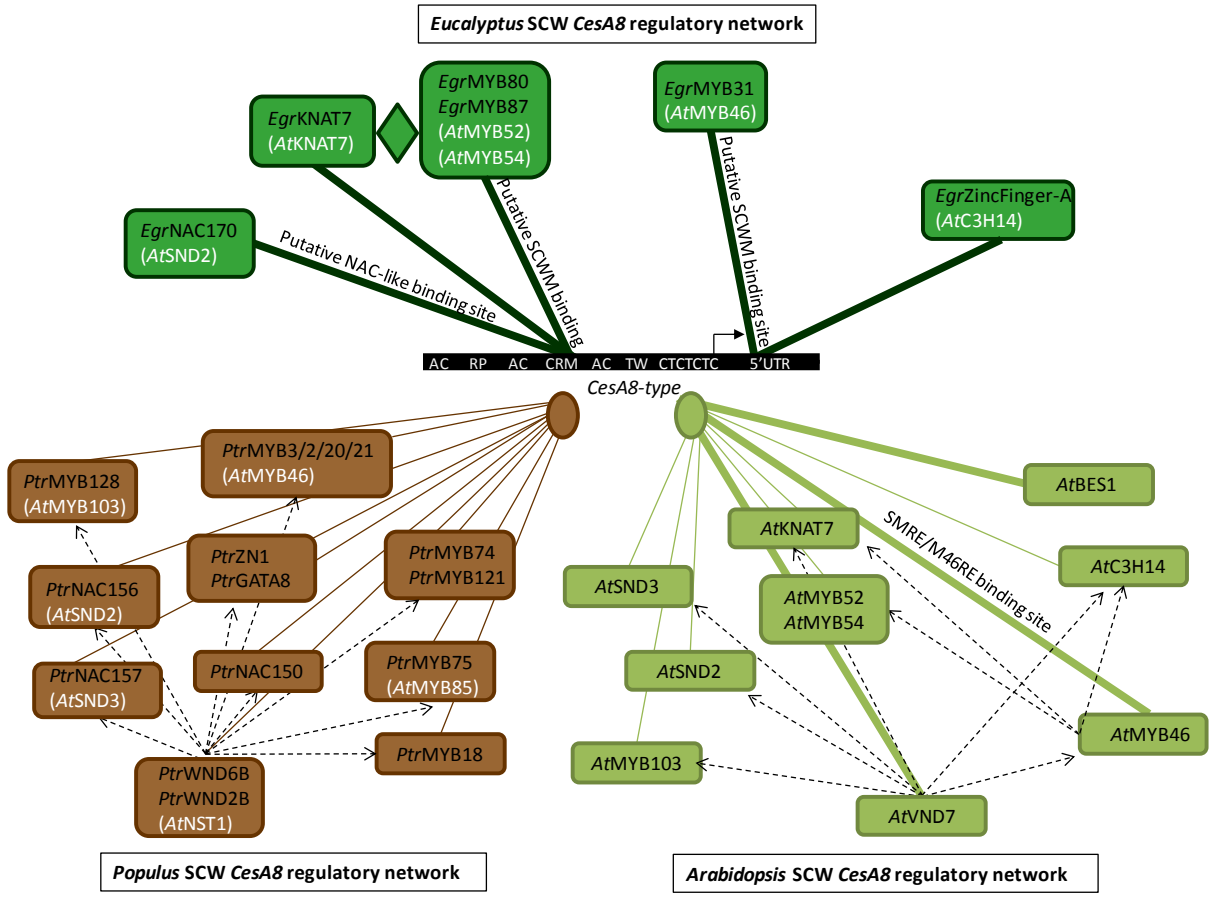


Figure 5.1 Networks of secondary cell wall (SCW)-associated transcription factors which interact with the *CesA8* promoter in *Eucalyptus* (dark green), *Populus* (brown) and *Arabidopsis* (light green). The horizontal black bar represents the *CesA8* promoter with known and novel regulatory modules indicated in white (AC-activator, RP-repressor, CRM-*cis*-regulatory module, TW-tension wood, $CT_{(11)}$-microsatellite, 5'UTR). Boxes represent the different transcription factors which

have been shown to directly interact (thick lines) or indirectly activate (thin lines) the *CesA8* promoter. White bracketed gene name in the *Populus* and *Eucalyptus* boxes indicate *Arabidopsis* homolog. Known and putative binding sites are indicated on the line representing interactions (SMRE, SCWM, and NAC binding sites). The dark green diamond indicates a putative protein complex regulating *CesA8* gene expression. All *Arabidopsis* and *Populus CesA8*-protein interactions lead to an oval indicating a lack of positional data for the protein-DNA interactions of these transcription factors and the *CesA8* promoter.

Several important features of the *EgrCesA8* promoter were also identified, including the $CT_{(11)}$-microsatellite and tissue-specific CRM which play a key role in *EgrCesA8* expression. These sequence elements could be used to investigate other *Eucalyptus* promoters and perhaps identify common features of the SCW-related cellulose biosynthetic gene promoters. These regulatory features could be used to identify cellulose-associated markers for breeding programmes or for other biotechnological applications such as constructing synthetic promoters. This is one of the most comprehensive studies of the *Eucalyptus* SCW-regulatory network, and we have substantially increased the number of known protein-DNA and putative protein-protein interactions in the *Eucalyptus* SCW regulatory network.

5.5 References

Bernard V, Brunaud V, Lecharny A (2010) TC-motifs at the TATA-box expected position in plant genes: a novel class of motifs involved in the transcription regulation. BMC Genomics 11: 166

Bhargava A, Mansfield SD, Hall HC, Douglas CJ, Ellis BE (2010) MYB75 functions in regulation of secondary cell wall formation in the *Arabidopsis* inflorescence stem. Plant Physiol. 154: 1428-1438

Creux NM, Ranik M, Berger DK, Myburg AA (2008) Comparative analysis of orthologous cellulose synthase promoters from *Arabidopsis*, *Populus* and *Eucalyptus*: evidence of conserved regulatory elements in angiosperms. New Phytol. 179: 722-737

Du Q, Pan W, Xu B, Li B, Zhang D (2013) Polymorphic simple sequence repeat (SSR) loci within cellulose synthase (*PtoCesA*) genes are associated with growth and wood properties in *Populus tomentosa*. New Phytol. 197: 763-776

Hussey SG, Mizrachi E, Spokevicius AV, Bossinger G, Berger DK, Myburg AA (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 Biol. 11: 173

Kim W-C, Ko J-H, Kim J-Y, Kim J, Bae H-J, Han K-H (2013) MYB46 directly regulates the gene expression of secondary wall-associated cellulose synthases in *Arabidopsis*. Plant J. 73: 26-36

Kullan ARK, Van Dyk MM, Hefer CA, Jones N, Kanzler A, Myburg AA (2012) Genetic dissection of growth, wood basic density and gene expression in interspecific backcrosses of *Eucalyptus grandis* and *E. urophylla*. BMC Genetics 13: 60

Lenhard B, Sandelin A, Carninci P (2012) Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nature Rev. Genet. 13: 233-245

Levine M, Davidson EH (2005) Gene regulatory networks for development. Proc. Natl. Acad. Sci.USA 102: 4936-4942

Myburg A, Griffin A, Sederoff R, Whetten R (2003) Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F1 hybrid based on a double pseudo-backcross mapping approach. Theor. Appl. Genet. 107: 1028-1042

Park MY, Kang J-y, Kim SY (2011) Overexpression of *At*MYB52 confers ABA hypersensitivity and drought tolerance. Mol. Cells 31: 447-454

Ramírez V, Agorio A, Coego A, García-Andrade J, Hernández MJ, Balaguer B, Ouwerkerk PB, Zarra I, Vera P (2011) MYB46 modulates disease susceptibility to *Botrytis cinerea* in *Arabidopsis*. Plant Physiol. 155: 1920-1935

Thumma BR, Baltunis BS, Bell JC, Emebiri LC, Moran GF, Southerton SG (2010) Quantitative trait locus (QTL) analysis of growth and vegetative propagation traits in *Eucalyptus nitens* full-sib families. Tree Genet. Genomes 6: 877-889

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

Yamaguchi M, Ohtani M, Mitsuda N, Kubo M, Ohme-Takagi M, Fukuda H, Demura T (2010) VND-INTERACTING2, a NAC domain transcription factor, negatively regulates xylem vessel formation in *Arabidopsis*. Plant Cell 22: 1249-1263

Yamamoto YY, Ichida H, Matsui M, Obokata J, Sakurai T, Satou M, Seki M, Shinozaki K, Abe T (2007) Identification of plant promoter constituents by analysis of local distribution of short sequences. BMC Genomics 8:67

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

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

Appendix A

Induced somatic sector analysis of cellulose synthase (*CesA*) promoter regions in woody stem tissues

N.M. Creux[1,2], A. Spokevicius[3], G. Bossinger[3], A.A. Myburg[1,2]

Department of Genetics[1], Faculty of Natural and Agricultural Sciences, Forestry Argicultural Biotechnology Institute (FABI)[2], University of Pretoria, Pretoria, 2000, South Africa

Department of Forest and Ecosystem Science[3], Melbourne School of Land and Environment, The University of Melbourne, Creswick, VIC, Australia

This appendix presents the optimization of Induced Somatic Sector Analysis (ISSA) for analysis of promoter::GUS constructs in the woody tissues of *Eucalyptus* and *Populus*. The work was performed in collaboration with the University of Melbourne and was published in the *Planta* (2013) 237:799-812. The article is represented here as it was published and for this reason it contains the original *Eucalyptus CesA* names. For the new names used throughout the dissertation please refer to Table 1.1.

