EXPRESSION PROFILING AND CHARACTERIZATION OF WOOD FORMATION GENES IN *EUCALYPTUS*

by

MARTIN RANIK

Submitted in partial fulfillment of the requirements for the degree

Magister Scientiae

In the Faculty of Natural of Agricultural Sciences

Department of Genetics

University of Pretoria

Pretoria

April 2005

Under the supervision of Dr Alexander A. Myburg

DECLARATION

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

Martin Ranik

April 2005

PREFACE

The genus *Eucalyptus* includes some of the most widely planted forest trees in the Southern hemisphere and interspecific hybrids as well as pure species of *Eucalyptus* are amongst the most prolific producers of wood known. Despite the fact that wood is an irreplaceable natural product which forms the basis of a massive global industry, relatively little is known about the structural and regulatory genes that govern its formation in forest trees. A thorough understanding of the molecular biology of wood development is required before any attempts can be made at improving wood and fibre quality of *Eucalyptus* trees. To achieve this, methods are needed to rapidly identify and analyse the function of wood formation genes and obtain an understanding of the genetic structure and evolution of wood formation gene families. These include not only structural genes, responsible for the biosynthesis of the cell wall biopolymers including lignin and cellulose, but also the regulatory and developmental genes that control cell fate and differentiation.

Chapter 1 of this dissertation comprises a brief overview of the literature relevant to our current knowledge of the molecular biology of wood formation. It highlights the important stages of wood development, both on the morphological and genetic level. Focus is also placed on current transcript profiling technologies and their relative applicability to studying gene expression during wood formation.

One of the major impediments in gene discovery research is the requirement for a high-throughput, cost-effective and flexible approach for transcript profiling. Establishing such systems is especially challenging in species for which little or no DNA sequence information is available, such as *Eucalyptus* tree species. **Chapter 2** of this dissertation describes the use of cDNA-AFLP transcript profiling and infrared detection technology for analysing the transcriptome of the developing secondary vasculature in the adult *Eucalyptus* stem. The usefulness of this technique for the isolation of novel genes as well as its accuracy in determining their expression levels are also evaluated in this chapter.

The cell walls of fibres in wood are composed almost entirely of cellulose and trees owe their potentially massive size to the strength and versatility of this relatively simple compound. Despite the fact that cellulose is the most abundant biopolymer on our planet, the molecular mechanism of its deposition in plant cell walls is poorly understood. It has recently become evident that a multisubunit transmembrane enzyme complex composed of distinct cellulose synthase catalytic subunits (*CesA*) carries out the polymerisation of glucose into cellulose. **Chapter 3** describes the isolation and comprehensive expression profiling of six distinct full-length cellulose synthase catalytic subunit genes, representing a major portion of the *Eucalyptus CesA* gene family, starting from a single short transcript-derived fragment isolated using cDNA-AFLP analysis as developed in Chapter 2.

The findings presented in this dissertation represent the outcomes of a study undertaken from March 2003 to March 2005 in the Department of Genetics, University of

Pretoria, under the supervision of Dr A.A. Myburg. Chapters 2 and 3 have been prepared as independent manuscripts and were submitted for review to the journals "Tree Physiology" and "Plant Molecular Biology" respectively. Therefore, a certain degree of redundancy may exist between the introductory sections of these chapters and Chapter 1. The following manuscripts, congress presentations and posters and were generated based on the results of this study.

RANIK, M., Creux, N. and Myburg, A.A. 2004. Transcriptome analysis of xylogenesis in *Eucalyptus* using cDNA-AFLP and LI-COR automated DNA analysers. Plant & Animal Genome XII Conference P610, January 10–14. San Diego, CA. (Poster presentation).

RANIK, M., Creux, N. and Myburg, A.A. 2004. High-throughput cDNA-AFLP profiling of gene expression during xylogenesis in *Eucalyptus*. South African Genetics Society Conference, April 4–7. Stellenbosch. (Verbal presentation).

RANIK, M., Creux, N. and Myburg, A.A. 2005. Transcript profiling across the *Eucalyptus* woody stem. American Society of Plant Biologists, Plant Biology Meeting, July 16–20. Seattle, WA. (Poster presentation).

RANIK, M., Creux, N.M. and Myburg, A.A. 2005. Within-tree transcriptome profiling in wood-forming tissues of a fast-growing *Eucalyptus* tree. Tree Physiology (In Press).

RANIK, M. and Myburg, A.A. 2005. Six new cellulose synthase genes from *Eucalyptus* are associated with primary and secondary cell wall biosynthesis. Plant Molecular Biology (Manuscript submitted).

ACKNOWLEDGEMENTS

I would like to express my gratitude to the following people, organizations and institutes for assisting me in the completion of this project:

- To Dr A.A. Myburg, for introducing me to the idea of studying the molecular biology of wood formation and for his professional, thorough, creative and insightful leadership of this project.
- To Nicky Creux, for her indefatigable dedication to the development of the transcript profiling system described in Chapter 2.
- To all my past and present colleagues in the Forest Molecular Genetics Laboratory: Elna Cowley, Anita Steyn, Adrene Laubscher, Yoseph Beyene, Minique DeCastro, Frank Maleka, Michelle Victor, John Kemp, Honghai Zhou, Luke Solomon and Dr Solomon Fekybelu, for their valuable insights and for forming a superb research collective.
- To Prof. Dave Berger, for excellent advice and encouragements.
- To Mondi Business Paper South Africa and Sappi Forest Products, for supplying the plant materials used in this study and for the first-rate assistance provided by their employees during collecting trips, as well as for funding contributions to this project.
- To the Genetics Department of the University of Pretoria and the Forestry and Agricultural Biotechnology Institute (FABI) for providing facilities and a sound academic environment.
- To the National Research Foundation of South Africa (NRF), for the grant holder-linked scholarship which I received.
- To The Human Resources and Technology for Industry Programme (THRIP), for financial support of the research.

TABLE OF CONTENTS

CHAPTER 1

LITERATURE REVIEW

THE MOLECULAR BIOLOGY OF WOOD FORMATION IN FOREST TREES

INTRODUCTION

Wood is one of the most important natural products with a multitude of applications. Despite the importance of the forest biome, the majority of wood is currently destructively harvested from natural forests. Sustainable harvesting, however, is comparatively low-yielding and complicated by the inaccessibility of a majority of natural forests (Fenning and Gershenzon, 2002). It is therefore plantation forestry which seems to hold the most promise for the future. The concept of improvement now becomes paramount: if trees are to become cultivated or domesticated species, there will be a need to create varieties in which there is a synthesis of the most desirable traits found in their wild ancestors as well as traits which are not normally evident in natural populations. A simple parallel can be drawn to the domestication of crop species such as maize or wheat. The current domesticated varieties are by far superior in many aspects to their wild ancestors, which were first utilized as food thousands of years ago. However, the fact that forest trees have naturally long generations times has precluded them from undergoing a similar agricultural revolution, as was the case with many other plant species (Campbell *et al.*, 2003). The problem posed is the finding of the most suitable and expedient methods to obtain these improved varieties. Biotechnology, it would seem, will have to play a central role.

The domestication of any organism is centred around breeding – crossing and selecting desirable traits in order to maximise output of the resource for which the organism is being exploited. Molecular breeding has become standard for many cultivated species (Koornneef and Stam, 2001). The underlying factor is a marker, or a molecular characteristic that is linked to a desirable phenotypic trait. Theoretically, the association between molecular markers and useful phenotypic traits can be used in designing breeding approaches. The application of marker-assisted breeding in forest trees is not as simple as in many other crop

species. Long generation times and difficulties associated with the inability to generate inbred lines pose severe problems to the employment of this approach in tree breeding (Strauss *et al.*, 1991; Van Raemdonck *et al.*, 2001). Despite this, forest tree molecular breeding has a promising future as more detailed phenotypic and marker data becomes available (Wu *et al.*, 2000). One of the main factors supporting the use of breeding rather than genetic modification is the negative public opinion to genetic engineering.

Recently, forest biotechnology has focused on more direct approaches aimed at improving forest trees: dissecting the process of wood formation, gene discovery, genome sequencing, the use of model plant species and the eventual genetic engineering of the target to achieve improvement (Bhalerao *et al.*, 2003). In the last decade, wood formation studies have shifted focus away from morphology. Current research is more intent on elucidating the genetic mechanisms that govern wood development and properties. This approach ranges from transcriptome analysis (Milioni *et al.*, 2001; Demura *et al.*, 2002; Milioni *et al.*, 2002) to comprehensive dissection of biochemical pathways by transgenic approaches (Li *et al.*, 2003). Perhaps the most significant impact on forest tree genomics will be made by the completed sequencing of tree genomes including that of *Populus trichocarpa* (Wullschleger *et al.*, 2002) and *Eucalyptus camaldulensis* (in progress). However, the genomes of certain economically important forest trees such as *Pinus* species will probably not be sequenced in the foreseeable future due to their sheer size (in excess of 20,000 Mb).

One of the main applications of genes characterised by functional genomic approaches is the genetic engineering of the target species in order to directly improve specific characteristics such as altering the chemical composition of wood. The main obstacle in manipulating forest trees is the difficulty associated with the transformation of some of the commercial trees including species of *Eucalyptus* and *Pinus*. In poplar (*Populus*), however, work on manipulating wood biochemistry aiming to demonstrate the individual and additive

effect of the various pathway enzymes was well demonstrated by Li *et al.* (2003). This study showed that fairly sophisticated combinations of transgene overexpression and simultaneous knock-down of endogenous genes can be used to analyse biochemical and regulatory pathways in forest tree species. Recently, progress in the field of RNA-induced silencing (RNAi) has provided a flexible system for the knockdown of gene expression by the expression of small interfering RNAs (Lu *et al.*, 2004).

Wood formation, or xylogenesis, has been studied on the anatomical level for decades. Only recently has significant progress been made towards the understanding of the complex chemistry of wood as well as the genes which govern its formation. This review highlights some of the recent work regarding the molecular biology of xylogenesis, focusing on the genetic basis of the main stages of wood development. An evaluation is also made of the current and potential methods that can be used to discover genes involved in wood formation.

GENETICS AND BIOCHEMISTRY OF WOOD FORMATION

Wood is the evolutionary answer to the colonisation of dry land by plants more than 350 million years ago and is principally a further step in the complexity of transport tissues of woody plants. The biochemical intricacy and variability of wood is the result of a highly organized expression of regulatory as well as structural genes during its formation. Recent research has adopted an integrative approach aimed at the understanding of this complex process – a synthesis of anatomy, chemistry and genetics.

Overview of the development of the woody stem

All vascular plants share the two basic transport tissues: xylem and phloem, which develop from a circular layer of undifferentiated cells, known as the vascular cambium. In trees and other wood-forming species, the development of these tissues is expanded to a secondary

level – where their functioning as a support is on par with the transport function of these tissues – without either the plant could not survive.

Plomion *et al.* (2001) describe the process of wood formation to encompass at least five major steps: 1) cell division (cambium cells divide to form xylem on the inner side and phloem on the outer side), 2) cell elongation 3) cell wall thickening 4) programmed cell death and 5) heartwood formation. Cellular and biochemical processes accompanying each of these main steps impact on the eventual structure and composition of wood. In turn, the structure and composition of wood have significant implications for the processing of wood. Notable characteristics include cell wall chemistry, fibre length and orientation. Understanding the genes that control these characteristics is essential to any program attempting to improve wood quality using biotechnology.

Basic vascular tissue organization

Xylem and phloem tissues arise from a narrow region, few cell layers thick, known as the vascular cambium which, in mature stems, forms a ring spanning the circumference of the stem. Vascular cambium, a secondary lateral meristem, has been well characterised and is highly conserved at the structural level in a wide variety of plants, as reviewed by Larson (1994) and Lachaud *et al.* (1999). The arrangement of the differentiating vascular tissues is concentric: cambial initial cells reside at the centre and give rise to xylem and phloem mother cells on opposite sides. The cambial initials can be classed into fusiform initial and ray initials. The elongated fusiform initial cells differentiate sequentially into progressively more mature xylem and phloem structures, whereas the shorter ray initials give rise to rays composed of longer-lived cells, which allow communication and transverse nutrient transport between the secondary xylem and phloem tissues (Plomion *et al.*, 2001).

Pits in xylem elements allow the interaction of cell contents between adjoining cells via plasmodesmata – cytoplasmic extensions between two cells. Chaffey and Barlow (2002)

have presented strong evidence of the involvement of myosin and associated cytoskeletal elements in the control of active transport of messages and nutrients across the plasmodesmata in ray and axial parenchyma cells. It seems that a complex three-dimensional, symplastic, continuum exists between the various vascular tissues – allowing active and controlled, transverse and lateral, communication and transport to occur in the secondary vasculature.

The similarity observed at the early levels ends with the differentiation of the mother cells into specialized xylem and phloem cell types. In angiosperms, xylem mother cells develop into xylem parenchyma, vessels and fibres whereas phloem mother cells give rise to sieve tubes and companion cells. Gymnosperm vasculature is less complex in terms of the variety of cell types present – mature xylem tissue is composed of tracheids alone.

Chemical composition of cell walls

One of the main stages in wood formation is the secondary thickening of the cell wall, which is also one of the major factors determining the composition and structure of wood. Secondary cell walls are composed of cellulose (a homopolymeric β-1,4-glucan), lignin (a polymer of various substituted phenolic compounds), hemicelluloses (which are polysaccharides composed of non identical carbohydrate monomers) and proteins (Hu *et al.*, 1999). The latter components (hemicelluloses and lignin) are heteropolymeric compounds with variable compositions. The resulting heterogeneity of the plant cell wall permits excessive specialization of the cells in woody tissues. This is especially notable in the variability of cell wall components observed between the primary and secondary cell wall as well as between different types of wood and wood from different species (Mellerowicz *et al.*, 2001). The presence of the three major constituents in an approximate ratio of 1:2:1 (lignin: cellulose: hemicelluloses) seems to be approximately common to wood and underlies one of the problems encountered during the processing of wood during paper manufacturing. The

cellulose content is desirable whereas the lignin content poses problems for the downstream processing of wood: its removal is costly and environmentally detrimental. The nature of wood is such that these three basic components interact in an impressively complex system.