Planta
DOI 10.1007/s00425-012-1792-x

ORIGINAL ARTICLE

Induced somatic sector analysis of cellulose synthase (*CesA*) promoter regions in woody stem tissues

Nicky M. Creux · Gerd Bossinger · Alexander A. Myburg · Antanas V. Spokevicius

Received: 2 July 2012 / Accepted: 17 October 2012

Abstract The increasing focus on plantation forestry as a renewable source of cellulosic biomass has emphasized the need for tools to study the unique biology of woody genera such as *Eucalyptus*, *Populus* and *Pinus*. The domestication of these woody crops is hampered by long generation times, and breeders are now looking to molecular approaches such as marker-assisted breeding and genetic modification to accelerate tree improvement. Much of what is known about genes involved in the growth and development of plants has come from studies of herbaceous models such as *Arabidopsis* and rice. However, transferring this information to woody plants often proves difficult, especially for genes expressed in woody stems. Here we report the use of induced somatic sector analysis (ISSA) for characterization of promoter expression patterns directly in the stems of *Populus* and *Eucalyptus* trees. As a case study, we used previously characterized primary and secondary cell wall-related cellulose synthase (*CesA*) promoters cloned from *Eucalyptus grandis*. We show that ISSA can be used to elucidate the phloem and xylem expression patterns of the *CesA* genes in *Eucalyptus* and *Populus* stems and also show that the staining patterns differ in *Eucalyptus* and *Populus* stems. These findings show that ISSA is an efficient approach to investigate promoter function in the developmental context of woody plant tissues and raise questions about the suitability of heterologous promoters for genetic manipulation in plant species.

Keywords CAMV35S promoter · *Eucalyptus* · GUS reporter gene · *Populus* · Secondary cell wall · Wood formation

Abbreviations

ATS^{-2}	Average of transformed sectors per cm^2
ISSA	Induced somatic sector analysis

Introduction

Plantation tree species such as those from the genera *Populus* and *Eucalyptus* are receiving worldwide attention for their capacity to produce cellulosic biomass which can be used for pulp and, potentially, biofuel production (Hinchee et al. 2011). Unlike first generation biofuel crops such as sugarcane and maize, forest trees are less likely to directly compete with food production and have a greater biomass production capacity (Rathmann et al. 2010), although the processing of lignin-rich woody biomass to liberate cell wall biopolymers remains a challenge (Mansfield 2009). Furthermore, the genetic improvement of forest trees is hindered by long generation times and late expression of mature traits. Tree breeders attempting to enhance properties such as wood quality and cellulose deposition will benefit from the application of molecular approaches such as marker-assisted breeding (MAB) and genetic modification (Grattapaglia et al. 2009; Seguin

N. M. Creux · A. A. Myburg
Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria 0028, South Africa

A. A. Myburg
e-mail: zander.myburg@fabi.up.ac.za

G. Bossinger (✉) · A. V. Spokevicius
Department of Forest and Ecosystem Science, Melbourne School of Land and Environment, The University of Melbourne, Creswick, VIC, Australia
e-mail: gerd@unimelb.edu.au

A. V. Spokevicius
e-mail: avjs@unimelb.edu.au

Published online: 07 November 2012

2011). These molecular approaches are now also benefiting from the application of next-generation genomics technologies, which can be used to study the genetics of wood formation as a system and to rapidly identify candidate genes for further functional analysis (Mizrachi et al. 2012).

Cellulose is deposited in plant cell walls by large, membrane-bound, protein complexes composed of several different cellulose synthase (CESA) proteins (Kimura et al. 1999), the identity of which depends on the type of cell wall being laid down (Song et al. 2010). In *Arabidopsis* and other plant species, three *CesA* genes (*CesA4*, *7* and *8*) have been associated with secondary cell wall deposition, while a different set of *CesA* genes were found to be involved in primary cell wall formation (Turner and Somerville 1997; Taylor et al. 2000, 2003; Hamann et al. 2004; Samuga and Joshi 2004; Ranik and Myburg 2006). During primary cell wall formation in *Arabidopsis* two *CesA* genes, *AtCesA1* and *3*, are essential for cell development with knock-out mutants being lethal (Arioli et al. 1998; Scheible et al. 2001). Five other *CesA* genes (*AtCesA2*, *4*, *5*, *6* and *9*) have been linked to primary cell wall formation in *Arabidopsis*, but these are functionally redundant when mutated and appear to be involved in tissue-specific primary cell wall formation (Beeckman et al. 2002; Desprez et al. 2002; Stork et al. 2010; Carroll and Specht 2011).

While there are many similarities in cellulose biosynthesis across plant genera (Popper et al. 2011), there are also a number of species-specific features. The *CesA* gene family has ten members in *Arabidopsis* (Richmond and Somerville 2000), while *Populus* has 18 expressed *CesA* genes (Djerbi et al. 2005; Suzuki et al. 2006; Kumar et al. 2009). A phylogenetic analysis of the *Populus CesA* gene family revealed that the 18 *CesA* genes grouped with the ten *Arabidopsis* orthologs in all of the primary and secondary cell wall-related clades and that *Populus* has two or more paralogs of some *Arabidopsis* genes (Kumar et al. 2009). In particular, it was noted that *Populus* has duplicated genes for the secondary cell wall-associated *AtCesA7* and *AtCesA8* genes. In each case, one of the two *Populus* paralogs (*PtiCesA7*-A or *PtiCesA8*-B) was more highly expressed in xylem, suggesting differential regulation of the paralogs and possible loss of regulation of the lower expressed paralog (Suzuki et al. 2006). Similarly, the primary cell wall-associated *AtCesA3* gene has four close orthologs in *Populus*, and each of these has a different expression pattern (Suzuki et al. 2006). The differentiated expression patterns of the duplicated *CesA* genes in *Populus* suggest that the *Populus* paralogs may be undergoing subfunctionalization.

Inter-specific differentiation can affect regulatory sequences in promoters and produce discordant results when different orthologs are used in transgene constructs. Fei et al. (2006) found that a promoter construct which increased glutamine synthase expression in both *Lotus japonicus* and *Sesbania rostrata*, did not produce a corresponding increase in expression in *Pisum sativum.* In an extensive study on mammalian and *Drosophila* cell lines eight supposedly constitutive promoters were tested and most promoters showed variation in reporter gene expression between both cell line and species (Qin et al. 2010). Even the highly utilized CAMV 35S promoter has been shown to have differential expression across different species (Benfey and Chua 1990; Zhang et al. 2003). This evidence suggests that for some promoters it may be desirable to perform functional analysis in the native genetic backgrounds; however, this may not be practical in species that are recalcitrant to genetic transformation.

Induced somatic sector analysis (ISSA), first proposed by Spokevicius et al. (2005) and developed further by Van Beveren et al. (2006), uses a novel *in planta* transformation method, which has been successfully applied in the analysis of transgenes in woody stem tissues of *Pinus*, *Populus* and *Eucalyptus* (Hussey et al. 2011). In this method, *Agrobacterium* carrying the promoter and transgene of interest is applied to the exposed cambium on the stem of a living tree. The gene construct is transferred by *Agrobacterium* into actively dividing cambial, xylem, phloem and ray initial cells, creating a number of transformants in this small section ($\sim$1 cm^2) of the tree stem (Van Beveren et al. 2006). When the cambium is resealed and the stem is allowed to grow for a few months where the transformed cells divide and multiply within the stem, producing somatic sectors of transformed cells. This area of transformed cells can then be analysed for transgene (e.g. β-glucuronidase) expression and changes in cell wall morphology by comparing transformed sectors with adjacent non-transformed stem cells. ISSA has great potential for functional genetic studies, as it allows for the analysis of transgenes and promoters directly in the stem tissues of the tree and, for wood-specific constructs, may give a more accurate picture of the native functions or expression patterns of transgenes in woody tissues (Spokevicius et al. 2007).

Here, we used ISSA to study the expression patterns of six previously characterized promoters of *Eucalyptus grandis CesA* genes (Creux et al. 2008) in woody stem tissues of *Eucalyptus* and *Populus* trees. The first objective of this study was to investigate the suitability of ISSA for the analysis of promoter function in various woody stem tissues. Second, we assessed whether ISSA could be used to compare reporter gene expression patterns in *Populus* and *Eucalyptus* stems. To our knowledge, this is the first study to directly compare the expression patterns of promoter::reporter gene constructs in woody tissues of *Populus*, the model tree genus for molecular studies, and *Eucalyptus*, a globally important fibre crop.

Materials and methods

Plant material

Three-month-old ramets of five *Eucalyptus camaldulensis* × *globulus* and six *Eucalyptus camaldulensis* × *grandis* clones were purchased from a specialist forestry nursery (Narromine Transplants, Narromine, NSW, Australia), potted in premium potting mix and maintained in a greenhouse for another 4 months. A single *Populus alba* (L.) 'pyramidalis' clone, growing at the University of Melbourne Creswick Campus (Vic. Australia), was used to generate plant material through rooted cuttings. Dormant stems were sourced and established in cutting beds following treatment with a commercial rooting hormone powder (Yates Striking Powder, Homebush, NSW), transplanted into premium potting mix after 6 weeks and maintained in the greenhouse for 3 months until required. Greenhouse temperatures were maintained between 14 and 17 °C at night and between 21 and 25 °C during the day. A 16-h photoperiod was kept through supplementary lighting. Supplementary lighting was supplied by six 1000 W Metal Halide globes in a glasshouse chamber of approximately 16 m^2. All plants were watered regularly with tap water (as required, depending on season) and fertilised with a slow release formulation (Osmocote Exact Mini, Scotts-Sierra Horticultural Products, Marysville, OH, USA) every 3 months.

Promoter isolation, vector and *Agrobacterium* preparation

Kumar et al. (2009) proposed a revision of the *CesA* gene nomenclature for the *Populus CesA* genes, which allows for the direct comparison of the *Arabidopsis* and *Populus CesA* genes (See Table 1 for *CesA* gene orthology). The change in nomenclature has not yet been applied to the *Eucalyptus CesAs* and for this reason we have retained the naming convention first published by Ranik and Myburg (2006). The *Eucalyptus CesA* promoter regions (*EgCesA1-5* and *7*) and the *Arabidopsis CesA8* promoter region were cloned into the pCR8/GW/TOPO entry vector (Invitrogen). Orientation of the inserts was determined using restriction endonuclease digestion. Promoter DNA was transferred from the entry vectors to the binary vector pMDC162 (Curtis and Grossniklaus 2003) using LR Clonase (Invitrogen) according to the manufacturer's instructions. The expression cassette consisting of the promoter and GUS (β-glucuronidase) reporter gene was confirmed by sequencing prior to *Agrobacterium* transformation (Creux et al. 2008).

Two CAMV 35S promoter vectors were used as a positive control. The first, 35S-F, was the pCAMBIA1305.1 vector (http://www.cambia.org/ verified 5/5/10) and the second, 35S-G, was based on the same pCAMBIA1305.1 vector backbone, but with a Gateway recombinase cassette in the multi-cloning region. An empty (promoter-less) pMDC162 vector was also used as negative control.

All vectors were transformed into AGL-1, a disarmed strain of *Agrobacterium tumefaciens* containing a derivative of pTiBO542 (Lazo et al. 1991), using an *E. coli* pulsar (BIO-RAD Laboratories, Gladesville, NSW, Australia), 2 mm cuvette and 2.5 kV, and following protocol 26 'transformation of *E. coli* by electroporation' as described in Sambrook and Russell (2001). Bacteria were grown for 48 h at 28 °C in LB medium containing 25 μg mL^{-1} rifampicin and 50 μg mL^{-1} kanamycin. The *Agrobacterium* suspension was then diluted 1:20 with fresh LB and grown to OD_{600} of 0.4–0.6 after which the cells were recovered by centrifugation (1,150*g* for 15 min) and resuspended in 1 mL of Murashige–Skoog (MS) media prior to inoculation (Table 2).

Table 1 *Arabidopsis thaliana* and *Populus trichocarpa CesA* orthologs and the corresponding *Eucalyptus grandis CesA* genes included in this study

	Arabidopsis[a]	*Eucalyptus*[b]	*Populus*[c]
Primary cell wall-associated *CesA* genes	*AtCesA1*	*EgCesA5*	*PtiCesA1*-A
	AtCesA2	*EgCesA7*	*PtiCesA6*-A
	AtCesA3	*EgCesA4*	*PtiCesA3*-D
Secondary cell wall-associated *CesA* genes	*AtCesA4*	*EgCesA2*	*PtiCesA4*
	AtCesA7	*EgCesA3*	*PtiCesA7*-A
	AtCesA8	*EgCesA1*	*PtiCesA8*-A

[a] All *Arabidopsis* ortholog information was obtained from TAIR (http://www.arabidopsis.org)

[b] All *Eucalyptus* ortholog information was obtained from Phytozome (http://www.phytozome.org) and nomenclature for the *EgCesA* genes taken from previous publications (Ranik and Myburg 2006; Yin et al. 2009)

[c] All *Populus* ortholog information was obtained from Kumar et al. (2009) in which the *Populus* and *Arabidopsis* naming conventions were unified

Inoculation and harvest of transformed tissues for Induced Somatic Sector Analysis

During the start of the growing season (early summer) 40 *P. alba* '*pyramidalis*', 20 *E. camaldulensis* × *globulus* (four ramets of each clone) and 24 *E. camaldulensis* × *grandis* (four ramets of each clone) potted plants were selected on the basis of good form and growth for experimentation (Table 2). Along the stem of each tree, 11 approximately 1 cm^2 cambial windows were opened using the in vivo stem ISSA method described in Van Beveren et al. (2006). Each cambial window was inoculated with 5 µL of *Agrobacterium* suspension containing one of the 11 promoter constructs under investigation (one promoter per window) and subsequently sealed using parafilm. Due to lower sector numbers for the *EgCesA1* and *EgCesA2* promoters (see "Results") a further 30 windows were produced for each promoter in both species (*Eucalyptus* and *Populus*) during the following year using the same clonal material. Plant height and stem diameter (measured at a height of 10 cm from the trunk base) were recorded. Plants were fertilised after inoculation and maintained in the greenhouse until harvest.

At harvest, plant height and stem diameter (measured at a height of 10 cm from the trunk base) were again recorded and stem sections harbouring cambial windows were excised from the main stem and placed in 10 ml Falcon tubes for transport. Un-inoculated stem tissue (outside the window area) was removed and the remaining cambial tissue was cut transversely into 1 mm half-discs and placed back into the 10 mL tubes for GUS assays. Cambial discs were washed twice with 0.1 M $NaPO_4$ buffer (pH 7) prior to the addition of 5 mL (approx) of GUS solution (0.1 M $NaPO_4$ buffer pH 7, 0.5 % v/v Triton X-100 (Sigma), 10 mM EDTA (Sigma), 0.5 mM potassium ferricyanide (III) (Sigma), 0.5 mM potassium hexacyanoferrate (II) trihydrate (Sigma), 0.5 mM X-Gluc). Cambial discs were incubated in a water bath in GUS solution at 55 °C for 10 min prior to being placed upright in the dark on a rotary shaker (150 rpm) at 37 °C overnight. The GUS solution was then replaced with 70 % ethanol and samples were stored at 4 °C until assessment (Spokevicius et al. 2005; Alwen et al. 1992).

Assessment of GUS staining patterns

For the purpose of promoter expression analysis, it is important to note that the observed "sector types" described in this study in all cases represent the net staining pattern produced by the combined effect of the initial cell transformed (determining the total sector of transformed cells) and the cell/tissue specificity of the promoter construct tested (specifying the subpopulation of cells within the sector that express GUS). Cambial windows were initially assessed for GUS staining using protocols described in previous ISSA studies where whole sectors were investigated (Spokevicius et al. 2005; Van Beveren et al. 2006). In most cases, the sector was transversely cut into two and each half was examined under a microscope. Generally, correct identification of sector type and staining pattern could be made without any further examination due to a combination of the intensity of the GUS staining, the transparency of the wood, sector size and the experience of investigators. Where identification was not certain upon initial investigation, serial sectioning was undertaken on the two halves until correct identification could be made. The sector categories described in the previous studies (Spokevicius et al. 2005; Van Beveren et al. 2006) were also used in this study in addition to new sector categories. The same 'tylose' and 'wound parenchyma' sector categories were used as before, but 'cambial' and 'phloem' sectors were redefined to include the addition of two new sector pattern types, indicative of the cell type that was initially transformed and the subsequent expression pattern of the promoter tested. In the case of 'cambial' sectors, these were reclassified into two subcategories. The first

Table 2 Growth parameters and overall transformation efficiency for *Eucalyptus* and *Populus* plants

Growth parameters	*Populus*	*Eucalyptus*
Average height at inoculation	91.7 cm (SE = 5.0 cm)	168.9 cm (SE = 6.7 cm)
Average height at harvest	205.8 cm (SE = 6.9 cm)	247.7 cm (SE = 9.186 cm)
Average diameter at inoculation (at stem height = 10 cm)	6.2 mm (SE = 0.09 mm)	7.5 mm (SE = 0.17 mm)
Average diameter at harvest (at stem height = 10 cm)	11.4 mm (SE = 0.22 mm)	11.2 mm (SE = 0.33 mm)
Average total radial growth of cambial window xylogenic tissue (from wound site)	2.55 mm (SE = 0.03 mm)	1.67 mm (SE = 0.069 mm)
Average total radial expansion rate	0.022 mm/day (SE = 0.001 mm/day)	0.013 mm/day (SE = 0.001 mm/day)
Total number of sectors counted	897	1,661
[a]ATS^{-2}	3.7	6.4

[a] ATS^{-2} is the average number of transformed sectors per cm^2 of inoculated stem tissue

subcategory, 'xylem mother sectors' (Fig. 1a), was characterized by GUS staining in the newly derived xylem cells extending for a short distance from the wound parenchyma. 'Xylem mother sectors' are indicative of the initial transformation of a xylem mother cell and the formation of a transgenic xylem sector which terminates after a number of cell divisions when the xylem mother cell is replaced by an adjacent, non-transgenic mother cell, or when it terminally differentiates and therefore ceases to be a mother cell. A second subcategory was defined as 'cambial sector proper' (Fig. 1a), where GUS staining was observed in and around the cambial region, which indicates that an undifferentiated cambial initial was transformed which continued to produce xylem and phloem mother cells and sectors of xylem and phloem cells derived from these mother cells. The definition of a 'cambial sector proper' had to be expanded for the promoter analysis because the tissue specificity of the promoters investigated could produce GUS staining patterns that differ from the original sector descriptions which were based on constitutive GUS expression (Fig. 1a—original descriptions and Fig. 1c—new cambial staining patterns). While a "cambial sector proper" may produce transgenic xylem and phloem sectors, GUS staining would only be seen in the xylem, for example if a xylem-specific promoter was tested. During this stage of assessment, the amount of new growth (wound parenchyma and xylem tissue) was measured in millimetres using a dissecting microscope to give an indication of the extent of growth that occurred post inoculation.