Hormonal regulation of vascular development

At each phase of xylogenesis, the process is regulated by the interaction of the differentiating cells with global impulses such as hormone signalling as well as local cell-cell interactions (Kuriyama and Fukuda, 2002). The regulation and timing of these processes is an intriguing aspect of plant development and may be one of the keys to the ability to control wood development of commercial cultivars in the future. Friml (2003) describes the interaction of environmental signals and an all-encompassing factor in the regulation of wood formation – the hormone auxin (and related hormones). Recently, genes involved in the polar transport of auxin in developing vascular tissues have been characterised by Schrader (2003), who found four genes actively involved in the development of an auxin gradient governing the stages of vascular cambium differentiation in the poplar stem.

On the morphological level, Little *et al.* (2002) have demonstrated that auxin flow from the apex promotes the growth of interfasicular cambium in *Arabidopsis* stems and is required for correct secondary xylem development. The currently accepted model of hormone-gene interaction cascade is centred around the interaction of Auxin Responsive Factors/ ARFs (auxin responsive transcription factors) and AUX/IAA transcriptional regulator proteins. The first ARF isolated and characterised was the *MONOPTEROS* (*MP*) auxin responsive transcription factor, which is required for correct vascular patterning in *Arabidopsis* (Hardtke and Berleth, 1998). MP possesses domains which have been shown to have affinity for specific promoter regions of auxin inducible genes. The functioning of ARF is itself regulated by the expression of the AUX/IAA proteins – ARFs are repressed by AUX/IAA proteins. Gray *et al.* (2001) demonstrated that AUX/IAA proteins degradation is

triggered by auxin – which causes the activation of ARFs and the removal of repression of the auxin response. However, subsequent induction of the AUX/IAA – also by auxin – leads to the restoration of the balance in the auxin response pathway. Most recently, Dharmasiri *et al.* (2005) found that the F-box protein TIR1 is required for auxin binding in *Arabidopsis*. This represents the first discovery of an auxin receptor which directly mediates the interaction between the AUX/IAA proteins and auxin.

Transcription factors in vascular development

An example of the complexity and redundancy of the various ARFs is made in the study of *ATHB-8*, an auxin responsive HD-ZIP (homeodomain leucine zipper) III class ARF in A*rabidopsis thaliana* (Baima *et al.*, 2001). *ATHB-8* is a member of a group of highly related transcription factors which are implicated in cambial proliferation and differentiation. The lack of distinctive phenotype in *ATHB-8* loss-of-function mutant lines suggested that there is a functional redundancy among the members of the HD-ZIP III transcription factors (Baima *et al.*, 2001). This is a somewhat unexpected situation, as it would be logical to anticipate high degrees of specialization at the regulation level. As developing cells have at their disposal more or less the same cellular machinery (structural enzymes, cytoskeleton), it is the determination of when, how and where these are applied that is controlled by the regulation via transcription factors and which in turns determines the fate of the cell. Overexpression of *ATHB-8* in transgenic *Arabidopsis* resulted in excessive proliferation of the cambium and its preferential differentiation into xylem tissue at both primary and secondary levels. It likely that auxin-regulated transcription factors do not function alone, but that there is a combinatorial effect of the transcription factors on the protein level. Recently, Kang and Dengler (2004) reported that *ATHB-8* is also involved in leaf vein pattern formation from the cambium.

The first definitive demonstration of a gene which determines whether a cambial cell

will differentiate into xylem or phloem tissues is the putative *Arabidopsis* transcription factor *ALTERED PHLOEM DEVELOPMENT* (*APL*) (Bonke *et al.*, 2003). This elegant study showed that the *apl* mutant line exhibited the lack of differentiation of protophloem cells into functional phloem cells, with the occurrence instead of cells with the distinct characteristics of xylem elements.

Other transcription factors recently proposed to play a role in regulating xylogenesis include the MADS-box class transcription factor genes which were previously shown to predominantly play roles in the control of flower and embryo development. Cseke *et al.* (2003) characterised an aspen MADS-box class transcription factor that is expressed in sites of primary and secondary xylem formation. The importance of identifying transcription factors specific to vascular development, especially in commercial and model forest trees, is clear – the possible future ability to control vascular development at a fine level is an attractive concept for wood improvement. In general, however, the current transcription factor molecular studies (at least in forest trees) are still in their infancy – targeted manipulation is yet but a concept that has to be pursued using model systems.

Cell expansion: expansins

Integral to the process of cambial differentiation into the major vascular tissues is the necessity of expanding the growing xylem mother cells to obtain the greatly elongated fibre cells. The plant cell wall, however, is a complex multi-layered structure, the perturbation of which requires complex mechanisms. Plomion *et al.* (2001) reviewed basic cell wall structure in wood-forming tissues: developing cells in the vascular tissues possess a primary cell wall composed of a relatively thin layer (0.1 micron thick) of cellulose microfibrils arranged in a random pattern and interwoven with hemicelluloses (branched heteropolymeric polysaccharides such as xylan or xyloglucan) and pectins (complex polysaccharides containing galactosyluronic acid). Outside this layer lies the middle lamella – the pectin-rich

layer filling the region between adjacent cells. It would be expected that a such dense heterogenic structure would be resistant to expansion during growth and distortion during the deposition of the thick secondary cell wall on its inside. Plants, however have evolved various proteins (proteins, in terms of total content form only a minor component of the cell wall) which enable the primary cell wall to expand and allow for wall expansion. One of the most important protein classes allowing this expansion are the expansins, as reviewed by Cosgrove (2000). Expansins have been identified to play a role in cell wall expansion in a variety of developmental processes such as pollen tube growth through the style (Pezzotti *et al.*, 2002) and fruit ripening (Civello *et al.*, 1999; Hiwasa *et al.*, 2003). Although the exact mode of action of expansins has not yet been decisively demonstrated, it is highly likely that expansins are not hydrolytic, but act by loosening non-covalent bonds between the hemicelluloses and/ or pectins, which bridge the large cellulose microfibrils. These proteins constitute a very large gene superfamily in plants and have been isolated from a wide variety of tissues (Cosgrove, 2000). Although a general review of expansins is beyond the scope of this work, it should be noted that expansins have, predictably, been isolated from developing vascular tissues. Im *et al.* (2000) report the discovery of three *Zinnia elegans* expansin genes which seem to play a role in the longitudinal intrusive cell expansion of developing xylem cells. This work suggests the necessity for multiple expansins during the growth of developing xylem elements prior to the deposition of the secondary cell wall. Recently, Gray-Mitsumune *et al.* (2004) have identified four new expansin genes in hybrid aspen, each showing specific expression in different tissues. One of the four expansins was shown to be highly expressed in developing xylem cells while another (highly similar) gene was found to be expressed in tension wood. The expression of an alternative expansin in tension wood (reaction wood to mechanical stress) suggests that a high degree of specialization exists within the expansin gene family.

Cell expansion: non expansin enzymes

Proteins other than expansins have been implicated to play a role in plant cell wall expansion, of which the two most notable are Xyloglucan Endotransglycosylase (XET) and β-1-4 endoglucanase, enzymes associated with the editing of the hemicellulose component of the cell wall. XET, as reported by Fry *et al.* (1992), catalyses the cleavage of xyloglucan chains and the transfer of the cleaved segments to non-reducing ends of other xyloglucan chains. This way, XET allows expansion of the cell wall by altering the length of the xyloglucan chains linking adjacent cellulose microfibrils. XET has been isolated from a variety of plant species such as cauliflower (Henriksson *et al.*, 2003), tomato (Albert *et al.*, 2004) and hybrid aspen (Johansson *et al.*, 2004). Enzyme purification in these cases has allowed more precise predictions of the proteins` functional domains than has been possible for other cell-wall enzymes (such as the case of the higher plant cellulose synthases, which are yet to be purified sufficiently for crystallographic analysis). Recently, a second hemicellulose endotransglycosylase was identified. Mannan transglycosylase was shown to act on mannan-containing hemicellulose chains in a manner very similar to XET (Schroder *et al.*, 2004). It is an indication that the complex heterogeneity of the plant cell wall is matched by an impressive array of similarly acting enzymes each with a specific substrate.

Egases (β-1-4 endoglucanases) act similarly to XETs but, rather than modifying the hemicellulose chains, they catalyse a hydrolytic cleavage of xyloglucans. In higher plant Egases, the most recent breakthrough has been the determination of the underlying gene in the *Arabidopsis* mutant *irx2*. Szyjanowicz *et al.* (2004) have shown that the *irregular xylem 2* (*irx2*) mutation is caused by the loss of function of a putative β-1-4 endoglucanase, previously characterised as KORRIGAN (Nicol *et al.*, 1998), which causes a dwarf phenotype coupled to a drastic reduction of cellulose synthesis in secondary cell walls. All other previously isolated mutant alleles of KORRIGAN have displayed reduction of cellulose

in primary, but not secondary cell walls. This study was particularly interesting as this specific *irx* mutation was not in a cellulose synthase gene, suggesting that other enzymes are required for correct cellulose microfibril deposition, though not necessarily being part of the cellulose synthase complex. These findings are complementary to the conclusions arrived at by Peng *et al.* (2002), who showed that a sitosterol linked primer is required for cellulose synthesis and proposed that an endoglucanase may be required to cleave off residues from this primer prior to their incorporation into the growing cellulose chain.

Cellulose biosynthesis

Cellulose is the most abundant polymer of natural origin (biopolymer) on the planet. The composition of cellulose is relatively simple (a (1→4)-β-D-glucan composed of polymerised glucose subunits), yet the process of its synthesis is fairly complex. Cellulose occurs naturally in the form of fibres which are composed of intertwined chains of the 1-4 glucan. It is ubiquitous in plant cell walls and is key in the determination of the vast variety of cell, and consequently, plant organ morphology.

Up to 36 continuous glucose chains combine to form a microfibril in plant cell walls. The space between cellulose microfibrils is occupied by lignin and hemicelluloses. Microfibril deposition occurs on the outside of the cell membrane and differs significantly in the two major cell wall types found in plants. Primary cell walls (common to most differentiated plant cells) are thin (0.1 μm), flexible layers of microfibrils deposited in a random manner during cell differentiation (Mellerowicz *et al.*, 2001; Plomion *et al.*, 2001). During the subsequent maturation stages, the secondary cell wall (which can be up to 100 times thicker than the primary wall) is deposited between the primary cell wall and the cell membrane.

The microfibrils in the secondary cell wall are arranged at very specific angles which differ greatly between species and types of wood (Barnett and Bonham, 2004). For example,

in reaction wood (wood formed in regions under mechanical stress such as the upper sides of branches) the microfibril angle will be smaller when compared to "normal" wood, thus increasing the strength and elasticity of the wood. Cellulose microfibril angle plays an important role in structural support but is also an important factor in the processing of wood into paper via pulping (Kibblewhite, 1999), by affecting the strength of fibres and thereby the paper.

Cellulose biosynthesis has been studied for years but major strides in the discovery of genes which play a role in the deposition of this remarkable polymer have only been made recently, reviewed in Delmer (1999) and Doblin *et al.* (2002). Although cellulose is common to all plant cells, the first advances in elucidating its biosynthesis were made in cellulose producing bacteria such as *Acetobacter xylinum* and *Agrobacterium tumefaciens* (Ross *et al.*, 1991), and included the isolation and partial sequencing of the first functional cellulose synthase protein. Interestingly, the first plant cellulose synthase genes were only characterised relatively recently (Pear *et al.*, 1996). In this study, two cellulose synthase cDNAs were discovered by screening clones from a cotton fibre cDNA library for the presence of amino acid motifs particular to the bacterial cellulose synthases.

In plants the cellulose microfibril is synthesised by a multisubunit rosette-shaped enzyme complex associated with the cell membrane (cellulose synthase complex, CSC). The rosette contains up to 36 subunits, each with cellulose synthase activity, which are encoded by cellulose synthase catalytic subunit genes or *CesA*s (Delmer, 1999). *CesA*s have been characterised in a number of plant species including *Arabidopsis* (Richmond, 2000), poplar (Joshi *et al.*, 2004), cotton (Pear *et al.*, 1996) and barley (Burton *et al.*, 2004). *CesA*s are enzymes with glycosyl transferase activity, specifically capable of catalysing β 1→4 bond formation, which involves the sequential inversion of the substrate by 180° (Saxena *et al.*, 1995). *CesA*s are present as a 8-10 member gene family in higher plants and the individual

members have become remarkably specialised. Studies of cell wall mutants in *Arabidopsis* have shown that three distinct *CesA*s are required for the synthesis of cellulose (Taylor *et al.*, 1999; Taylor *et al.*, 2000; Taylor *et al.*, 2003). Inactivation of any one of these three distinctive single-copy genes brings about drastic reduction in cellulose content coupled to massive phenotypic abnormalities. Moreover, it is clear that different *CesA*s are involved in the synthesis of the primary cell wall than are required for the synthesis of the secondary cell wall. The main obstacle currently facing the community of researchers studying cellulose synthases is that none of the cellulose synthases have yet been purified sufficiently for crystallographic analysis. This is primarily due to the fact that CESAs are large (ca. 120 kDa) membrane-bound proteins with a high turnover rate (Doblin *et al.*, 2002).

Cellulose biosynthesis is a carbon sink – following polymerisation, the carbon of the glucose substrate is no longer available for recycling within the cell as it is irreversibly extruded out of the cell membrane in the form of microfibrils. The production of the substrate used in the synthesis of cellulose (UDP-glucose) is catalysed by sucrose synthases (SuSy) and is a rate-limiting factor in cellulose biosynthesis. Plants possess a small SuSy gene family containing 6 to 10 members. Interestingly, SuSy proteins can occur as cytosolic as well membrane-bound entities within the cell, with the localization being controlled by the phosphorylation of a particular serine residue (Haigler *et al.*, 2001). The membrane-associated form of this enzyme may be directly channelling UDP-glucose to the site of cellulose biosynthesis (Amor *et al.*, 1995; Haigler *et al.*, 2001) and it is likely that it functions as one of the regulators of cellulose biosynthesis, particularly under stress conditions (Albrecht and Mustroph, 2003). Currently it is not known whether the membrane-bound SuSy directly interacts with the cellulose synthase complex.