Cambial sectors were further analysed to gain insight into the temporal and spatial activity of the promoters under investigation, referred to here as 'cambial sector ratio'. For this analysis, cambial sectors were assessed for presence or absence of GUS staining in three distinct regions defined as

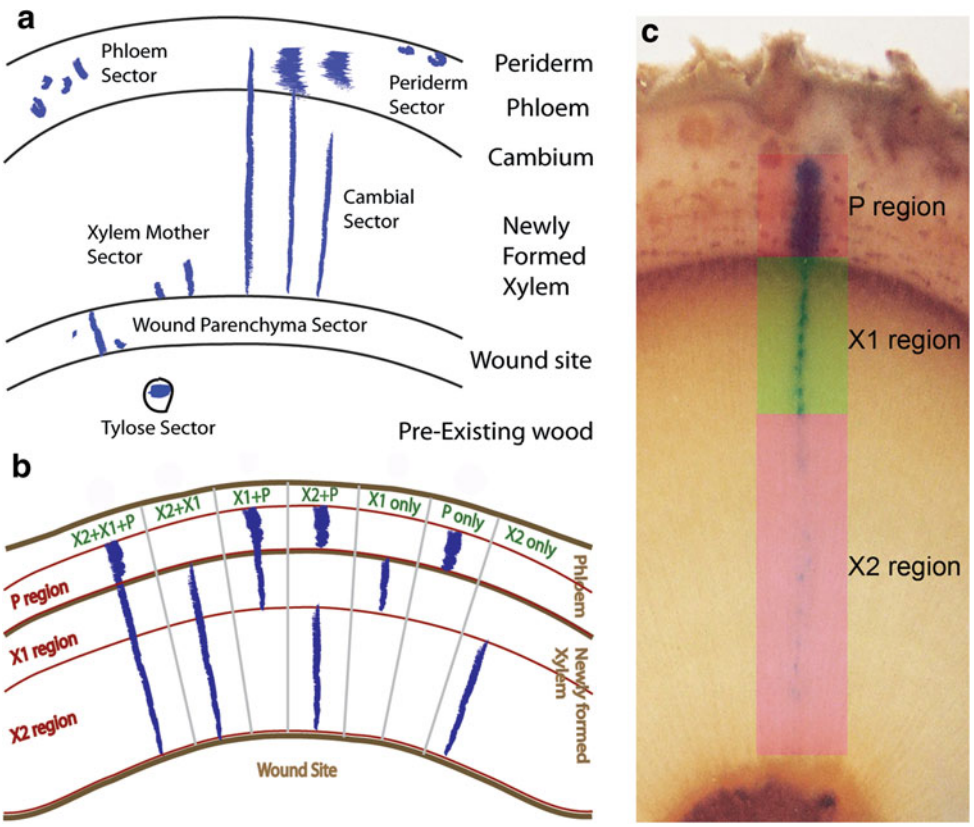


Fig. 1 Representative somatic sectors (expected and observed) for different promoter types in woody stem tissues. **a** Schematic representation of the different somatic sectors that are typically observed in cross sections of transformed woody stem tissue during induced somatic sector analysis (ISSA) with constitutive CAMV35S driven GUS expression. Only tylose sectors are observed in the pre-existing xylem at the centre of the stem. All other sector types are observed in the wound site and across the newly formed cambial zone. **b** Reclassification of the different sector types (staining patterns) that can be formed during ISSA in woody stem tissues depending on the type of cell transformed (cambial initial, phloem mother cell or xylem mother cell) and the cell or tissue specificity of the promoter used to drive GUS expression. Sector types were classified depending on the presence of GUS in the phloem (P), immature xylem (X1) mature xylem (X2) or a combination of these. The only sector types observed during this study were X2 + X1 + P, X1 + X2, X1 + P, X1 only and P only. **c** Cross section of a *Populus* stem showing a transformed cambial sector with GUS expression in the phloem (P), immature xylem (X1) and the mature xylem (X2) driven by the CAMV35S promoter, indicating that a cambial initial was transformed which continued to divide and produce xylem and phloem cells

X1, X2, and P (Fig. 1b). The X2 region extended outward from the initial wound parenchyma cells up to the end of the mature xylem. Staining observed in this region was indicative of promoter activity in ray cells which extend radially through the stem. Most xylem fibres and vessels in the X2 region have already undergone programmed cell death (PCD) and one would therefore only expect X2 GUS expression and staining in ray cells which have not undergone PCD. The X1 region was characterized by staining in developing xylem cells close to and including the cambial zone (but no staining on the phloem side). Staining observed in the X1 region was indicative of promoter activity in differentiating xylem cells (before the onset of PCD). The P region comprised all phloem tissues and GUS staining in this region was indicative of promoter activity in phloem tissue in general. It is important to again note here that the final staining pattern observed (combination of P, X1 and X2) was determined by the cell type initially transformed and the specificity of the promoter tested.

Statistical analysis of ISSA results

Details of the statistical analysis are outlined in the Results section. Confidence intervals (95 %) were calculated for growth data using Minitab (Minitab Inc., State College, PA, USA) to compare growth rates. Chi-squared tests were performed using Minitab to compare the frequency of GUS expressing sectors observed in the X1, X2 and P regions (at $\alpha = 0.05$). Promoter constructs for which fewer than ten sectors were observed were excluded from the statistical analysis, as was the case for the *EgCesA2* and *EgCesA4* in *Populus* stems. However, the majority of sectors observed for these promoters were cambial sectors and data for these promoters were included in the graphs, but should be validated in future experiments.

Results

The number of transformed sectors varied dependent on species and promoter

In total, inoculated stem tissue with a surface area of 559 cm^2 (*Eucalyptus:* 258 cm^2 and *Populus:* 240 cm^2) was harvested and stained for GUS activity, with 2,558 transformed tissue sectors identified for the ten promoter constructs investigated (Table 2). Overall, more GUS expressing sectors were identified in *Eucalyptus* stem tissues with a total of 1,661 and an average of 6.4 transformed sectors per cm^2 of inoculated tissue (ATS^{-2}). In *Populus* stems only 897 sectors were counted, with an ATS^{-2} of 3.7 (Table 2). In the *Populus* and *Eucalyptus* stems, the *35S*::GUS (F and G) constructs resulted in the highest ATS^{-2} values ranging from 12.0 to 27.1 (Table 3), which is a measure of overall transformation efficiency as the 35S promoters are expressed ubiquitously in most plants. No sectors were observed in windows inoculated with the promoter-less pMDC162 vector (negative control). The highest ATS^{-2} value for the *CesA* promoters was observed for the *AtCesA8* promoter construct in *Eucalyptus* ($ATS^{-2} = 9.4$) and *Populus* stems ($ATS^{-2} = 5.2$), while the lowest values were observed for the *EgCesA2* promoter with an ATS^{-2} value of 0.6 for *Eucalyptus* and ATS^{-2} of 0.3 in *Populus* stems (Table 3). It is important to note that the ATS^{-2} values for the *CesA* promoters are more likely to reflect the spatio-temporal regulation of these promoters and the lower values are likely due to the smaller subset of tissues in which the promoters are active.

Sector type and frequency differed between promoter constructs

Overall, the most abundant sector types observed were phloem, xylem mother and cambial sectors, while periderm sectors only occurred at very low frequencies in *Populus* and *Eucalyptus* stems (Fig. 2a, b). As expected, the constitutive 35S promoter constructs (F and G) produced a wider range of sector types than the *CesA* promoter constructs (Fig. 2a, b). All of the major sector types were represented in *Populus* stems inoculated with the two 35S constructs, including periderm and tylose sectors, which are induced upon wounding (Van Beveren et al. 2006). In *Eucalyptus* stems, no periderm or tylose sectors were recorded for the 35S constructs, and the ratio of sector types observed for the 35S constructs was distinctly different to that of the *CesA* promoter constructs (Fig. 2b). In both species, the *CesA* promoter constructs produced a high frequency of cambial and xylem mother sectors (Fig. 2a, b). In *Populus* stems, the *CesA* promoter constructs also produced phloem sector types at high frequencies

Table 3 Average number of transformation events per cm^{-2} of tissue (ATS^{-2}) observed for the different promoter constructs

Promoter	*Populus*[a]	*Eucalyptus*[a]
EgCESA1	0.5 (0.07)	0.7 (0.13)
EgCESA2	0.3 (0.17)	0.6 (0.12)
EgCESA3	1.6 (0.28)	2.0 (0.45)
EgCESA4	0.6 (0.17)	8.1 (1.26)
EgCESA5	1.3 (0.28)	2.7 (0.82)
EgCESA7	3.0 (0.45)	5.3 (1.01)
AtCESA8	5.2 (0.75)	9.4 (1.58)
35SF	19.1 (0.85)	27.1 (2.20)
35SG	12.0 (0.98)	17.8 (2.10)

[a] ATS^{-2} (values in brackets show the standard error for each promoter tested in *Eucalyptus* and *Populus*)

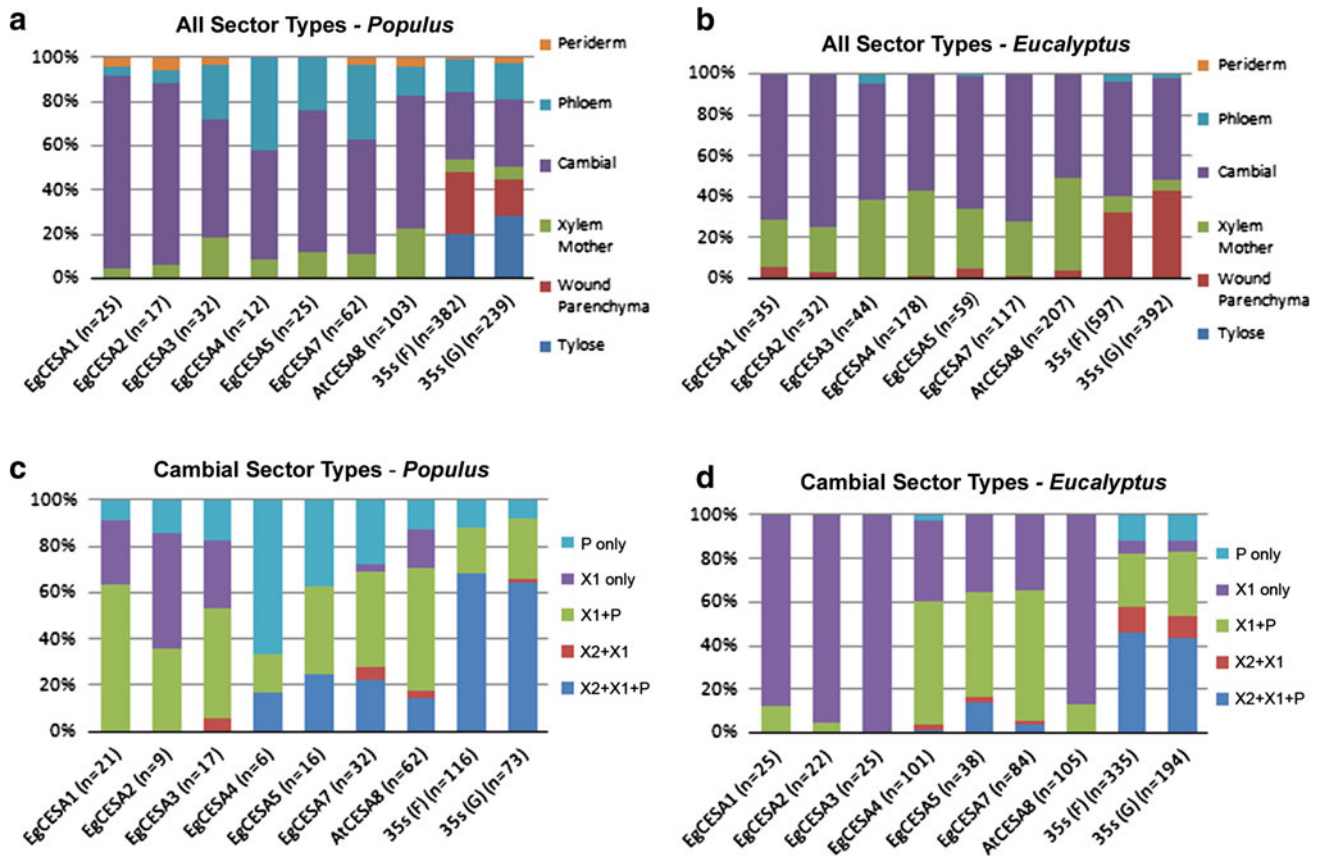


Fig. 2 The overall observed frequency of somatic sector types and cambial sector types observed in *Populus* and *Eucalyptus* stem tissues. The frequency of the different sector types for each promoter in *Populus* (**a** and **c**) and *Eucalyptus* (**b** and **d**) plants is indicated on the *y* axis, while the promoters used for each transformation are indicated on the *x* axis. The number next to each promoter name (*n*) indicates the total number of sectors observed for that promoter. Of all sector types (**a** and **b**) counted, cambial sectors were found to be most highly abundant in *Populus* and *Eucalyptus* stems. The cambial sectors (**c** and **d**) were further classified into different subtypes (Fig. 1c) depending on GUS staining patterns in phloem (P), immature xylem (X1) and mature xylem (X2) regions: P + X1 + X2, X1 + X2, P + X1, X1 only and P only

(15–40 %) in five (*EgCesA3*, *4*, *5*, *7* and *AtCesA8*) of the seven *CesA* promoter constructs investigated (Fig. 2a). This included the secondary cell wall-related *CesA* promoters (*EgCesA3* and *AtCesA8*) for which low numbers of phloem sectors were observed in *Eucalyptus* stems (Fig. 2b).

Cambial sectors were the most abundant and varied among species and promoter constructs

Cambial sectors were highly abundant in *Populus* and *Eucalyptus* stems (Fig. 2a, b) and these sectors were further classified into expression patterns (Fig. 2c, d). Five different cambial expression patterns were observed in the two species which included X2 + X1 + P, X2 + X1, X1 + P, X1-only and P-only (Fig. 1b). The X2 + X1 + P sector type was most likely produced by the transformation of a cambial initial, which subsequently gave rise to a ray sector extending into the P, X1 and X2 region followed by promoter activity in all three regions (Fig. 1c). The X2 + X1 sector type could be the result of transformation of a ray initial on the xylem side, or xylem-specific promoter activity in a cambial sector giving rise to ray cells. The X1 + P sector type was most likely the result of transformation of a cambial initial differentiating into phloem (P) and xylem (X1), but terminating at the zone of PCD (X1/X2 border), and subsequent promoter activity in phloem and xylem cells. X1-only and P-only sectors could be produced by the transformation of a cambial initial followed by xylem or phloem promoter activity, or the transformation of a xylem or phloem mother cell, respectively, followed by promoter activity in the resulting xylem or phloem sector. No X2-only or X2 + P staining patterns were detected in either species.

In the case of the two *35S* promoter constructs (F and G) all cambial sector types (Fig. 1b) were found in *Eucalyptus*

stems (Fig. 2d), but three (X2 + X1 + P, X1 + P and P-only) accounted for the majority of cambial staining patterns observed in *Populus* stems (Fig. 2c). In both species, the X2 + X1 + P sector type was the highest frequency cambial staining pattern observed for the 35S promoter constructs (approx 40 % in *Eucalyptus* and 65 % in *Populus* stems). For the *CesA* promoter constructs, all cambial staining patterns (Fig. 1b) were identified, but with distinct differences in the frequencies of staining patterns between *Populus* and *Eucalyptus* stems.

In *Eucalyptus* stems, cambial sector X1-only was the most frequent type observed for the promoters of secondary cell wall-related genes *EgCesA1*, *2*, *3* and *AtCesA8*, whereas for the promoters of the primary cell wall-related genes *EgCesA 4*, *5* and *7* much higher frequencies of the X1 + P cambial sector types in addition to X1-only were observed (Fig. 2d). In *Populus* stems, the secondary cell wall-related *CesA* promoters also showed a high frequency of X1-only staining patterns; however, in *Populus* these promoters also displayed a higher frequency of the X1 + P type sectors. The primary cell wall-related *CesA* promoters produced a high proportion of X1 + P sectors in the *Populus* stems, similar to the pattern observed in *Eucalyptus*, but there was also a number of X2 + X1 + P and P only sectors present (Fig. 2c).

Some *CesA* promoters showed similar activity to the *35S* promoter

We next investigated whether any of the *CesA* promoter constructs exhibited similar or different cambial sector (staining pattern) ratios when compared with the *35S* promoter constructs (Chi-squared tests, Table 4). Cambial sector ratios were derived from the spatial temporal data sourced from cambial sectors (ratio of X2, X1, P). In *Eucalyptus* stems, all of the *CesA* promoter constructs exhibited significantly different ($\alpha = 0.05$) cambial sector ratios from that of the *35S* promoter constructs (Fig. 3b; Table 4). In *Populus* stems, the cambial sector ratios of some *CesA* promoters, such as *EgCesA4* and *5*, were not significantly different from that of the *35S* promoter constructs (Table 4) and in general the expression pattern seemed to be more variable than in *Eucalyptus* (Fig. 2c). This result suggests that there is a difference in *Populus* and *Eucalyptus* stems, either on a developmental or anatomical level as a result of different cell/tissue patterning,

Table 4 Comparison of β-glucuronidase (GUS) expression frequencies observed for the *CesA* promoter constructs and for the CAMV35S promoter constructs in *Eucalyptus* and *Populus* stem tissues

Chi-square values[a]	*Populus*							
	EgCesA1	*EgCesA2*	*EgCesA3*	*EgCesA4*	*EgCesA5*	*EgCesA7*	*AtCesA8*	35SA[b]
Eucalyptus								
EgCesA1		0.295[e]	1.373[e]	6.169[c]	6.46[c]	6.986[c]	3.863[d]	13.754[c]
EgCesA2	0.708[e]		1.112[e]	4.79[d]	4.518[d]	5.317[d]	2.855[e]	9.186[c]
EgCesA3	2.839[e]	1.11[e]		2.223[e]	2.449[e]	2.941[e]	1.003[e]	7.555[c]
EgCesA4	8.856[c]	11.05[c]	15.151[c]		0.676[e]	0.661[e]	1.463[e]	1.869[e]
EgCesA5	9.976[c]	12.008[c]	16.123[c]	4.852[e]		0.015[e]	1.047[e]	1.591[e]
EgCesA7	9.18[c]	11.363[c]	15.488[c]	0.108[e]	3.275[e]		1.871[e]	3.952[d]
AtCesA8	0.024[e]	1.122[e]	3.258[e]	27.224[c]	27.855[c]	27.32[c]		14.897[c]
35SA	29.165[c]	30.234[c]	37.791[c]	46.858[c]	11.278[c]	36.082[c]	103.367[c]	

Values below the diagonal are pair-wise comparisons within *Eucalyptus* and above the diagonal are within *Populus*

[a] Chi-squared values determined by pair-wise comparison of the frequency of GUS expression for different promoter constructs across the different stem tissues (P: phloem, X1: immature xylem and X2: mature xylem) with null hypothesis of equal expected frequency ratios for each comparison

[b] Only one of the two 35S promoter data sets was used as there was no significant difference between the two datasets

[c] Significance determination: >5.99 shows significant difference with two degrees of freedom (light grey shading)

[d] Significance determination: >3.84 shows significant difference with one degree of freedom (dark grey shading)

[e] White cells indicate promoter comparisons where there was no significant difference in expression patterns

and/or on a genetic level with different transcriptional regulation of the *CesA* promoters.