Another component in the biosynthesis of cellulose in plants is the cytoskeleton, particularly the microtubules. Recent evidence indicated that the cellulose synthase complex

co-localized with the microtubules and that microtubules were required for the direction of the CSC to the cell membrane (Gardiner *et al.*, 2003). These findings are intriguing as they suggest that the CSC (which is associated with the cell membrane), is also somehow coupled to the microtubule network, which could be instrumental in the determination of the highly ordered structure of cellulose microfibrils in cell walls.

Hydrolytic enzymes also play a role in cellulose biosynthesis. A putative β-1→4 endoglucanase from *Arabidopsis* was found to be required for the biosynthesis of cellulose in the secondary cell wall (Szyjanowicz *et al.*, 2004). It is probable that this enzyme is required for the cleavage of the nascent microfibrils from CSCs, but is not directly associated with the complex itself.

Lignin biosynthesis

The secondary cell wall consists of the highly ordered cellulose microfibrils embedded in a matrix of lignin and hemicelluloses (Boudet *et al.*, 2003). In a way, lignin forms the "glue" which assists in the bonding of the cell wall components and renders the cell wall impermeable to water. However, the function of this substance is not limited to rigidity but is also implicated in the defence response of the plant to wounding. Lignin is a heterogeneous polymer, the biosynthesis of which is the end product of a well-described pathway, involving highly specialised enzymes. As reviewed by Boudet (2000), lignin biosynthesis is (energy-wise) a very expensive process with a product of great variability.

In flowering plants, the basic lignin biosynthetic pathway begins at the aromatic amino acids phenylalanine and tyrosine, and through a number of steps terminates in the formation of three different monolignols: *p*-coumaryl, coniferyl or synapyl alcohols (Anterola and Lewis, 2002). These three aromatic alcohols are then polymerised into the heteropolymer known as lignin. As much as is known about the enzymes involved in the biosynthesis of the lignin precursors, comparatively little is known about how the monolignols are polymerised

into lignin (Goujon *et al.*, 2003). It is possible that the group of enzymes known as laccases may be involved (O'Malley *et al.*, 1993), but to this day clear evidence is lacking (mainly due to the absence of *Arabidopsis* laccase mutants), suggesting redundant enzyme functions of the laccase family members. Another large enzyme family thought to play a role in the polymerisation of lignin are the peroxidases. Blee *et al.* (2003), showed that by suppressing a specific tobacco peroxidase, significantly reduced lignin deposition could be achieved. It is likely that a number of redundant enzymes control the last step of lignin biogenesis.

Lignin of gymnosperms and angiosperms differs with respect to the types of monomers it incorporates (Lewis and Yamamoto, 1990). Angiosperms utilize all three monolignols, whereas gymnosperms lack the pathway component leading to the synthesis of synapyl alcohol. One of the peculiarities of lignin, which remains to be elucidated, is the conflicting issue regarding the composition of lignin in different plants under a variety of conditions. The opposing views are centred on the supposition that lignin biosynthesis involves the incorporation of varying monomers. A study of loblolly pine in which cinnamyl alcohol dehydrogenase (CAD) - the last enzyme in monolignol biosynthesis - is depleted, showed that the mutated trees incorporated unusual monomers into the lignin polymer (Ralph *et al.*, 1997). The CAD enzyme seems essential to the lignin biosynthesis pathway, yet the plants were able to overcome the problem via other not yet determined pathways. In this study the presence of the novel monomers is explained on the basis of a “metabolic plasticity”, i.e. the plants are able to adapt to changing conditions by re-routing biosynthetic drive to other pathways.

Even though lignin is a lesser and more variable component of the cell wall, compared to cellulose, the understanding of the genes underlying its biosynthetic pathway is far more advanced than the knowledge of the cellulose biosynthesis machinery of the cell. This is primarily due to the fact that purification of the lignin biosynthetic enzymes has been

achieved (Halpin *et al.*, 1992), and that complex manipulation of the lignin biosynthetic pathway has been demonstrated in forest trees (Li *et al.*, 2003).

METHODS FOR THE DISCOVERY AND ANALYSIS OF WOOD FORMATION GENES

In a very broad sense, gene discovery and characterisation approaches can be described in terms of so-called forward and reverse genetics. "Forward genetics" usually refers to the search for a gene that underlies a specific trait of interest – such as a mutant phenotype. "Reverse genetics" applies to approaches taken to elucidate the functioning and role of a gene by its manipulation (usually the manipulation of its expression patterns). Forest biotechnology in the past decade has seen the application of both of these principles ranging from whole-genome transcript profiling to multiple gene knockdown.

Arabidopsis thaliana as a model for wood formation

Studying the molecular biology of wood formation in forest trees has traditionally been difficult due a number of factors including slow growth, large size and lack of mutant lines. Recently, *Arabidopsis thaliana* has been used for gene discovery in areas where a herbaceous species would not normally be even considered a model: such as secondary vascular development (Chaffey, 1999). Despite the intuitive prediction that herbaceous plants do not possess the capacity to produce wood (secondary xylem), evidence exists to the contrary. Lev-Yadun (1994) demonstrated that *A. thaliana* could be induced to produce wood by subjecting wild-type plants to pruning resulting in delayed flowering and senescence. This and similar findings contributed to a bout of research efforts concentrating on the use of *A. thaliana* as a model for vascular development (Zhao *et al.*, 2000; Sabatini *et al.*, 2003; Ko *et al.*, 2004).

Chaffey *et al.* (2002) demonstrated that the structural characteristics of secondary xylem in *A. thaliana* are quite similar to that seen in young poplar stems, suggesting that vascular development in *Arabidopsis* is achievable and can be used as an equivalent model for woody plants. These uses include the identification of candidate genes (by studying mutants), the study of gene function (by gene inactivation) and large scale gene expression analyses (comparing wood formation in mutants to normal plants by microarray analysis).

A. thaliana has been used to study vascular development by the development of numerous mutants which exhibit varying degrees of unusual cell-wall characteristics (Fagard *et al.*, 2000). A number of the key lignin (Jones *et al.*, 2001) and cellulose (Scheible *et al.*, 2001; Taylor *et al.*, 2003; Szyjanowicz *et al.*, 2004) biosynthetic enzymes were identified by means of *Arabidopsis* cell wall mutant studies. One of the first genes implicated in the regulation of vascular development of *A. thaliana* discovered was the *MONOPTEROS* gene as described by Hardtke and Berleth (1998). This transcription factor seems to play a major role in the vascular development in *Arabidopsis*: the *mp* mutant plants exhibit abnormal vascular development, including the disruption of the hypocotyls/ root axis, as a result of the lack of this auxin inducible transcription factor. The *MP* gene is only one of a group of regulatory proteins which play a role in the vascular development of *Arabidopsis*. Other studies initiated on the use of *A. thaliana* as a model include the approach of Zhao *et al.* (2000), in which the construction and screening of cDNA libraries from xylem and phloem led to the identification of a number of protease-encoding genes expressed in the vascular tissues. Recently, a microarray transcript profiling study of *Arabidopsis* stems undergoing secondary thickening identified clusters of genes activated during the transition from primary growth to secondary thickening (Ko *et al.*, 2004). Studies such as this would be difficult to orchestrate in woody plants, which begin secondary xylem deposition spontaneously and continuously. On the whole, the drive to identify candidate genes for wood formation in

Arabidopsis has produced valuable results. The challenge that remains is the extrapolation of the results from *Arabidopsis* to the species of interest.

Currently, the major advantage of *Arabidopsis* is that the sequence of its genome is available (Arabidopsis Genome Initiative, 2000). Annotation, both algorithmic and based on molecular studies has allowed the development of integrated internet databases such as The Arabidopsis Information Resource (TAIR, http://www.arabidopsis.org) (Huala *et al.*, 2001; Rhee *et al.*, 2003).

One of the limiting factors of using *Arabidopsis* and other herbaceous plants as model species is the fact that many genes expressed during wood formation in trees do not exhibit homology with *Arabidopsis* genes (Allona *et al.*, 1998; Sterky *et al.*, 1998). Even though it is these genes that are potentially most interesting as candidates for characterisation, the risk of investing time and resources in the analysis of genes with no homology to known proteins is substantial.

Mutant analysis

One of the classical forward genetic methods of gene discovery is the study of mutants. Much of what is known about the genes underlying the vascular development genes has been learned through the study of mutants, mostly in model plants such as *Arabidopsis*, but also in crop plants. The common approach here entails the identification (phenotypic) of a plant after germination of seed, which has been mutated (by chemical, radiation or transposon insertion methods). The mutant phenotype generally exhibits characteristics that may be of particular interest. The next stage involves the identification and characterisation of the gene the mutation of which resulted in the phenotype.

Some of the most important cell wall biosynthetic and regulatory genes were characterised based on mutant analysis. These include the secondary cell wall-associated cellulose synthase catalytic subunit genes which were identified following the analysis of a

number of *irregular xylem* (*irx*) mutants in *Arabidopsis thaliana* (Taylor *et al.*, 1999; Taylor *et al.*, 2000; Taylor *et al.*, 2003). Inactivation of a gene in a completely separate biochemical pathway can often produce mutant phenotypes with very similar characteristics. Jones *et al.* (2001) found that the *irx4* mutant suffered from greatly reduced levels of lignin due to the mutation of the cinnamoyl-CoA reductase (CCR) gene.

Mutants in crop species have also been used to identify important cell wall biosynthetic genes as is the case with the secondary cell wall-associated cellulose synthases identified as the mutated genes in the rice brittle culm mutants (Tanaka *et al.*, 2003), generated by transposon mutagenesis. Although mutant lines of forest trees are not available, random "natural" mutations in cell wall biosynthetic enzyme genes can occur as was observed by Ralph *et al.* (1997) in a loblolly pine with abnormal lignin.

It should be noted that this forward genetics approach is fairly labour-intensive and that one of the foremost issues is the identification of the phenotype. If we consider that a mutation may inactivate a gene involved in vascular development, but the gross morphological phenotype may not be noticeably affected, it would be likely that this gene would be overlooked. This is also complicated by the fact that there is often a degree of redundancy and phenotypic plasticity observed in plants. Conversely, important single copy or non redundant genes may be lethal if mutated and thus be overlooked. However, high throughput techniques are playing a role in this approach to gene discovery as well. Boyes *et al.* (2001) developed an automated technique for measuring *Arabidopsis* growth during development that has been applied to generate data which could be used to automatically identify mutants or record continuous phenotypic characteristics to be correlated with gene expression analyses.

Transcriptome analysis

A complementary approach to mutant analysis is the global analysis of gene expression by

screening large quantities of transcripts. Various techniques have been used to measure and interpret the levels of gene expression in tissues. They all share the common goal of identifying genes/ transcripts that are expressed in ways which indicates them as possible candidate genes for a specific process or function. The goal is to analyse a pool of transcripts and hopefully gain a number of candidates for further analysis. Another crucial aspect of the differential expression experiments is their design – a well contrasted pair of samples could result in the identification of genes which are, at least in part, responsible for the differences between the samples.

Expressed sequence tag analysis

With the advances in high-throughput capillary sequencing in the last decade, it is now possible to sequence thousands of clones from cDNA libraries with relative ease. The generation of Expressed Sequence Tags – short low-quality, single-pass sequences is often the first step towards analysing the types of genes expressed in particular tissues (Adams *et al.*, 1993). Until recently, little was known about the genes expressed in woody tissues. Expressed Sequence Tag (EST) projects in pine (Allona *et al.*, 1998) and poplar (Sterky *et al.*, 1998) provided the first insights into the xylem transcriptome. The focus of these early studies was to obtain a substantial representation of the transcriptome in the tissues of interest – developing vascular tissues.

EST sequencing is one of the necessary steps on the road to large-scale transcriptome profiling, particularly in species with limited sequence availability. However, this approach has limitations, particularly those associated with the redundancy of the data generated. Strongly expressed transcripts (often corresponding to constitutively expressed genes) can be represented in cDNA libraries at frequencies many times higher than rare transcripts (which are often those interesting regulatory genes). Diatchenko *et al.* (1996) introduced the technique of suppression subtractive hybridisation (SSH), which allowed the reduction of

redundancy and enrichment of rare, differentially expressed, messages in cDNA libraries. Application of this principle in *Eucalyptus* xylem cDNA library construction was demonstrated by Paux *et al.* (2004), who observed that a high proportion of a xylem SSH clones corresponded to cell wall-associated genes, with relatively low redundancy, following subtraction with young leaf RNA.

Generation of high-quality ESTs is important not only for obtaining a global view of the transcriptome but also for generating good probes for expression profiling using microarray technology.

Expression microarrays

The transcript-profiling technique that has received perhaps the most attention in recent years has been the cDNA microarray method of comparing the relative levels of gene expression of two discrete samples, as first demonstrated by Schena *et al.* (1995). Essentially, the principle involves the competitive hybridisation between two pools of contrasting fluorescently labelled cDNA and a large number of immobilised probes (usually short cDNA fragments) printed on glass slides. cDNA microarrays have been used successfully to profile gene expression during wood-formation in trees (Hertzberg *et al.*, 2001). This study focused on vascular tissues in poplar trees and described the developmental regulation of genes involved in xylogenesis. By plotting the expression patterns of lignin biosynthetic genes onto the lignin pathway, it was possible to picture the global perspective of lignification across the woody stem. Microarray studies of gene expression in gymnosperm species, such as the analysis of the *Pinus taeda* xylem transcriptome (Yang *et al.*, 2004), have yielded similar results to those seen in poplar. Recently, Paux *et al.* (2004) provided the first analysis of the genes expressed in *Eucalyptus* xylem by cDNA array analysis (using nylon membranes as support for the probes instead of glass slides). Genes associated with cell wall biosynthesis as well as auxin-dependent signalling were found to be strongly upregulated in developing xylem of

Eucalyptus. By using the same *Eucalyptus* array, Paux *et al.* (2005), identified a number of genes upregulated during tension wood development. As tension wood contains extremely high proportions of crystalline cellulose, it is not surprising that one of the most strongly induced transcripts in tension wood corresponded to a putative cellulose synthase.