Patterns of individual *CesA* promoter activity between tree species

We directly compared the cambial staining patterns observed for each promoter construct between the two species. We found that the *35S* promoter staining patterns were not significantly different in *Populus* and *Eucalyptus* tissues irrespective of the vector backbone (Table 5). This confirmed that the *Populus* and *Eucalyptus* stem tissues did not have significantly different relative transformation efficiencies in cells giving rise to X1, X2 and P staining patterns, although the *Populus* stems exhibited lower overall transformation efficiency (Table 3). The similarity of the staining patterns observed for the 35S promoter in *Eucalyptus* and *Populus* stems also suggests that differences in the amount of diffusion of GUS observed in the *Populus* and *Eucalyptus* stems did not greatly influence the resulting sector frequencies. In contrast, all the *CesA* promoter constructs (except the *EgCesA5* promoter) exhibited statistically significant differences ($\alpha = 0.05$) in cambial sector ratios between *Eucalyptus* and *Populus* stems (Table 5). Differences in *EgCesA1*, *2* and *3* and *AtCesA8* promoter activity could be attributed to activity being confined mostly to the X1 (developing xylem) region of *Eucalyptus* (Figs. 3b, 4a, b), whereas in *Populus* stems activity was observed at similar frequencies in both the X1 and P regions (Figs. 3c, 4e, f). In the case of the *EgCesA4* and *7* promoters, the majority of activity was observed in the X1 and P regions in *Eucalyptus* (Fig. 3b), while in *Populus* stems a higher proportion of observations were in the P and X2 (mature xylem) regions (Fig. 3a).

Discussion

ISSA provides a rapid and efficient approach to evaluate promoter expression in woody stems

Gene and promoter testing in tree genera such as *Eucalyptus, Populus* or *Pinus* require time-consuming and laborious manipulation through tissue culture and greenhouse studies. In this study we investigated the use of ISSA (Spokevicius et al. 2005; Van Beveren et al. 2006) as an approach for rapid functional genetic analysis of promoter expression patterns in developing woody tissues based on large numbers of independent transgenic events. We demonstrate the suitability of ISSA for promoter expression analysis of six *Eucalyptus* cellulose synthase (*CesA*) genes in the stems of *Eucalyptus* and *Populus* trees. We show that in the *Eucalyptus* genetic background the *EgCesA* promoters produced distinct staining patterns, which were consistent with the primary and secondary cell wall-associated expression patterns previously demonstrated for these genes (Samuga and Joshi 2004; Ranik and Myburg 2006), whereas in the heterologous *Populus* genetic background the staining patterns of the two groups of *Eucalyptus CesA* genes were less distinct.

There are a number of key advantages to using ISSA for functional genetic analysis of wood formation genes

Table 5 The inter-specific comparison of β-glucuronidase (GUS) expression patterns observed in *Populus* and *Eucalyptus* cambial tissues for the *EgCesA* and CAMV35S promoters

Chi-squared value[a]	*Populus*							
	EgCesA1	***EgCesA2***	***EgCesA3***	***EgCesA4***	***EgCesA5***	***EgCesA7***	***AtCesA8***	**35SA[b]**
Eucalyptus								
EgCesA1	8.584[c]	4.613[d]	9.488[c]	14.438[c]	15.776[c]	20.837[c]	16.503[c]	32.828[c]
EgCesA2	–	7.125[b]	4.613[d]	17.072[c]	17.69[c]	22.588[c]	18.394[c]	33.858[c]
EgCesA3	–	–	7.125[c]	22.917[c]	22.418[c]	27.903[c]	23.5[c]	41.617[c]
EgCesA4	–	–	–	5.462[d]	11.024[c]	16.697[c]	9.004[c]	49.714[c]
EgCesA5	–	–	–	–	2.437[e]	4.369[d]	1.476[e]	13.97[c]
EgCesA7	–	–	–	–	–	12.978[c]	6.421[c]	39.142[c]
AtCesA8	–	–	–	–	–	–	47.554[c]	101.493[c]
CAMV35SA	–	–	–	–	–	–	–	1.568[e]

[a] Chi-squared values determined by comparison of the frequency of GUS expression for different promoter constructs across the different stem tissues (P: phloem, X1: immature xylem and X2: mature xylem) of *Populus* and *Eucalyptus* plants with null hypothesis of equal expected frequency ratios for *Populus* and *Eucalyptus* tissues

[b] Only one of the two 35S promoter data sets was used as there was no significant difference observed between the two datasets

[c] Significance determination: >5.99 shows significant difference with two degrees of freedom (light grey shading)

[d] Significance determination: >3.84 shows significant difference with one degree of freedom (dark grey shading)

[e] White cells indicate promoter comparisons where there was no significant difference in expression patterns

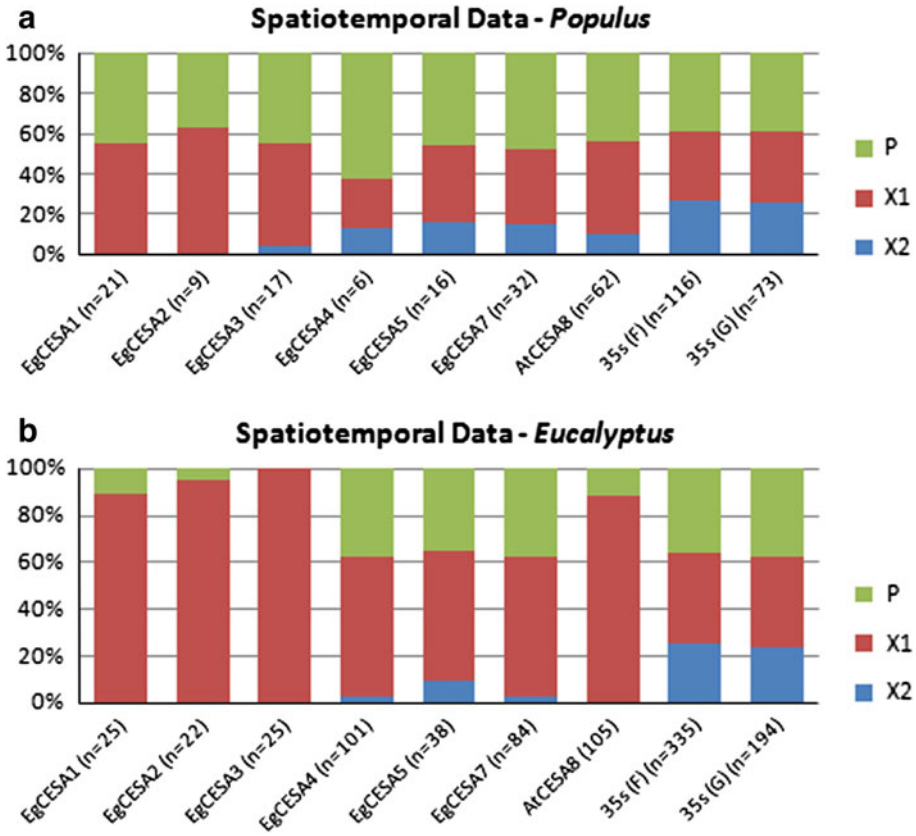


Fig. 3 Spatiotemporal frequencies of cambial staining patterns observed for the different promoter constructs in the woody stem tissues of *Populus* (**a**) and *Eucalyptus* (**b**) plants. The sector frequency of GUS expression in the cambially derived stem tissues (phloem P-*green*, immature xylem X1-*red* and mature xylem X2-*blue*) is indicated on the *y* axis and the promoter constructs are listed on the *x* axis. The *n* indicates the number of sectors counted for each promoter

and promoters. First, it allows the testing of transgenic constructs directly in native woody tissues (e.g. *Pinus*, *Populus* and *Eucalyptus*) in a relatively short period of time yielding measurable results within a few months (Spokevicius et al. 2005). Second, this technique requires little greenhouse space because multiple constructs or replicate transformations (up to 10 windows) can be performed on a single tree stem. Finally, a major advantage of ISSA is that each transgenic sector represents an independent transformation event, and with ten inoculation windows per tree a large number of independent events are produced, which can then be statistically analysed. For these reasons ISSA can be a useful technique to quickly screen promoter constructs for expression in woody stem tissues, to select candidate promoters for more detailed whole-plant analyses in model species such as *Arabidopsis* or *Populus*.

An important aspect to consider when analysing ISSA data is that stem tissue is comprised of different cell types at different developmental stages (Plomion et al. 2001) each of which may respond differently to transformation by *Agrobacterium.* Similarly, genetic background and species-specific developmental patterns may affect transformation efficiency. For example the periderm sectors, which are a result of transformed cells near the cut surface of the cambial window and have undergone rounds of division during the wounding response, were observed for most of the promoter constructs transformed into *Populus* stems. No such sectors were observed in *Eucalyptus* stems (Fig. 2) suggesting that this tissue is recalcitrant to transformation or responds differently to wounding in *Eucalyptus*. Another important consideration for analysing promoter regions using ISSA is the cell fate of the initially transformed cell and the cell- or tissue-specificity of the promoter construct. These two factors determine the final staining pattern observed and have to be jointly considered in the analysis of tissue- or cell type-specific promoters. We found it useful to compare the sector type frequencies obtained for the *CesA* promoters with those obtained for the CAMV35S promoter constructs (Fig. 3), which for the purpose of this study we assumed to be constitutively expressed in all cells derived from transformed initials. The latter is supported by the observation of a more diverse set of sector types for the CAMV35S promoter constructs including a higher frequency of wound parenchyma and tylose sectors (Fig. 2a, b), suggesting that these tissues are indeed susceptible to transformation, but that tissue-specific regulation resulted in low sector counts for these tissues when transformed with the *CesA* promoter constructs.