The main advantages of the cDNA microarray technique lie in the reproducibility, ability to compare the results of separate experiments and relative simplicity once the system has been developed. One of the main benefits of the cDNA microarray approach is the ability to set up experimental designs, which have high levels of repetition allowing for high-accuracy inference. Conversely, microarray analysis is often fraught with difficulties which range from the technical aspects of the procedure (large RNA quantities required) to issues regarding the processing of microarray data. To this day, microarray data is often analysed using simple log-transformed ratios – often overlooking the lack of significance of data. This is despite the availability of techniques which allow the assessment of data significance in microarray experiments (Kerr *et al.*, 2000; Kerr and Churchill, 2001; Wolfinger *et al.*, 2001). Perhaps the most limiting aspect, though, is the fact that the microarray is a "closed" system. This is due to the fact that the set of probes on the microarray slide is either a subset of a cDNA library or an oligonucleotide. Hence, it is possible to test only for the expression of genes which are represented on the slide. Use of whole genome or transcriptome arrays alleviates this problem to a certain extent, but these measures are obviously not available for the vast majority of species.

Serial analysis of gene expression

One of the first approaches for studying large-scale gene expression levels was achieved by Velculescu *et al.* (1995), who coined the phrase Serial Analysis of Gene Expression (SAGE). The technique entails counting short (ca. 15 bp) tags (corresponding to specific transcripts) in cDNA libraries and hence gives an absolute level of expression of that particular tagged

transcript. The ability to determine the absolute (rather than relative) abundance of a transcript is possibly its main advantage. Some difficulty lies in the need for laborious sequencing of many 1000s of tags coupled to the possible difficulties encountered with identifying some of the tags (i.e. differentiating between gene family members). Analysis of SAGE data also depends on the presence of large well-annotated EST or genome sequence data. Lorenz and Dean (2002) generated a very large set of SAGE tags from loblolly pine xylem representing up to 40,000 different genes, many of which exhibited similarity to known wood formation genes. With the lack of a comprehensively-annotated sequence resource for *Pinus taeda*, however, the data generated in this study is yet to attain full significance.

Massively parallel signature sequencing

Recently, Brenner *et al.* (2000) developed a related technique to SAGE, termed Massively Parallel Signature Sequencing (MPSS) which uses a combination of bead array technology with sequencing of 20 bp cDNA tags. This approach is a technological advancement of SAGE and is capable of analysing millions of tags in a comparatively short time. MPSS was applied to the expression analysis of the maize cellulose synthase gene family (Appenzeller *et al.*, 2004). The major shortcoming of this approach is the requirement of specialized equipment and the generally high cost. As such, MPSS is more suited to large corporations and consortia than academic research groups.

cDNA-AFLP

Bachem *et al.* (1996) developed a transcript profiling method based on the Amplified Fragment Length Polymorphism technique (Vos *et al.*, 1995). The cDNA-AFLP technique was originally developed to obtain an expression profile fingerprint, but has since been adapted for various purposes, most particularly for novel gene discovery.

Briefly, mRNA is isolated from the tissues of choice and is converted to cDNA. The

cDNA is then digested with two different restriction endonucleases after which adapters are ligated to the cleaved ends. Amplification of the adaptor-ligated cDNA with primers specific to the adaptors (pre-amplification) follows. The next step involves amplification using adapter specific primers with 2-3 bp extensions at the 3' end. The selective nucleotides ensure that only a small subset of the pre-amplified fragments are amplified during the selective amplification. The selective amplicons are then resolved on a polyacrylamide gel where it becomes possible to identify differential expression merely by the presence or absence of a band in a specific location. These, termed Transcript Derived Fragments (TDFs), can be detected by radioactive nuclide incorporation (Bachem *et al.*, 1998) or silver-staining of the gel (Dubos and Plomion, 2003). With the recent development of an AFLP technique using infrared detection (Myburg *et al.*, 2001), it became possible to apply fluorescence detection to cDNA-AFLP profiling (Figure 1).

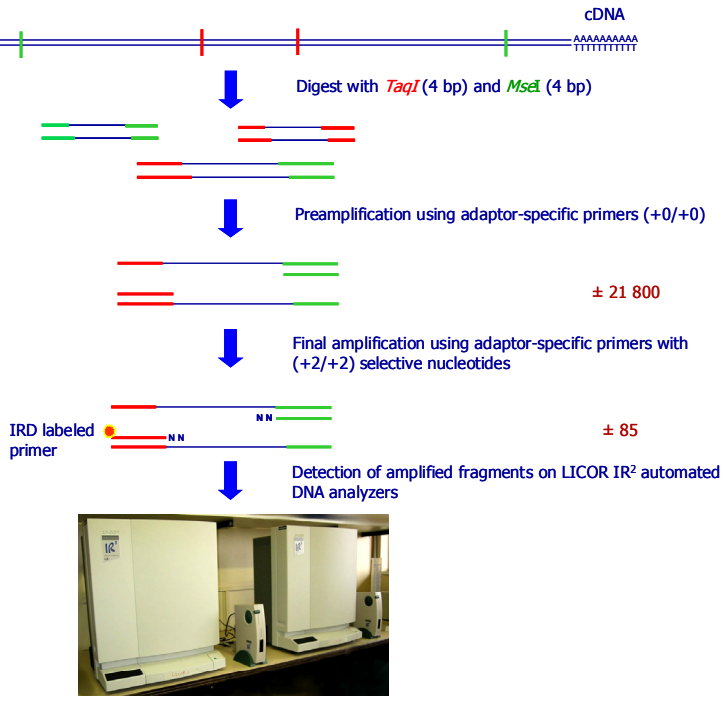


Figure 1. cDNA-AFLP profiling using LI-COR IR^2 infrared DNA analysers. The approximate number of fragments generated after the pre-amplification and selective amplification steps are indicated in red.

In essence, the cDNA-AFLP procedure using selective primers is designed to reduce the complexity of the transcript profile to an extent where the number of TDFs is sufficiently small permitting resolution by polyacrylamide gel electrophoresis (PAGE). TDFs of specific interest (e.g. TDFs upregulated in a specific tissue) can be excised from the PAGE gels and further characterised by re-amplification, cloning and sequencing (Bachem *et al.*, 1996).

cDNA-AFLP is particularly useful in systems which are poorly characterised or for which there is limited available DNA sequence. Specific examples of cDNA-AFLP applied to transcript profiling in tree species include apple (Jensen *et al.*, 2003), pine (Dubos and Plomion, 2003) and almond (Campalans *et al.*, 2001). Even though cDNA-AFLP is yet to be applied to the transcriptome analysis of woody tissues in forest trees, it has been used for the analysis of gene expression during the transdifferentiation of *Zinnia* mesophyll cells into tracheary elements (Milioni *et al.*, 2001; Milioni *et al.*, 2002). These studies generated a number of interesting TDFs with homology to important cell wall biosynthesis genes including cellulose synthase and laccase as well as putative regulatory genes such as the auxin efflux carrier *PIN1* and the auxin responsive transcription factor *MONOPTEROS*.

In addition to enabling the isolation of TDFs with interesting expression profiles, cDNA-AFLP is also capable of accurately quantifying the relative abundance of thousands of transcripts. Breyne *et al.* (2003) have successfully demonstrated this application of cDNA-AFLP in the quantitative analysis of gene expression levels in tobacco with results which clearly indicated that cDNA-AFLP is on par with typical microarray experimental results. Parallel analysis of gene expression in yeast using two hybridisation techniques (cDNA microarray and GeneChip) and cDNA-AFLP revealed that this PCR-based technique was a good alternative to the array-based approaches (Reijans *et al.*, 2003). One problem associated with the use of cDNA-AFLP for genome-wide expression analysis lies in the fact that the

identity of individual TDFs is not known until they are excised and further characterised. However, cDNA-AFLP expression profiling may be useful as a starting point, or in cases where an estimate of the overall transcript variability is required (such as divergent tissues in a poorly understood organism).

The major advantage of cDNA-AFLP lies in its ability to act as an "open" system: rather than using a set of probes to which gene expression is compared, the two samples are directly contrasted. Additionally, there is an almost unlimited number of enzyme-primer combinations to explore, permitting excessive customisation. Another advantage of cDNA-AFLP over hybridisation techniques is the ability to detect rare transcripts (due to its PCR nature) which might not be assayed with hybridisation-based technologies (Reijans *et al.*, 2003).

CONCLUSIONS AND FUTURE PROSPECTS

Recent years have seen rapid advances in the molecular biology of wood formation and forest biotechnology in general. Many of the key enzymes which synthesize the major components of the cell wall have been characterised and understanding of the genetic regulation of wood development has progressed significantly. The public availability of the genome sequence of forest trees such as *Populus* and *Eucalyptus* will enable the rapid isolation of genes orthologous to those characterised previously in model plants and provide the opportunity of characterising them in a woody species.

One of the great remaining challenges in cell wall molecular biology, which has evaded scientists for nearly a decade since the discovery of the first plant cellulose synthases, is the elucidation of the cellulose biosynthetic complex in plants. A large proportion of tree mass is made up by cellulose, yet a majority of the advances in understanding cellulose biosynthesis research have been made in herbaceous systems. Functional characterisation of

cellulose biosynthetic genes from trees can be used to answer the simple question which has arisen recently: if trees and weeds seem to possess the same cellulose biosynthetic machinery, why is it that a single tree can produce more cellulose than an entire forest of *Arabidopsis*? Whether the difference is at the developmental, regulatory or structural level, answering this and related questions will enhance our ability to produce improved trees.

Another implication of the genome sequence becoming available, will the ability to perform much more accurate and informative transcript profiling studies. It will be possible to generate whole-genome microarrays and use these to focus on specific aspects of wood formation, rather than the general studies performed to date. Continuing annotation of the genome sequence will afford hypothetical functions to more of the currently uncharacterised ESTs. The availability of technologies such as laser capture microdissection (Asano *et al.*, 2002), coupled to whole-genome transcript profiling could possibly enable the identification of specific genes expressed in the key vascular cambium cell layer as well as the other major tissue layers and cell types in developing wood.

As much as the understanding of wood formation is advancing, the current ability to transform some of the major commercially important forest trees such as *Eucalyptus* and *Pinus* species is lagging significantly behind. This field of forest biotechnology will probably have to undergo drastic advances before the understanding of wood development genes can be put to significant use. Even more disconcerting is the current general public opinion of genetically modified organisms which will probably hinder efforts to produce and commercialise transgenic trees. It is possible that instead of widespread planting of transgenic trees, the future may bring advances in *in vitro* “designer fibre”. However, it will first be necessary to further the understanding of the molecular genetics and biochemistry of wood formation specifically in forest trees, which is currently in its relative infancy.

REFERENCES

Adams, M.D., Soares, M.B., Kerlavage, A.R., Fields, C. and Venter, J.C. 1993. Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. Nat Genet 4: 373-380.

Albert, M., Werner, M., Proksch, P., Fry, S. and Kaldenhoff, R. 2004. The cell wall-modifying xyloglucan endotransglycosylase/hydrolase LeXTH1 is expressed during the defence reaction of tomato against the plant parasite *Cuscuta reflexa*. Plant Biol 6: 402-407.

Albrecht, G. and Mustroph, A. 2003. Localization of sucrose synthase in wheat roots: increased in situ activity of sucrose synthase correlates with cell wall thickening by cellulose deposition under hypoxia. Planta 217: 252-260.

Allona, I., Quinn, M., Shoop, E., Swope, K., St. Cyr, S., Carlis, J., Riedl, J., Retzel, E., Campbell, M.M., Sederoff, R. and Whetten, R.W. 1998. Analysis of xylem formation in pine by cDNA sequencing. Proc Natl Acad Sci U S A 95: 9693-9698.

Amor, Y., Haigler, C.H., Johnson, S., Wainscott, M. and Delmer, D.P. 1995. A membrane-associated form of sucrose synthase and its potential role in synthesis of cellulose and callose in plants. Proc Natl Acad Sci U S A 92: 9353-9357.

Anterola, A.M. and Lewis, N.G. 2002. Trends in lignin modification: a comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity. Phytochemistry 61: 221-294.

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

Arabidopsis_Genome_Initiative 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796-815.

Asano, T., Masumura, T., Kusano, H., Kikuchi, S., Kurita, A., Shimada, H. and Kadowaki, K. 2002. Construction of a specialized cDNA library from plant cells isolated by laser capture microdissection: toward comprehensive analysis of the genes expressed in the rice phloem. Plant J 32: 401-408.

Bachem, C.W.B., Oomen, R.J.F.J. and Visser, R.G.F. 1998. Transcript imaging with cDNA-AFLP: a step-by-step protocol. Plant Mol Biol Rep 16: 157-173.

Bachem, C.W.B., van der Hoeven, R.S., de Bruijn, S.M., Vreugdenhil, D., Zabeau, M. and Visser, R.G.F. 1996. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. Plant J 9: 745-753.

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

Barnett, J. and Bonham, V. 2004. Cellulose microfibril angle in the cell wall of wood fibres. Biol Rev Camb Philos Soc 79: 461-472.

Bhalerao, R., Nilsson, O. and Sandberg, G. 2003. Out of the woods: forest biotechnology enters the genomic era. Curr Opin Biotechnol 14: 206-213.

Blee, K.A., Choi, J.W., O'Connell, A.P., Schuch, W., Lewis, N.G. and Bolwell, G.P. 2003. A lignin-specific peroxidase in tobacco whose antisense suppression leads to vascular tissue modification. Phytochemistry 64: 163-176.

Bonke, M., Thitamadee, S., Mahonen, A.P., Hauser, M.T. and Helariutta, Y. 2003. APL regulates vascular tissue identity in *Arabidopsis*. Nature 426: 181-186.

Boudet, A.M. 2000. Lignins and lignification: Selected issues. Plant Physiol Biochem 38: 81-96.

Boudet, A.M., Kajita, S., Grima-Pettenati, J. and Goffner, D. 2003. Lignins and lignocellulosics: a better control of synthesis for new and improved uses. Trends Plant Sci 8: 576-581.

Boyes, D.C., Zayed, A.M., Ascenzi, R., McCaskill, A.J., Hoffman, N.E., Davis, K.R. and Gorlach, J. 2001. Growth stage-based phenotypic analysis of *Arabidopsis*: a model for high throughput functional genomics in plants. Plant Cell 13: 1499-1510.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S.R., Moon, K., Burcham, T., Pallas, M., DuBridge, R.B., Kirchner, J., Fearon, K., Mao, J. and Corcoran, K. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol 18: 630-634.

Breyne, P., Dreesen, R., Cannoot, B., Rombaut, D., Vandepoele, K., Rombauts, S., Vanderhaeghen, R., Inze, D. and Zabeau, M. 2003. Quantitative cDNA-AFLP analysis for genome-wide expression studies. Mol Genet Genomics 269: 173-179.

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

Campalans, A., Pages, M. and Messeguer, R. 2001. Identification of differentially expressed genes by the cDNA-AFLP technique during dehydration of almond (*Prunus amygdalus*). Tree Physiol 21: 633-643.

Campbell, M.M., Brunner, A.M., Jones, H.M. and Strauss, S.H. 2003. Forestry's fertile crescent: the application of biotechnology to forest trees. Plant Biotech J 1: 141-154.

Chaffey, N. 1999. Wood formation in forest trees: from *Arabidopsis* to *Zinnia*. Trends Plant Sci 4: 203-204.

Chaffey, N. and Barlow, P. 2002. Myosin, microtubules, and microfilaments: co-operation between cytoskeletal components during cambial cell division and secondary vascular differentiation in trees. Planta 214: 526-536.

Chaffey, N., Cholewa, E., Regan, S. and Sundberg, B. 2002. Secondary xylem development in *Arabidopsis*: a model for wood formation. Physiol Plant 114: 594-600.

Civello, P.M., Powell, A.L., Sabehat, A. and Bennett, A.B. 1999. An expansin gene expressed in ripening strawberry fruit. Plant Physiol 121: 1273-1280.

Cosgrove, D.J. 2000. Loosening of plant cell walls by expansins. Nature 407: 321-326.

Cseke, L.J., Zheng, J. and Podila, G.K. 2003. Characterization of PTM5 in aspen trees: a MADS-box gene expressed during woody vascular development. Gene 318: 55-67.

Delmer, D.P. 1999. Cellulose biosynthesis: exciting times for a difficult field of study. Annu Rev Plant Physiol Plant Mol Biol 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. and Fukuda, H. 2002. Visualization by comprehensive microarray analysis of gene expression programs during transdifferentiation of mesophyll cells into xylem cells. Proc Natl Acad Sci U S A 99: 15794-15799.

Dharmasiri, N., Dharmasiri, S. and Estelle, M. 2005. The F-box protein TIR1 is an auxin receptor. Nature. 435:441-445.

Diatchenko, L., Lau, Y.F.C., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D. and Siebert, P.D. 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci U S A 93: 6025-6030.

Doblin, M.S., Kurek, I., Jacob-Wilk, D. and Delmer, D.P. 2002. Cellulose biosynthesis in plants: from genes to rosettes. Plant Cell Physiol 43: 1407-1420.

Dubos, C. and Plomion, C. 2003. Identification of water-deficit responsive genes in maritime pine (*Pinus pinaster* Ait.) roots. Plant Mol Biol 51: 249-262.

Fagard, M., Hofte, H. and Vernhettes, S. 2000. Cell wall mutants. Plant Physiol Biochem 38: 15-25.

Fenning, T.M. and Gershenzon, J. 2002. Where will the wood come from? Plantation forests and the role of biotechnology. Trends Biotechnol 20: 291-296.

Friml, J. 2003. Auxin transport - shaping the plant. Curr Opin Plant Biol 6: 7-12.

Fry, S.C., Smith, R.C., Renwick, K.F., Martin, D.J., Hodge, S.K. and Matthews, K.J. 1992. Xyloglucan endotransglycosylase, a new wall-loosening enzyme activity from plants. Biochem J 282: 821-828.

Gardiner, J.C., Taylor, N.G. and Turner, S.R. 2003. Control of cellulose synthase complex localization in developing xylem. Plant Cell 15: 1740-1748.

Goujon, T., Sibout, R., Eudes, A., MacKay, J. and Joulanin, L. 2003. Genes involved in the biosynthesis of lignin precursors in *Arabidopsis thaliana*. Plant Physiol Biochem 41: 677-687.

Gray, W.M., Kepinski, S., Rouse, D., Leyser, O. and Estelle, M. 2001. Auxin regulates SCFTIR1-dependent degradation of AUX/IAA proteins. Nature 414: 271-276.

Gray-Mitsumune, M., Mellerowicz, E.J., Abe, H., Schrader, J., Winzell, A., Sterky, F., Blomqvist, K., McQueen-Mason, S., Teeri, T.T. and Sundberg, B. 2004. Expansins abundant in secondary xylem belong to subgroup A of the α-expansin gene family. Plant Physiol 135: 1552-1564.

Haigler, C.H., Ivanova-Datcheva, M., Hogan, P.S., Salnikov, V.V., Hwang, S., Martin, K. and Delmer, D.P. 2001. Carbon partitioning to cellulose synthesis. Plant Mol Biol 47: 29-51.

Halpin, C., Knight, M.E., Grima-Pettenati, J., Goffner, D., Boudet, A. and Schuch, W. 1992. Purification and characterization of cinnamyl alcohol-dehydrogenase from tobacco stems. Plant Physiol 98: 12-16.

Hardtke, C.S. and Berleth, T. 1998. The Arabidopsis gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development. EMBO J 17: 1405-1411.

Henriksson, H., Denman, S.E., Campuzano, I.D., Ademark, P., Master, E.R., Teeri, T.T. and Brumer, H. 2003. N-linked glycosylation of native and recombinant cauliflower xyloglucan endotransglycosylase 16A. Biochem J 375: 61-73.

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

Hiwasa, K., Rose, J.K., Nakano, R., Inaba, A. and Kubo, Y. 2003. Differential expression of seven α-expansin genes during growth and ripening of pear fruit. Physiol Plant 117: 564-572.

Hu, W.J., Harding, S.A., Lung, J., Popko, J.L., Ralph, J., Stokke, D.D., Tsai, C.J. and Chiang, V.L. 1999. Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. Nat Biotechnol 17: 808-812.

Huala, E., Dickerman, A.W., Garcia-Hernandez, M., Weems, D., Reiser, L., Lafond, F., Hanley, D., Kiphart, D., Zhuang, M.Z., Huang, W., Mueller, L.A., Bhattacharyya, D., Bhaya, D., Sobral, B.W., Beavis, W., Meinke, D.W., Town, C.D., Somerville, C. and

Rhee, S.Y. 2001. The *Arabidopsis* Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Res 29: 102-105.

Im, K.H., Cosgrove, D.J. and Jones, A.M. 2000. Subcellular localization of expansin mRNA in xylem cells. Plant Physiol 123: 463-470.

Jensen, P.J., Rytter, J., Detwiler, E.A., Travis, J.W. and McNellis, T.W. 2003. Rootstock effects on gene expression patterns in apple tree scions. Plant Mol Biol 53: 493-511.

Johansson, P., Brumer, H., Baumann, M.J., Kallas, A.M., Henriksson, H., Denman, S.E., Teeri, T.T. and Jones, T.A. 2004. Crystal structures of a poplar xyloglucan endotransglycosylase reveal details of transglycosylation acceptor binding. Plant Cell 16: 874-886.

Jones, L., Ennos, A.R. and Turner, S.R. 2001. Cloning and characterization of *irregular xylem4* (*irx4*): a severely lignin-deficient mutant of *Arabidopsis*. Plant J 26: 205-216.

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

Kang, J. and Dengler, N. 2004. Vein pattern development in adult leaves of *Arabidopsis thaliana*. Int J Plant Sci 165: 231-242.

Kerr, M.K. and Churchill, G.A. 2001. Experimental design for gene expression microarrays. Biostatistics 2: 183-201.

Kerr, M.K., Martin, M. and Churchill, G.A. 2000. Analysis of variance for gene expression microarray data. J Comput Biol 7: 819-837.

Kibblewhite, R.P. 1999. Designer fibres for improved papers through exploiting genetic variation in wood microstructure. Appita J 52: 429-433.

Ko, J.H., Han, K.H., Park, S. and Yang, J.M. 2004. Plant body weight-induced secondary growth in *Arabidopsis* and its transcription phenotype revealed by whole-transcriptome profiling. Plant Physiol 135: 1069-1083.

Koornneef, M. and Stam, P. 2001. Changing paradigms in plant breeding. Plant Physiol 125: 156-159.

Kuriyama, H. and Fukuda, H. 2002. Developmental programmed cell death in plants. Curr Opin Plant Biol 5: 568-573.

Lachaud, S., Catesson, A.M. and Bonnemain, J.L. 1999. Structure and functions of the vascular cambium. C R Acad Sci III 322: 633-650.

Larson, P.R. 1994 The vascular cambium. Development and structure. Springer-Verlag, Berlin.

Lev-Yadun, S. 1994. Induction of sclereid differentiation in the pith of *Arabidopsis thaliana* (L.) Heynh. J Exp Bot 45: 1845-1849.

Lewis, N.G. and Yamamoto, E. 1990. Lignin: occurrence, biogenesis and biodegradation. Annu Rev Plant Physiol Plant Mol Biol 41: 455-496.

Li, L., Zhou, Y., Cheng, X., Sun, J., Marita, J.M., Ralph, J. and Chiang, V.L. 2003. Combinatorial modification of multiple lignin traits in trees through multigene cotransformation. Proc Natl Acad Sci U S A 100: 4939-4944.

Little, C.H.A., MacDonald, J.E. and Olsson, O. 2002. Involvement of indole-3-acetic acid in fascicular and interfascicular cambial growth and interfascicular extraxylary fiber differentiation in *Arabidopsis thaliana* inflorescence stems. Int J Plant Sci 163: 519-529.

Lorenz, W.W. and Dean, J.F. 2002. SAGE profiling and demonstration of differential gene expression along the axial developmental gradient of lignifying xylem in loblolly pine (*Pinus taeda*). Tree Physiol 22: 301-310.

Lu, S., Shi, R., Tsao, C.C., Yi, X., Li, L. and Chiang, V.L. 2004. RNA silencing in plants by the expression of siRNA duplexes. Nucleic Acids Res 32: e171.

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

Milioni, D., Sado, P.E., Stacey, N.J., Domingo, C., Roberts, K. and McCann, M.C. 2001. Differential expression of cell-wall-related genes during the formation of tracheary elements in the *Zinnia* mesophyll cell system. Plant Mol Biol 47: 221-238.

Milioni, D., Sado, P.E., Stacey, N.J., Roberts, K. and McCann, M.C. 2002. Early gene expression associated with the commitment and differentiation of a plant tracheary element is revealed by cDNA-amplified fragment length polymorphism analysis. Plant Cell 14: 2813-2824.

Myburg, A.A., Remington, D.L., O'Malley, D.M., Sederoff, R.R. and Whetten, R.W. 2001. High-throughput AFLP analysis using infrared dye-labeled primers and an automated DNA sequencer. BioTechniques 30: 348-357.

Nicol, F., His, I., Jauneau, A., Vernhettes, S., Canut, H. and Hofte, H. 1998. A plasma membrane-bound putative endo-1,4-β-D-glucanase is required for normal wall assembly and cell elongation in *Arabidopsis*. EMBO J 17: 5563-5576.

O'Malley, D.M., Whetten, R., Bao, W.L., Chen, C.L. and Sederoff, R.R. 1993. The role of laccase in lignification. Plant J 4: 751-757.

Paux, E., Carocha, V., Marques, C., Mendes de Sousa, A., Borralho, N., Sivadon, P. and Grima-Pettenati, J. 2005. Transcript profiling of *Eucalyptus* xylem genes during tension wood formation. New Phytol In Press.

Paux, E., Tamasloukht, M., Ladouce, N., Sivadon, P. and Grima-Pettenati, J. 2004. Identification of genes preferentially expressed during wood formation in *Eucalyptus*. Plant Mol Biol 55: 263-280.

Pear, J.R., Kawagoe, Y., Schreckengost, W.E., Delmer, D.P. and Stalker, D.M. 1996. Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. Proc Natl Acad Sci U S A 93: 12637-12642.

Peng, L., Kawagoe, Y., Hogan, P. and Delmer, D. 2002. Sitosterol-β-glucoside as primer for cellulose synthesis in plants. Science 295: 147-150.

Pezzotti, M., Feron, R. and Mariani, C. 2002. Pollination modulates expression of the PPAL gene, a pistil-specific β-expansin. Plant Mol Biol 49: 187-97.

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

Ralph, J., MacKay, J.J., Hatfield, R.D., O'Malley, D.M., Whetten, R.W. and Sederoff, R.R. 1997. Abnormal lignin in a loblolly pine mutant. Science 277: 235-239.

Reijans, M., Lascaris, R., Groeneger, A.O., Wittenberg, A., Wesselink, E., van Oeveren, J., de Wit, E., Boorsma, A., Voetdijk, B., van der Spek, H., Grivell, L.A. and Simons, G. 2003. Quantitative comparison of cDNA-AFLP, microarrays, and genechip expression data in *Saccharomyces cerevisiae*. Genomics 82: 606-618.

Rhee, S.Y., Beavis, W., Berardini, T.Z., Chen, G.H., Dixon, D., Doyle, A., Garcia-Hernandez, M., Huala, E., Lander, G., Montoya, M., Miller, N., Mueller, L.A., Mundodi, S., Reiser, L., Tacklind, J., Weems, D.C., Wu, Y.H., Xu, I., Yoo, D., Yoon, J. and Zhang, P.F. 2003. The *Arabidopsis* Information Resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. Nucleic Acids Res 31: 224-228.

Richmond, T. 2000. Higher plant cellulose synthases. Genome Biol 1: 1-6.

Ross, P., Mayer, R. and Benziman, M. 1991. Cellulose biosynthesis and function in bacteria. Microbiol Rev 55: 35-58.

Sabatini, S., Heidstra, R., Wildwater, M. and Scheres, B. 2003. SCARECROW is involved in positioning the stem cell niche in the *Arabidopsis* root meristem. Genes Dev 17: 354-358.

Saxena, I.M., Brown, R.M. Jr., Fevre, M., Geremia, R.A. and Henrissat, B. 1995. Multidomain architecture of β-glycosyl transferases: implications for mechanism of action. J Bacteriol 177: 1419-1424.

Scheible, W.R., Eshed, R., Richmond, T., Delmer, D. and Somerville, C. 2001. Modifications of cellulose synthase confer resistance to isoxaben and thiazolidinone herbicides in *Arabidopsis ixr1* mutants. Proc Natl Acad Sci U S A 98: 10079-10084.

Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270: 467-470.