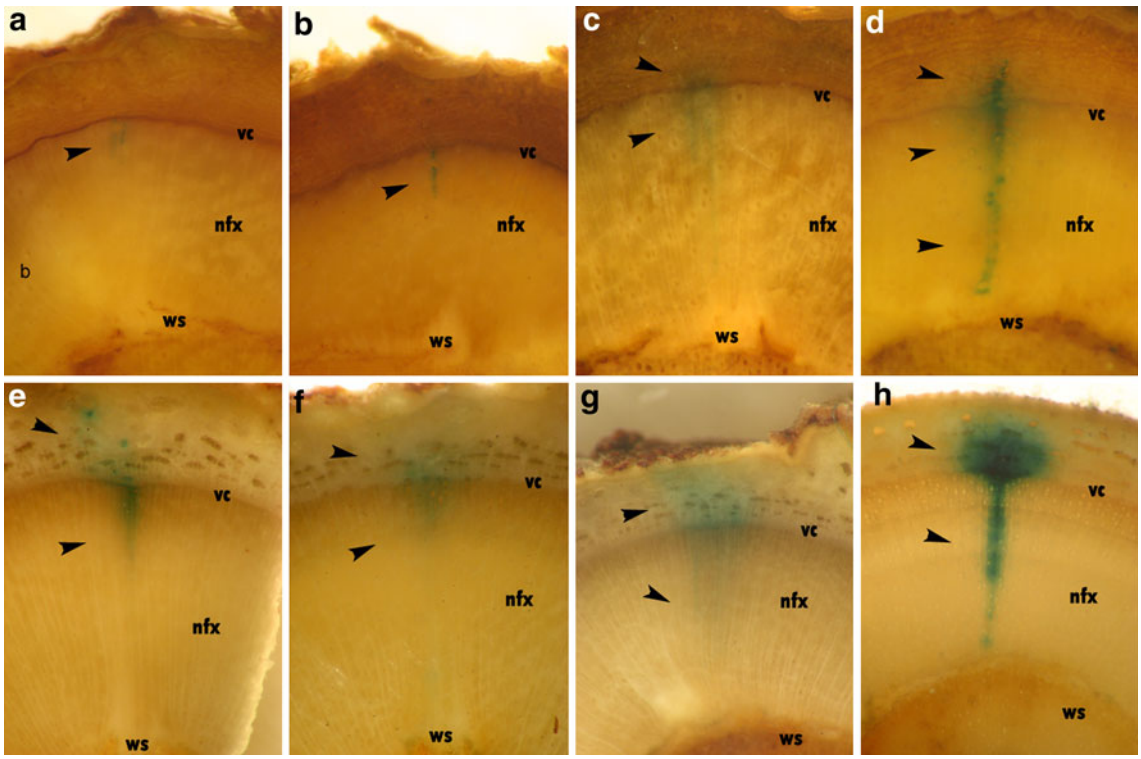


Fig. 4 Examples of cambial staining patterns observed in *Eucalyptus* and *Populus* stems. *EgCesA3* (**a**) and *AtCesA8* (**b**) promoter constructs showing activity in the X1 region only (*black arrows*) in *Eucalyptus* stems, whereas the same promoters (**e** and **f**, respectively) showed activity in the X1 and P regions in *Populus* stems. The *EgCesA5* promoter construct showing activity in the X1 and P regions in *Eucalyptus* (**c**) and Populus (**g**) stems. *35S* promoter activity was often seen in all three regions (X2, X1 and P) in *Eucalyptus* (**d**) and *Populus* (**h**) stems. *nfx* newly formed xylem, *vc* vascular cambium and *ws* wound site

The CAMV35S control promoter was expressed in all stem tissues and exhibited similar cambial expression patterns in *Eucalyptus* and *Populus* stems

One of the aims of this study was to investigate the suitability of ISSA for assessing and comparing promoter activity in woody stems across plant species and genera. Towards this end, we first asked whether the observed sector types and staining patterns were indeed comparable among species, because it has been noted before that even constitutive promoters such as CAMV35S can show variable expression across species (Qin et al. 2010). Similar to previous results (Van Beveren et al. 2006) we found that the CAMV35S promoter was active in most sampled stem tissue types (Fig. 2a, b) and this was to be expected as the CAMV35S promoter is constitutive and will express GUS in most plant tissues (Odell et al. 1985; Jefferson et al. 1987; Benfey and Chua 1990). The comparison of CAMV35S driven GUS expression patterns in cambial derived sectors observed in *Eucalyptus* and *Populus* stems did not show any significant differences (Table 5); however, other sector types such as the tylose sectors exhibited very different frequencies presumably due to differences in the susceptibility of cell types to transformation (Fig. 2a, b). Together, these results suggest that ISSA can be used to compare promoter expression patterns across species using either sector type (Fig. 2) or cambial staining patterns (Fig. 3), provided that the inherent differences in transformation efficiency are accounted for by using a standard constitutive promoter construct such as CAMV35S.

Staining patterns for *CesA* promoters in cambial derived tissues showed clear grouping of primary and secondary cell wall-related promoters

In a previous study, the expression patterns of the *Arabidopsis* (*AtCesA8*) and *Eucalyptus* (*EgCesA1*) promoters were analysed using promoter::GUS assays in *Arabidopsis* plants (Creux et al. 2008). This confirmed the secondary cell wall-related expression patterns of these two functional orthologs (Ranik and Myburg 2006). We included the same two promoter constructs in this study to allow comparison of the ISSA results to that obtained by whole-plant transformation in *Arabidopsis*. We found that the cambial staining patterns obtained in *Eucalyptus* stems for these two promoter constructs, as well as for the other *CesA* genes (Fig. 3), were consistent with the expression patterns previously observed for these genes (Taylor et al. 2003;

Taylor 2008; Ko et al. 2012). The *EgCesA1*, *2*, *3* and *AtCesA8* promoters produced GUS staining mostly in developing xylem cells (X1) which actively deposit secondary cell walls before the onset of PCD (Plomion et al. 2001) and are thus expected to show high *EgCesA1*, *2* and *3* expression levels. Their distinct expression patterns may explain the lower net ATS^{-2} values ($ATS^{-2} = 0.7$ for *EgCesA1* to $ATS^{-2} = 2.0$ for *EgCesA3*) observed for the secondary cell wall-associated *Eucalyptus CesA* promoters (Table 3). In contrast, higher ATS^{-2} values were observed for the primary cell wall-related promoters (*EgCesA4*, *5* and *7*), which reflected their expression in a wider range of cell types such as phloem (P), developing xylem (X1) and ray cells in mature xylem (X2) tissues. These results demonstrate that the ISSA approach was able to discriminate the distinct expression patterns of the *Eucalyptus CesA* genes in woody stem tissues.

The staining patterns of the *CesA* promoter constructs were not as distinctive in *Populus* stems as was observed for the primary and secondary cell wall-associated *CesA* genes in *Eucalyptus* stems (Figs. 2c, d, 3a). In particular, the three secondary cell wall-related *Eucalyptus CesA* promoters (*EgCesA1, 2* and *3*) did not predominantly produce developing xylem (X1) expression in *Populus* stems, but were expressed at equal frequency in phloem (P) and developing xylem (X1) tissues. This could be the result of differences between the regulatory networks of the two genera and has been reported in a number of other plant promoter studies (Zhang et al. 2003; Fei et al. 2006; Qin et al. 2010). While the transcriptional network regulating secondary cell wall deposition is thought to be largely conserved across plant species and genera (Zhong et al. 2010), there may be important differences in promoter sequence and transcription factor binding sites of these species. In well-studied models such as humans, fruit flies and yeast, cis-regulatory variation has been shown to be relatively common (Ho et al. 2009; Dowell 2010; Mu et al. 2011) and could underlie differences in reporter gene expression observed for the same promoter construct in different species, as was found in this study. Cis-element evolution within promoter sequences can give rise to sub-functionalization of duplicated gene loci in organisms such as *Populus*, which have undergone genome-wide or segmental duplications (Tuskan et al. 2006). Furthermore, the NAC domain transcription factor family harbouring many of the key transcription factors involved in secondary cell wall formation is highly expanded in some plant genomes and the duplicated genes may be under different evolutionary pressures (Hu et al. 2010). These differences may explain the variation observed in reporter gene expression from different genetic backgrounds. Other possible sources of variation in the reporter gene expression observed for these two species could be on an anatomical or development level, but would require further investigation to elucidate this complex issue.

Conclusion

In this study we show that ISSA is an efficient approach to investigate promoter expression in the stems of woody plants such as *Populus* and *Eucalyptus*. ISSA requires less time and space to test promoters in woody stems than whole-plant transformation and regeneration and provides ample independent transformation events for statistical analysis. However, it is important to include appropriate controls to interpret the ISSA staining patterns produced by transforming multiple cell types and using promoters with cell type- or developmental stage-specific expression. We found that the *CesA* promoter constructs produced distinct staining patterns in woody stem tissues consistent with the predicted roles of the corresponding *CesA* genes in primary and secondary cell wall formation. Our results suggest that, while many aspects of the secondary cell wall transcriptional network are conserved (Zhong et al. 2010), there are regulatory differences which should be considered when testing promoters in heterologous systems. ISSA should be applicable to a wider range of woody plants and various secondary cell wall-related promoters could be analysed in this manner, which will be important for elucidating the transcriptional control of woody biomass production.