Schrader, J., Baba, K., May, S.T., Palme, K., Bennett, M., Bhalerao, R.P. and Sandberg, G. 2003. Polar auxin transport in the wood-forming tissues of hybrid aspen is under simultaneous control of developmental and environmental signals. Proc Natl Acad Sci U S A 100: 10096-10101.

Schroder, R., Wegrzyn, T.F., Bolitho, K.M. and Redgwell, R.J. 2004. Mannan transglycosylase: a novel enzyme activity in cell walls of higher plants. Planta 219: 590-600.

Sterky, F., Regan, S., Karlsson, J., Hertzberg, M., Rohde, A., Holmberg, A., Amini, B., Bhalerao, R., Larsson, M., Villarroel, R., Van Montagu, M., Sandberg, G., Olsson, O., Teeri, T.T., Boerjan, W., Gustafsson, P., Uhlen, M., Sundberg, B. and Lundeberg, J. 1998. Gene discovery in the wood-forming tissues of poplar: Analysis of 5,692 expressed sequence tags. Proc Natl Acad Sci U S A 95: 13330-13335.

Strauss, S.H., Howe, G.T. and Goldfarb, B. 1991. Prospects for genetic engineering of insect resistance in forest trees. For Ecol Manage 43: 181-209.

Szyjanowicz, P.M.J., McKinnon, I., Taylor, N.G., Gardiner, J., Jarvis, M.C. and Turner, S.R. 2004. The *irregular xylem 2* mutant is an allele of *korrigan* that affects the secondary cell wall of *Arabidopsis thaliana*. Plant J 37: 730-740.

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

Taylor, N.G., Howells, R.M., Huttly, A.K., Vickers, K. and Turner, S.R. 2003. Interactions among three distinct CesA proteins essential for cellulose synthesis. Proc Natl Acad Sci U S A 100: 1450-1455.

Taylor, N.G., Laurie, S. and Turner, S.R. 2000. Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in *Arabidopsis*. Plant Cell 12: 2529-2539.

Taylor, N.G., Scheible, W.R., Cutler, S., Somerville, C.R. and Turner, S.R. 1999. The *irregular xylem 3* locus of *Arabidopsis* encodes a cellulose synthase required for secondary cell wall synthesis. Plant Cell 11: 769-779.

Van Raemdonck, D., Jaziri, M., Boerjan, W. and Baucher, M. 2001. Advances in the improvement of forest trees through biotechnology. Belg J Bot 134: 64-78.

Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. 1995. Serial analysis of gene expression. Science 270: 484-487.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. 1995. AFLP - A new technique for DNA-fingerprinting. Nucleic Acids Res 23: 4407-4414.

Wolfinger, R.D., Gibson, G., Wolfinger, E.D., Bennett, L., Hamadeh, H., Bushel, P., Afshari, C. and Paules, R.S. 2001. Assessing gene significance from cDNA microarray expression data via mixed models. J Comput Biol 8: 625-637.

Wu, R.L., Hu, X.S. and Han, Y.F. 2000. Molecular genetics and developmental physiology: implications for designing better forest crops. CRC Crit Rev Plant Sci 19: 377-393.

Wullschleger, S.D., Tuskan, G.A. and Difazio, S.P. 2002. Genomics and the tree physiologist. Tree Physiol 22: 1273-1276.

Yang, S.H., van Zyl, L., No, E.G. and Loopstra, C.A. 2004. Microarray analysis of genes preferentially expressed in differentiating xylem of loblolly pine (*Pinus taeda*). Plant Sci 166: 1185-1195.

Zhao, C.S., Johnson, B.J., Kositsup, B. and Beers, E.P. 2000. Exploiting secondary growth in *Arabidopsis*. Construction of xylem and bark cDNA libraries and cloning of three xylem endopeptidases. Plant Physiol 123: 1185-1196.

CHAPTER 2

WITHIN-TREE TRANSCRIPTOME PROFILING IN WOOD-FORMING TISSUES OF A FAST-GROWING EUCALYPTUS TREE

Martin Ranik, Nicky M. Creux and Alexander A. Myburg

Accepted for publication: Tree Physiology (2005)

ABSTRACT

The biochemical and morphological complexity of wood is the result of the coordinated, tissue-specific expression of a large number of genes. Despite the availability of high-throughput transcript profiling technology, little is known about tissue-specific gene expression patterns in the wood-forming tissues of *Eucalyptus* plantation tree species. We used cDNA-amplified fragment length polymorphism (AFLP) analysis in combination with infrared fragment detection and semi-automated band quantification to profile gene expression in a fast-growing, commercial *Eucalyptus* tree. The expression profiles of 6385 transcript-derived fragments (TDFs) were analysed across four major woody tissues (mature xylem, immature xylem, phloem and cork) collected from two positions along the stem of a six-year-old tree. This provided a global view of transcript abundance and variability in the mature *Eucalyptus* stem. Approximately 21% of the TDFs were differentially expressed and could be grouped into clusters representing co-expressed genes. A total of 71 TDFs representing different gene clusters were isolated and characterized. We found that genes implicated in cell fate, signal transduction and cell wall biosynthesis, processes closely associated with xylogenesis, were significantly upregulated in differentiating xylem tissues. Analysis of the expression levels of selected TDF using quantitative RT-PCR corroborated the TDF quantification and confirmed that cDNA-AFLP analysis is a highly efficient and accurate tool for transcript profiling and gene discovery in wood-forming tissues of forest tree species.

INTRODUCTION

Fast-growing hybrids of *Eucalyptus* tree species are some of the most efficient producers of wood fiber and cellulose on earth. Clonal plantations of these hybrid genotypes can produce

up to 40 m^3 of wood per hectare per year in subtropical areas of countries such as Brazil and South Africa (FAO, http://apps.fao.org). Such high rates of biomass production are supported by a very efficient cell proliferation and differentiation system originating in the lateral meristem (i.e. vascular cambium) of these trees. The morphology and chemical composition of mature xylem cell walls are important factors that affect pulp and paper processing (Rudie, 1998), as well as the quality of solid wood products derived from these trees (Evans and Ilic, 2001; Barnett and Bonham, 2004). Currently, our understanding of the genes that determine the size, shape and chemical makeup of xylem cells lags behind our knowledge of wood biochemistry and anatomy. It has already been shown in tree species such as pine and poplar that a large number of genes are differentially expressed in the relatively confined region of developing tissues surrounding the vascular cambium (Hertzberg *et al.*, 2001; Israelsson *et al.*, 2003; Yang *et al.*, 2004b). This variability constitutes a potentially rich source of novel genes underlying the process of wood formation.

Expressed sequence tag (EST) sequencing has proved to be an efficient approach to obtain information about the types of genes and identify novel genes expressed during wood formation. For example, 10% of approximately 3700 unique ESTs obtained from poplar xylem (Sterky *et al.*, 1998), showed no similarity to any known sequence and were thought to represent novel genes involved in xylem differentiation. Using a similar approach in pine, Allona *et al.* (1998) found significant representation of cellulose, lignin and other cell-wall biosynthesis genes and a comparable percentage of ESTs with no suggested function.

Interestingly, the herbaceous model plant *Arabidopsis thaliana* has also proven to be a useful model for xylem development, as it can be induced to form secondary xylem (Lev-Yadun, 1994; Zhao *et al.*, 2000; Chaffey *et al.*, 2002). Expressed wood formation genes appear to show high functional conservation across plant genera as diverse as *Arabidopsis* and *Pinus*. This is illustrated by a recent finding that up to 90% of genes expressed in loblolly

pine appear to have homologs in *Arabidopsis* (Kirst *et al.* 2003). A detailed comparison of the *Arabidopsis* genome and the recently completed poplar tree genome may help to further clarify the conservation of wood formation genes and may allow the identification of genes that are unique to the genomes of large woody plants.

The establishment and sequencing of large cDNA libraries derived from vascular tissues have allowed the construction of cDNA microarray chips, which could be used to study gene expression during xylogenesis in woody and non-woody plants. In a milestone experiment, Hertzberg *et al.* (2001) used microarray analysis to quantitatively analyse transcript levels in differentiating woody tissues of poplar and described the expression patterns of a large number of wood formation genes. Further evidence of the identity of genes involved in xylogenesis has been gained from microarray analysis of gene expression (Demura *et al.* 2002) during the transdifferentiation of tracheary elements in the *Zinnia elegans* model (Fukuda and Komamine, 1980). Demura *et al.* (2002) used cDNA microarrays to identify clusters of *Z. elegans* genes that punctuate the major morphological and biochemical events of the transdifferentiation process. Microarray analysis is now increasingly being used to study gene expression during xylogenesis in woody and herbaceous species and is providing insight into the regulatory and structural pathways leading to wood formation (Israelsson *et al.*, 2003; Ko *et al.*, 2004; Yang *et al.*, 2004a; Yang *et al.*, 2004b).

Large-scale gene discovery approaches such as EST sequencing and subsequent microarray analysis require a great initial outlay of costs and labour. In contrast, cDNA-AFLP analysis (Bachem *et al.*, 1996) is an RT-PCR-based technique that allows high-throughput transcript profiling in virtually any gene expression system without the need for prior library construction or sequencing. Recent transcriptome studies in species with limited genomic resources available reflect this advantage (Jensen *et al.*, 2003; Ko *et al.*, 2003).

cDNA-AFLP analysis generates transcript profiles from RNA samples by assaying the abundance of transcript-derived cDNA fragments (TDFs) using polyacrylamide gel electrophoresis. Identities are assigned to selected TDFs by fragment isolation, sequencing and homology searches. PCR amplification of cDNA-AFLP fragments with mostly unique sizes helps to overcome the under-representation of rare messages and redundancy of highly expressed messages, limitations that are often associated with EST sequencing. If combined with dedicated image analysis software, cDNA-AFLP analysis can be used for accurate and reproducible quantification of relative expression levels, on par with cDNA and oligonucleotide microarray analysis (Reijans *et al.*, 2003). Finally, microarray transcript profiling is a closed system: transcripts not represented on the array are not quantified, while cDNA-AFLP analysis represents an open gene discovery system with an essentially unlimited number of primer-enzyme combinations to be assayed (Breyne *et al.*, 2003). The technique has been used very successfully to profile gene expression in the *Zinnia* model system and to discover novel genes involved in tracheary element formation (Milioni *et al.*, 2001; Milioni *et al.*, 2002).

The high-throughput identification and partial isolation of cDNA-AFLP fragments with interesting expression patterns enables a targeted approach to characterize novel genes underlying different aspects of wood development. The aim of this study was to identify transcripts that show tissue-specific expression patterns across the vascular cambium of a fast-growing *Eucalyptus* plantation tree. Accordingly, we used cDNA-AFLP analysis to profile more than 6300 TDFs in woody tissues of a fast-growing *Eucalyptus* tree and identified approximately 1300 TDFs with tissue-specific expression patterns, thus obtaining an overview of the gene expression pattern in the woody stem of *Eucalyptus*.

MATERIALS AND METHODS

Plant materials

Tissue samples were collected from one 6-year old ramet of a *E. grandis* x *E. nitens* hybrid commercial clone (NH0000, Mondi Business Paper South Africa). Immediately after felling of the tree, two stem sections, called Lower Half (L, 2 to 5 m) and Upper Half (U, 7 to 10 m) were marked on the main stem. Each stem section was progressively debarked and the following tissue layers were immediately collected into liquid nitrogen: **Immature Xylem (IX)**, the *circa* 2 mm outer glutinous layer coating the main stem following the removal of the bark (encompassing mostly xylem mother cells and some cambial initials); **Mature Xylem (X)**, after complete removal of the IX layer, the next 2-3 mm wet, fibrous layer containing secondary xylem tissue in varying stages of maturity; **Phloem (P)**: the 2-5 mm layer adhering to the inner surface of the bark following removal from the stem, composed mainly of developing phloem and cambial initials; **Cork (C)**, after complete removal of the Phloem layer, the entire spongy bark material (varying in thickness ca. 0.5 - 1 cm), including the cork cambium. The tissue samples were preserved on-site in liquid N_2 and retained at –80°C for long-term storage.

RNA Extraction, quality control and cDNA synthesis

Total RNA was extracted as described by Chang *et al.* (1993) and assayed by agarose gel electrophoresis. The total RNA was digested with RNase-free DNaseI I (Roche Diagnostics GmbH) for 30 minutes at 37°C prior to column-purification using the QIAGEN RNeasy Plant Mini Kit (QIAGEN, Valencia, CA) according to manufacturer's instructions. Poly A^+ RNA was isolated using the Oligotex mRNA Mini Kit (QIAGEN) according to the manufacturer's instructions. Total and mRNA were stored in RNase-free water at –80°C. Double-stranded (ds) cDNA was prepared from 300 ng of purified poly A^+ RNA using the cDNA Synthesis

System of Roche. The ds-cDNA was column-purified using the QIAquick PCR Purification Kit (QIAGEN). The cDNAs were assayed for genomic DNA contamination by PCR using an intron-exon boundary spanning primer pair: EgCAD-F (CACTGATTCGCTCGACTACG) and EgCAD-R (TCGCCAACCACTATCTCACCAG), specific for the *E. gunni* cinnamyl alcohol dehydrogenase (CAD2) gene known to be expressed in wood-forming tissues (Grima-Pettenati *et al.*, 1993). Ten nanogram of cDNA were used as template in the following thermal cycling reaction: initial denaturation for 2 minutes at 94°C followed by 30 cycles of 20 s denaturation at 94°C, 30 s annealing at 54°C and 1 minute of elongation at 72°C. Products were analysed by agarose gel electrophoresis to verify absence of genomic DNA-derived products in cDNA samples.

cDNA-AFLP analysis

cDNA-AFLP analysis was performed as described by Bachem *et al.* (1996), but using the AFLP Expression Analysis Kit of LI-COR (LI-COR Biosciences, Lincoln, NE). One hundred nanogram of double-stranded cDNA was used as initial template and the manufacturer's instructions were followed in the generation of *Taq*I+0/*Mse*I+0 pre-amplification PCR products, which were assayed for quality and quantity by electrophoresis on 1% agarose gels. The pre-amplification products were diluted 1:300 in sterile water (SABAX) and used as template for final selective amplification. Selective PCRs were performed using all 64 +2/+2 primer combinations afforded by the eight *Taq*I+2 primers and eight *Mse*I+2 primers (+GA, +GT, +TC, +TG, +CT, +CA, +AG and +AC on both adaptor primers) provided in the AFLP Expression Analysis Kit. The *Taq*I+2 primers in this kit are fluorescently labelled with infrared dye (IRDye700, LI-COR) for the purpose of fragment visualization. Selective PCR products were resolved on 8% denaturing polyacrylamide gels in model 4200S LI-COR DNA Analysers as previously described (Myburg *et al.*, 2001). cDNA-AFLP images were saved in 16-bit TIFF format for image analysis.

Image analysis and TDF quantification

Quantity One 1-D Analysis Software (Bio-Rad Laboratories) was used to crop the primer fronts from LI-COR TIFF images (while preserving 16-bit image depth) before cDNA-AFLP band sizes and intensities were determined using the AFLP-Quantar*Pro* software (KeyGene products B.V., Wageningen, The Netherlands). Lane finding, band finding and sizing were performed as described in the AFLP-Quantar*Pro* user manual, with band finding and scoring parameters previously described for LI-COR gels (Myburg *et al.*, 2001). Only differentially expressed TDFs (based on visual inspection) were quantified in AFLP-Quantar*Pro*. Band intensities were automatically lane-to-lane normalized by the software based on the total lane intensity to correct for loading inconsistencies and other technical artefacts. Band intensities were exported to Microsoft Excel for further analysis.

Clustering and identification of gene expression patterns

Cluster analysis was performed on the normalized band intensities using the Cluster program (Eisen *et al.*, 1998) and the open-source software Java TreeView (Saldanha, 2004), in order to identify groups of TDFs with similar expression patterns across the four tissues and two height levels in the tree. Following mean centring and standardization, distances were calculated using the standard Pearson's correlation and the expression profiles clustered using the hierarchical centroid linkage algorithm in Cluster. The output of the clustering algorithm was visualized using Java TreeView.

TDF isolation and identification

Following partial electrophoresis on LI-COR DNA Analysers, polyacrylamide gels containing fragments of interest were scanned using the Odyssey Infrared Imager (LI-COR). These fragments were excised and the PCR products eluted in TE buffer. Elution was achieved by 5 cycles of freezing (-20°C) and thawing. Eluted PCR products were re-

amplified using the same primer combination used in the particular final amplification reaction. Following re-amplification, the PCR products were resolved on a LI-COR gel alongside the original cDNA-AFLP fragments to confirm re-amplification of the correct TDF. Re-amplified TDFs were cloned using the TOPO TA Cloning Kit for Sequencing (Invitrogen). TDF inserts were sequenced using ABI Bigdye terminator chemistry on ABI3100 instruments with standard M13 vector primers. Sequences were assigned putative identities by translating BLAST (BLASTX) (Altschul *et al.*, 1990) against the non-redundant protein database in Genbank. Additionally, all sequences were subjected to similarity searches using nucleotide blast (BLASTN), mainly to identify hits to possible contaminants such as rRNA and genomic DNA products.

Quantitative RT-PCR confirmation of differential gene expression

Quantitative, reverse-transcription PCR (qRT-PCR) was performed using a LightCycler version 1.2 instrument (Roche) to confirm the TDF expression patterns identified by cDNA-AFLP, using a two-step RT-PCR approach. First strand cDNA was synthesized using the ImProm-II Reverse Transcription System (Promega, Madison, WI) starting from 1 µg of total RNA extracted from the same tissues that were used for cDNA-AFLP analysis. The LightCycler FastStart DNA Master^{PLUS} SYBR Green I system (Roche) was used for real-time PCR using the first strand cDNAs as template, as specified by the instruction manual. All PCR reactions were performed in triplicate. Gene-specific primers were designed based on selected TDF sequences using Primer Designer 5 software (Scientific & Educational Software, Cary, NC). Relative quantification was performed using the LightCycler software version 3.5.3 (Roche).

RESULTS

RNA and cDNA quality

RNA purified from woody tissues of *Eucalyptus* was found to be of high quality and the absence of contaminating genomic DNA was confirmed for all cDNA samples (Figure 1). The amplification of a region of the CAD2 gene from cDNA yielded the expected 410 bp mRNA-derived amplicon, which was clearly distinguishable from the 700 bp genomic DNA-derived, intron-containing fragment.

cDNA-AFLP expression patterns

cDNA-AFLP analysis on LI-COR DNA analysers allowed high-throughput, high-resolution identification of differentially expressed TDFs (Figure 2). Generally, fragments ranging in sizes from 100 bp to over 700 bp were visualized and scored. Many different expression profiles were generated using the 64 primer combinations across 8 tissues, including absolute presence or absence of TDFs and changes in relative abundance of TDFs (Figure 2). Selective amplifications using a single +2/+2 primer combination yielded on average 100 discrete bands per lane. Of these, approximately 15 to 30 were found to be differentially expressed. Of the 6385 TDFs generated, 1374 (21%) exhibited variable expression levels in the different tissues and sections of the tree being analysed. To obtain quantitative expression data for the identified TDFs, the 1374 TDFs were semi-automatically scored using the AFLP-Quantar*Pro* software. Differentially expressed fragments identified by visual inspection corresponded to TDF intensities that varied at least 4-fold in one of the tissues.

Clustering and tissue-specific expression patterns

The 1374 differentially expressed TDFs were clustered according to their expression patterns across the 8 tissue samples (Figure 3A). Based on the TDF clustering, groups of co-expressed

TDFs could be annotated (Figure 3A and B). The tissues were also clustered according to the expression profiles of the 1374 TDFs within each tissue. In all cases, upper and lower-half samples of the same tissues were grouped together. At a higher level, the Mature Xylem and Immature Xylem expression profiles were grouped, as were Phloem and Cork profiles (Figure 3A). Within tissues, the upper and lower half expression profiles were much more similar (R^2 values ranging from 0.45 to 0.68) than comparisons among different tissues within the same height class (R^2 values ranging from 0.017 to 0.43). Expression profile similarity between tissues separated by a number of layers such as Mature Xylem and Cork was very low, while neighbouring tissues (e.g. Mature Xylem and Immature Xylem) showed a much higher degree of similarity.

TDF isolation and characterization

TDFs representing clusters with interesting expression patterns were excised from polyacrylamide gels with the aid of an Odyssey infrared imager. This instrument allowed re-scanning of gels to confirm isolation of the correct bands (Figure 4). A total of 71 TDFs were successfully excised, re-amplified, cloned, sequenced and assigned putative identities using BLASTX. Based on the putative functions of the proteins inferred by similarity, the TDFs were broadly classified into eight functional categories: Defence, Signalling and signal transduction, Transport (intra- and inter-cellular), Cell wall biosynthesis, General primary metabolism (including bioenergetics), Protein biosynthesis and modification, Cell differentiation and development and, lastly, Gene expression regulation (Figure 5). Three additional categories included: TDFs with no significant similarities to any proteins in the non-redundant database; TDFs with similarities to hypothetical proteins or proteins of unknown function; and TDFs with similarity to rRNA. Figure 6 depicts the characterized TDFs (excluding TDFs with no significant similarity and similarity to rRNA) along with the measure of similarity and representation of the actual expression pattern of each TDF. These

TDF sequences were deposited in the dbEST database in Genbank (Genbank accessions AY770746 and DN596730 to DN596773, Figure 6).

Confirmation of TDF expression quantification by qRT-PCR

To verify the accuracy of cDNA-AFLP fragment quantification, we performed qRT-PCR of five sequenced TDFs (Figure 7). The five TDFs that were chosen to represent varying expression profiles across the tissues included fragments with similarity to a secondary xylem-specific cellulose synthase (Genbank accession DN596749) , an S-adenosyl methionine synthase (SAMS, DN596748), an ADP ribosylation factor (AY770746), α tubulin (DN596737) and a TDF similar to an uncharacterised membrane protein from rice (DN596757). cDNA-AFLP analysis suggested that the TDF with similarity to the ADP ribosylation factor was expressed constitutively. Following the confirmation of its constitutive expression profile by qRT-PCR, this gene was chosen as a control for normalization of the qRT-PCR data. The expression patterns as determined using qRT-PCR agreed well with those generated by cDNA-AFLP (Figure 7).

DISCUSSION

Identification and characterization of novel genes in the wood forming tissues of *Eucalyptus* is a priority for the domestication of this genus, which includes many commercially important tree species. The *de novo* description of such genes requires an overview of the gene expression variation in differentiating woody tissues. Accordingly, our goals included: demonstrating that reliable expression profiles could be generated in a high-throughput fashion in woody tissues of *Eucalyptus* using cDNA-AFLP analysis and identifying and partially isolating candidate genes for further characterization.

Ensuring that RNA is entirely free of genomic DNA (as part of quality control prior to cDNA-AFLP analysis) is crucial, as contaminating DNA could be amplified along with the

cDNA, producing false positives. Before performing cDNA-AFLP analysis we subjected the RNA to quality control by amplifying a region of the CAD gene using an intron-spanning primer pair. No contaminating DNA was detected and none of the TDF sequences derived from these tissues suggested genomic DNA contamination.

Initially, we compared the expression profiles generated using cDNA synthesized from total and poly A selected RNA and observed the occurrence of additional bands, possibly attributable to amplified rRNA, in the profiles generated from total RNA (results not shown). Care was therefore taken to ensure that as much ribosomal RNA (rRNA) as possible was removed from total RNA by selection for the poly A tailed messenger RNA. The fact that only two of the cloned TDF sequences were homologous to rRNA sequences, indicated that the poly A selection process was successful in eliminating the majority of the rRNA species.

cDNA-AFLP was performed using automated DNA analysers. The use of infrared detection on the LI-COR and Odyssey systems avoided the use of radioactive isotopes for cDNA-AFLP band visualization (Bachem *et al.*, 1998) or the need for post-electrophoresis staining of gels prior to band excision (Dubos and Plomion, 2003). Use of the infrared detection technology allowed the generation of expression data which was highly replicable with biologically and technically (PCR) repeated experiments generating nearly identical profiles (Figure 2).

One of the major criticisms regarding PCR-based expression profiling techniques (such as cDNA-AFLP) has been fuelled by the observation that conventional PCR is not quantitative. This is despite the evidence of studies clearly demonstrating the ability of cDNA-AFLP to accurately quantify relative gene expression levels at the whole-genome level (Breyne *et al.*, 2003; Reijans *et al.*, 2003). This problem can be circumvented using quantitative PCR which is based on monitoring the accumulation of product during the

exponential phase of the PCR as first described in principle by Higuchi *et al.* (1993). Our examination of the expression profiles of five TDFs using qRT-PCR clearly showed that cDNA-AFLP was accurately quantifying the relative expression levels of the TDFs. The limited number of PCR cycles performed during pre-amplification and selective amplification ensures that the relative TDF quantities were directly proportional to the initial transcript abundance in the RNA samples.

Based on the number of TDFs generated using 64 selective primer combinations, the estimated coverage of the transcriptome can be assessed. Assuming a redundancy of approximately 50% as calculated for an *in silico* scenario by Breyne *et al.* (2003), the 6385 TDFs generated correspond to approximately 3000 unique genes. Under these assumptions, it would be possible to analyse a large proportion of all genes expressed in *Eucalyptus* woody tissues by using the full complement of 256 selective primer combinations possible for the *Taq*I/*Mse*I enzyme combination.

The quantification of TDF expression levels by cDNA-AFLP suggested that a significant proportion of the genes expressed in the wood forming tissues of *Eucalyptus* are strongly up or down-regulated in one or more of the tissue layers sampled. This high variability in gene expression patterns demonstrates the extent to which the major wood forming tissues differ in function, biochemistry and morphology. Additionally, it indicates that our sampling strategy succeeded at separating the main tissue types present in the mature stem. Gene expression patterns in identical tissues sourced from the two regions of the stem (upper and lower half) showed a high level of similarity and these could be thought of as biological replicates. There was very limited similarity between distantly located tissues such as mature xylem and cork suggesting that the distinct morphological and functional differences between these tissues are paralleled at the gene expression level. The accuracy and speed of TDF quantification was aided by the use of the AFLP-Quantar*Pro* software,

without which the generation of consistently normalized expression levels would be problematic.

Clustering of TDFs across the eight tissue samples led to the identification of highly distinct groups of co-expressed genes. These may contain co-ordinately regulated genes which underlie the main metabolic or developmental processes occurring during tissue differentiation. Identification of tissue-specific expression patterns allows the informed selection of TDFs for excision and characterization. This approach minimizes the risk of sequencing numerous fragments all corresponding to the same transcript, as is often the case with other approaches such as EST sequencing.

In addition to generating gene expression profiles across the tree stem, we assigned provisional identities to selected TDFs in order to identify the types of genes expressed in woody tissues of *Eucalyptus* and identify candidates for further analysis. TDFs with homology to characterized proteins were grouped according to broad functional categories. The most numerous group represented proteins involved in general primary cell metabolism. These proteins included bioenergetic pathway enzymes such as phosphoenol pyruvate carboxylase and malate dehydrogenase. Additionally, this group also contained enzymes which catalyse reactions upstream of certain secondary metabolic processes such as lignin biosynthesis – notably hydroxymethyl transferase and s-adenosyl methionine synthase (SAMS). This finding seems consistent with EST sequencing studies in *Populus* and *Pinus*, which suggested that SAMS is one of the most highly expressed genes in woody tissues (Allona *et al.*, 1998; Sterky *et al.*, 1998). Three of our cloned TDFs exhibited significant homology to SAMS, all being strongly up-regulated in mature and immature xylem. The fact that these three TDFs were obtained with different selective primer combinations suggests that a number of different SAMS isoforms or alleles may be expressed in wood-forming tissues. The occurrence of allele-specific TDFs could be attributed to the hybrid genotype of

the *Eucalyptus* tree studied, which would be expected to exhibit a higher degree of divergence at gene loci than pure species individuals. Allele-specific gene expression variation may be a more common phenomenon in plant genomes than previously thought (Guo *et al.*, 2004). Studies of allele-specific gene expression via a cDNA-AFLP approach may be possible in highly heterozygous organisms such as interspecific hybrids of *Eucalyptus* species.