Acknowledgments We are grateful to Martin Ranik for providing the *CesA*-promoter::GUS constructs which made this study possible and Minique De Castro for aiding in the construction of the second CAMV35S-promoter::GUS construct. We also thank Julio Najera, Valerie Frassiant and Angelique Manuel for laboratory assistance. This work was supported through funding provided by Mondi and Sappi to the Forest Molecular Genetics (FMG) Programme, the Technology and Human Resources for Industry Programme (THRIP) and the National Research Foundation (NRF) of South Africa as well as a Linkage Grant from the Australian Research Council (LP0776563) to GB, AAM and AVS.

References

Alwen A, Moreno RMB, Vincente O, Heberle-Bors E (1992) Plant endogenous β-glucuronidase activity: how to avoid interference with the use of the *E. coli* β-glucuronidase as a reporter gene in transgenic plants. Transgenic Res 1:63–70

Arioli T, Peng L, Betzner AS, Burn J, Wittke W, Herth W, Camilleri C, Hofte H, Plazinski J, Birch R, Cork A, Glover J, Redmond J, Williamson RE (1998) Molecular analysis of cellulose biosynthesis in *Arabidopsis*. Science 279:717–720

Beeckman T, Przemeck Gerhard KH, Stamatiou G, Lau R, Terryn N, De Rycke R, Inze D, Berleth T (2002) Genetic complexity of CESA gene function in *Arabidopsis* embryogenesis. Plant Physiol 130:1883–1893

Benfey PN, Chua N-H (1990) The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants. Science 250:959–966

Carroll A, Specht CD (2011) Understanding plant cellulose synthases through a comprehensive investigation of the cellulose synthase family sequences. Front Plant Sci 2:5

Creux NM, Ranik M, Berger DK, Myburg AA (2008) Comparative analysis of orthologous cellulose synthase promoters from *Arabidopsis*, *Populus* and *Eucalyptus*: evidence of conserved regulatory elements in angiosperms. New Phytol 179:722–737

Curtis MD, Grossniklaus U (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol 133:462–469

Desprez T, Vernhettes S, Fagard M, Refrégier G, Desnos T, Aletti E, Py N, Pelletier S, Höfte H (2002) Resistance against herbicide isoxaben and cellulose deficiency caused by distinct mutations in same cellulose synthase isoform CESA6. Plant Physiol 128:482–490

Djerbi S, Lindskog M, Arvestad L, Sterky F, Teeri TT (2005) The genome sequence of black cottonwood (*Populus trichocarpa*) reveals 18 conserved cellulose synthase (*CesA*) genes. Planta 221:739–746

Dowell RD (2010) Transcription factor binding variation in the evolution of gene regulation. Trends Genet 26:468–475

Fei H, Chaillou S, Hirel B, Polowick P, Mahon JD, Vessey JK (2006) Effects of the overexpression of a soybean cytosolic glutamine synthetase gene (*GS15*) linked to organ-specific promoters on growth and nitrogen accumulation of pea plants supplied with ammonium. Plant Physiol Biochem 44:543–550

Grattapaglia D, Plomion C, Kirst M, Sederoff RR (2009) Genomics of growth traits in forest trees. Curr Opin Plant Biol 12:148–156

Hamann T, Osborne E, Youngs HL, Misson J, Nussaume L, Somerville C (2004) Global expression analysis of *CESA* and *CSL* genes in *Arabidopsis*. Cellulose 11:279–286

Hinchee M, Rottmann W, Mullinax L, Zhang C, Chang S, Cunningham M, Pearson L, Nehra N, Tomes D, Lakshmanan P, Songstad D (2011) Short-rotation woody crops for bioenergy and biofuels applications. In: Dwight T, Lakshmanan P, songstad D (eds) Biofuels. Global impact on renewable energy, production agriculture, and technological advancements. Springer, New York, pp 139-156

Ho MCW, Johnsen H, Goetz SE, Schiller BJ, Bae E, Tran DA, Shur AS, Allen JM, Rau C, Bender W, Fisher WW, Celniker SE, Drewell RA (2009) Functional evolution of *cis*-regulatory modules at a homeotic gene in *Drosophila*. PLoS Genet 5:e1000709

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

Hussey S, Mizrachi E, Spokevicius A, Bossinger G, Berger D, 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 Biol 11:173

Jefferson RA, Kavanagh TA, Bevan MW (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J 6:3901–3907

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

Ko J-H, Kim H-T, Hwang I, Han K-H (2012) Tissue-type-specific transcriptome analysis identifies developing xylem-specific promoters in poplar. Plant Biotechnol 10:587–596

Kumar M, Thammannagowda S, Bulone V, Chiang V, Han K-H, Joshi CP, Mansfield SD, Mellerowicz E, Sundberg B, Teeri T, Ellis BE (2009) An update on the nomenclature for the cellulose synthase genes in *Populus*. Trends Plant Sci 14:248–254

Lazo G, Stein P, Ludwig R (1991) A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. Biotechnology 9:963–967

Mansfield SD (2009) Solutions for dissolution-engineering cell walls for deconstruction. Curr Opin Biotech 20:286–294

Mizrachi E, Mansfield SD, Myburg AA (2012) Cellulose factories: advancing bioenergy production from forest trees. New Phytol 194:54–62

Mu XJ, Lu ZJ, Kong Y, Lam HYK, Gerstein MB (2011) Analysis of genomic variation in non-coding elements using population-scale sequencing data from the 1000 Genomes Project. Nucleic Acids Res 39:7058–7076

Odell JT, Nagy F, Chua N-H (1985) Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature 313:810–812

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

Popper ZA, Michel G, Herve C, Domozych DS, Willats WGT, Tuohy MG, Kloareg B, Stengel DB (2011) Evolution and diversity of plant cell walls: from algae to flowering plants. Annu Rev Plant Biol 62:567–590

Qin JY, Zhang L, Clift KL, Hulur I, Xiang AP, Ren B-Z, Lahn BT (2010) Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. PLoS ONE 5:e10611

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

Rathmann R, Szklo A, Schaeffer R (2010) Land use competition for production of food and liquid biofuels: an analysis of the arguments in the current debate. Renew Energ 35:14–22

Richmond TA, Somerville CR (2000) The cellulose synthase superfamily. Plant Physiol 124:495–498

Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual. Cold Spring Harbour Press, New York

Samuga A, Joshi CP (2004) Differential expression patterns of two new primary cell wall-related cellulose synthase cDNAs, *PtrCesA6* and *PtrCesA7* from aspen trees. Gene 334:73–82

Scheible WR, Eshed R, Richmond T, Delmer D, Somerville C (2001) Modifications of cellulose synthase confer resistance to isoxaben and thiazolidinone herbicides in *Arabidopsis ixr1* mutants. Proc Natl Acad Sci USA 98:10079–10084

Seguin A (2011) How could forest trees play an important role as feedstock for bioenergy production? Curr Opin Environ Sustain 3:90–94

Song D, Shen J, Li L (2010) Characterization of cellulose synthase complexes in *Populus* xylem differentiation. New Phytol 187:777–790

Spokevicius AV, Van Beveren K, Leitch MM, Bossinger G (2005) *Agrobacterium*-mediated in vitro transformation of wood-producing stem segments in eucalypts. Plant Cell Rep 23:617–624

Spokevicius AV, Southerton SG, MacMillan CP, Qiu D, Gan S, Tibbits JFG, Moran GF, Bossinger G (2007) β-Tubulin affects cellulose microfibril orientation in plant secondary fibre cell walls. Plant J 51:717–726

Stork J, Harris D, Griffiths J, Williams B, Beisson F, Li-Beisson Y, Mendu V, Haughn G, DeBolt S (2010) CELLULOSE SYNTHASE9 serves a nonredundant role in secondary cell wall synthesis in Arabidopsis epidermal testa cells. Plant Physiol 153:580–589

Suzuki S, Li LG, Sun YH, Chiang VL (2006) The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthase-like genes in *Populus trichocarpa*. Plant Physiol 142:1233–1245

Taylor NG (2008) Cellulose biosynthesis and deposition in higher plants. New Phytol 178:239–252

Taylor NG, Laurie S, Turner SR (2000) Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in *Arabidopsis*. Plant Cell 12:2529–2539

Taylor NG, Howells RM, Huttly AK, Vickers K, Turner SR (2003) Interactions among three distinct CESA proteins essential for cellulose synthesis. Proc Natl Acad Sci USA 100:1450–1455

Turner SR, Somerville CR (1997) Collapsed xylem phenotype of *Arabidopsis* identifies mutants deficient in cellulose deposition in the secondary cell wall. Plant Cell 9:689–701

Tuskan GA, DiFazio S, Jansson S et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596–1604

Van Beveren K, Spokevicius AV, Tibbits J, Wang Q, Bossinger G (2006) Transformation of cambial tissue in vivo provides efficient means for Induced Somatic Sector Analysis (ISSA) and gene testing in stems of woody plants species. Funct Plant Biol 33:629–638

Yin Y, Huang J, Xu Y (2009) The cellulose synthase superfamily in fully sequenced plants and algae. BMC Plant Biol 9:99

Zhang P, Bohl-Zenger S, Puonti-Kaerlas J, Potrykus I, Gruissem W (2003) Two cassava promoters related to vascular expression and storage root formation. Planta 218:192–203

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