TDFs with functions related to protein processing, synthesis and degradation were also significantly represented. This group included peptides mostly functioning in general protein turnover such as ribosomal subunit proteins, as well as a number of proteins with functions related to protein folding such as chaperonins. One TDF with increased abundance in mature xylem tissues showed homology to an aspartic protease. Aspartic proteinases have been identified in *Arabidopsis* (D'Hondt *et al.*, 1997) and it has been suggested that these and other proteolytic enzymes may be involved in cell autolysis during xylogenesis (Zhao *et al.*, 2000).

Proteins involved in intra- and inter-cellular transport included a number of water channel related proteins such as aquaporins. While the majority of water translocation occurs along the vertical axis via dead xylem elements, water distribution between living cells is also required in the lateral dimension. Thus, the abundance of these membrane proteins in the water transporting tissues of the tree stem is not surprising.

The group of TDFs with possible functions in development and cell differentiation is of great interest as it may contain key genes associated with the differentiation of the vascular cambium. A notable member of this group is a TDF showing homology to the ERECTA protein from *Arabidopsis* – a receptor kinase implicated in the differentiation of tissues from the shoot apical meristem (Torii *et al.*, 1996). The single cell layer thick vascular cambium gives rise to the wide variety of cell types forming the major tissues through a developmental

cascade. Developmental regulators shared between the apical and lateral meristems may provide insight into the basic molecular mechanisms of cell differentiation in emerging plant organs. TDFs with similarities to transcriptional regulators are also good candidates for further characterization. A number of cDNAs encoding distinct members of the HD-ZIP transcription factor family have been isolated from developing vascular tissues in the *Zinnia elegans* model (Ohashi-Ito *et al.*, 2002; Ohashi-Ito and Fukuda, 2003). HD-ZIP family members represent some of the key genes involved in the regulation of lateral meristem differentiation (Bowman *et al.*, 2002).

Compared to the understanding of the regulation of development during lateral differentiation, the precise regulation of key metabolic pathways such as cellulose and lignin biosynthesis during secondary cell wall development remains poorly understood. A key group of TDFs identified in this study represents proteins directly associated with cell wall biosynthesis, with both lignin and cellulose biosynthetic pathway enzymes being represented. These TDFs were all found to be up-regulated in maturing xylem – the site of secondary cell wall deposition. Caffeoyl-CoA O-methyltransferase is an enzyme catalysing the methylation of caffeoyl-CoA to feruloyl-coA and 5-hydroxyferuloyl-CoA to sinapoyl-CoA. It is a key component in the lignin biosynthetic pathway and has been characterized in tobacco (Martz *et al.*, 1998) and poplar (Zhong *et al.*, 2000). A single TDF was similar to a laccase precursor from tobacco. Laccases have been implicated in the oxidative polymerisation of cinnamyl alcohols into lignin – the last step in the lignin biosynthetic pathway. However, a definite link between this process and a gene encoding a laccase has to date not been clearly demonstrated (Boudet, 2000; Boudet *et al.*, 2003). A TDF which was highly upregulated in immature and mature xylem tissues exhibited strong homology to cellulose synthase catalytic subunits. The TDF was highly similar to a secondary cell wall-specific cellulose synthase isolated from *Populus tremuloides* (Wu *et al.*, 2000) and to the *Arabidopsis thaliana* gene AtCesA8

(encoding the cellulose synthase catalytic subunit IRX1), also associated with secondary cell wall deposition (Taylor *et al.*, 2000). TDFs associated with cell wall biosynthesis provide important candidates for further characterization in *Eucalyptus*.

The largest group of characterized TDFs contained open reading frames with no significant similarity to any known protein sequences. Many of these TDFs showed limited similarity to characterized proteins but fell below the 1e-10 threshold imposed - this is an inherent limitation of similarity searches with short sequences. Some of these TDFs are probably derived from the 3' regions of transcripts due to the fact that cDNA synthesis is primed at the poly A tail and full-length cDNA copies are less abundant than truncated partial cDNAs. In contrast, the majority of ESTs in databases are sequenced from the 5' end, thus lowering the likelihood of overlap with 3' biased TDFs. Lastly, 14% of the sequenced TDFs exhibited significant homology to proteins of unknown function, possibly representing transcripts which are uniquely expressed in secondary xylem and not yet fully characterized in herbaceous model species.

Our knowledge of gene expression patterns in the wood-forming tissues of commercially important hardwood forest tree species such as *Eucalyptus* is relatively limited when compared to the wealth of information available for other tree genera such as *Populus* and *Pinus*. Analysing gene expression levels in the major woody tissues of species with limited genomic resources requires a technique that enables the simultaneous generation of expression data along with sequence information. By performing expression analysis across the major wood-forming tissues of the tree stem, we have obtained a comprehensive perspective of the types of genes active in *Eucalyptus* wood, together with their tissue-specific expression levels. We demonstrated that cDNA-AFLP performed using automated DNA analysers is a powerful, fast and relatively inexpensive technique for the analysis of gene expression coupled to gene discovery in forest trees. By quantifying and clustering

TDFs across tissues it is possible to partially isolate candidate genes for further characterization without the requirement for excessive sequencing. It would be relatively simple to adapt this approach for transcriptome analysis in other tree species.

ACKNOWLEDGEMENTS

The authors would like to thank Wietse Mulder for valuable advice on the isolation of TDFs from LI-COR acrylamide gels, as well as Ross Whetten and Dave Berger for excellent suggestions and critical reading of the manuscript. This work was supported by funding provided by Sappi Forest Products and Mondi Business Paper South Africa, through the Wood and Fibre Molecular Genetics Programme. Additional funding was provided by the Technology and Human Resources for Industry Programme (THRIP grant no 2546) and the National Research Foundation of South Africa (NRF grant no 5391).

FIGURES

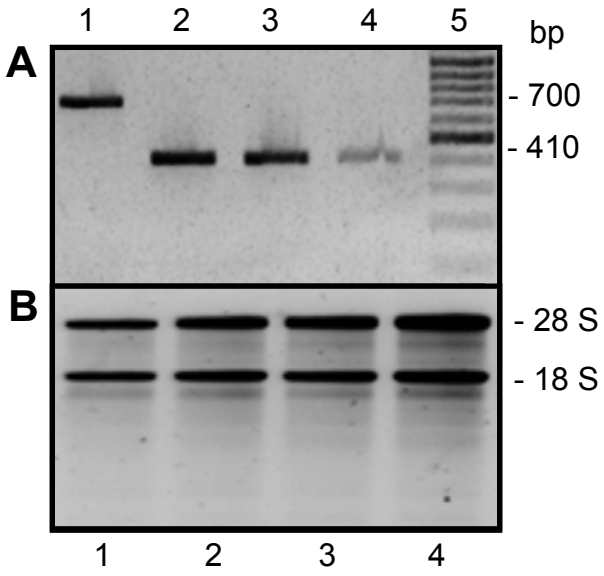


Figure 1

A. Quality control of cDNA using genomic DNA contamination assay. CAD2 intron-spanning PCR was performed using as template control genomic DNA (Lane 1), purified cDNA from mature xylem, immature xylem and phloem (Lanes 2,3,4). Lane 5: molecular weight standard (100 bp ladder, Fermentas). **B**: Total RNA from four woody tissues assayed by agarose gel electrophoresis. Lane 1: mature xylem, Lane 2: immature xylem, Lane 3: phloem, Lane 4: cork.

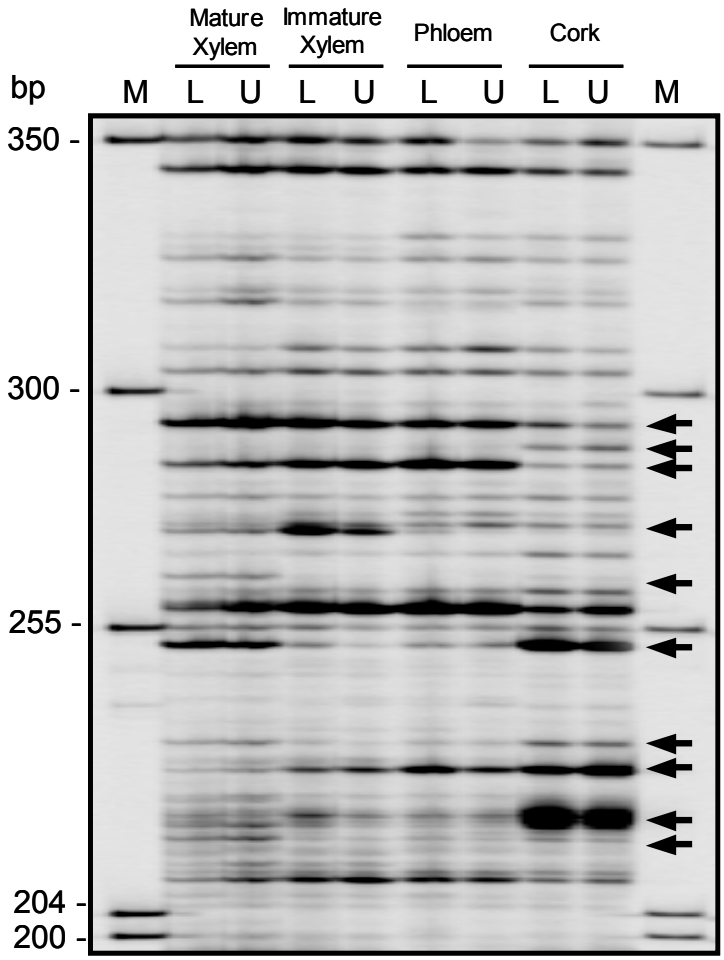
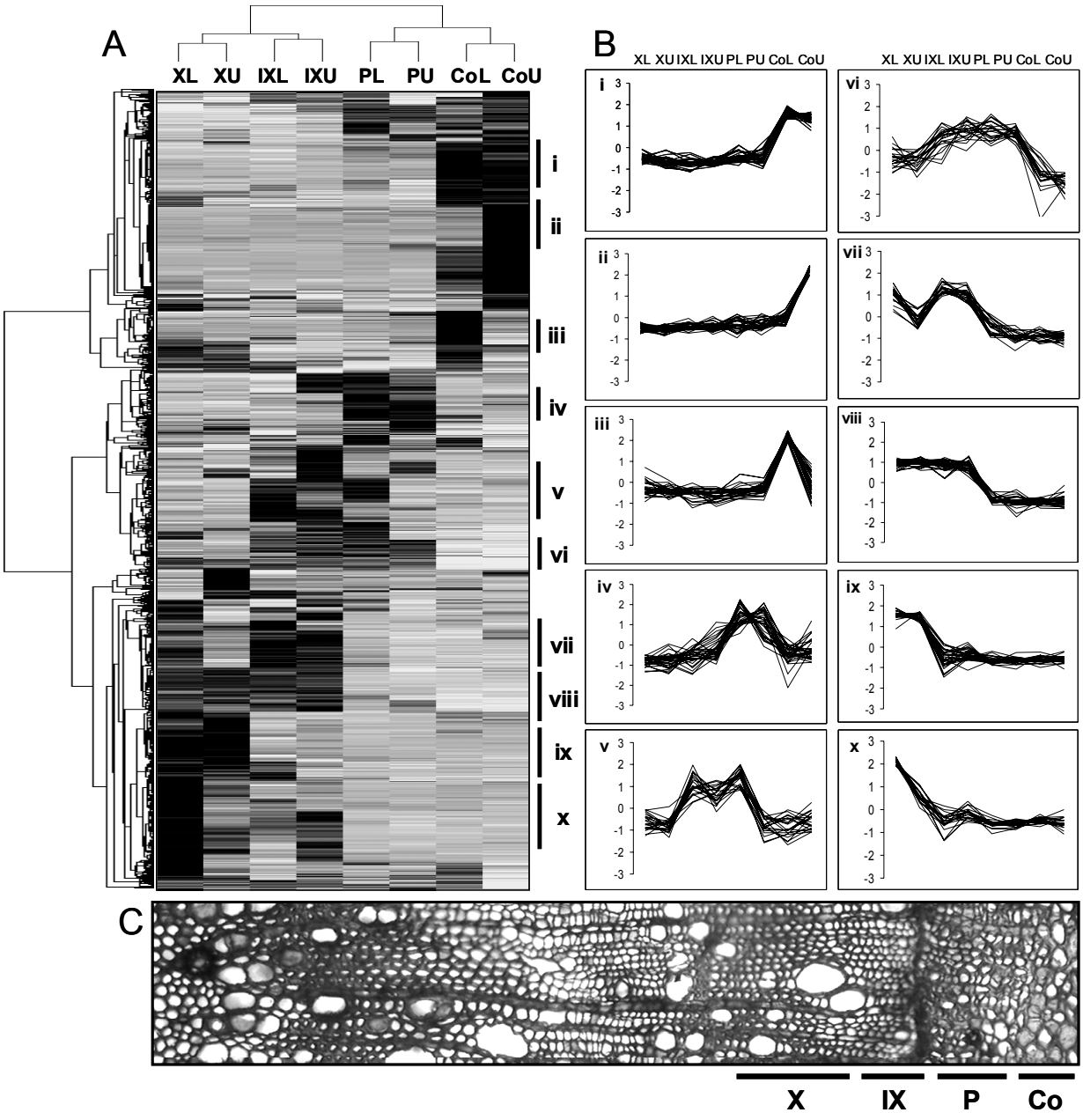


Figure 2

Section of a typical cDNA-AFLP expression profile across the eight different tissues, from the upper (U) or lower (L) sections of the tree. Arrows indicate differentially expressed TDFs. M: IRD700-labeled molecular size standard.

Figure 3

A. Hierarchical clustering of 1374 differentially expressed TDFs and tissue profiles. Relative expression levels are represented by a greyscale continuum with white signifying absence of the TDF in the tissue and black indicating strong upregulation in the respective tissues. The rows correspond to the 1374 quantified TDFs and columns to the respective tissue profiles. **B**. Groups of TDFs with similar expression patterns (location of each group indicated in A by vertical bars numbered i to x). Vertical axes in B represent standard deviation from the mean expression level of each gene. XL and XU: mature xylem, IXL and IXU: immature xylem, PL and PU: phloem, and CL and CU: cork, all collected from either the lower-half or upper-half of the stem. **C**. Section through a young *Eucalyptus* stem with the four major sampled tissue layers highlighted.


A
XL XU IXL IXU PL PU CoL CoU
i
ii
iii
iv
v
vi
vii
viii
ix
x
B
XL XU IXL IXU PL PU CoL CoU
XL XU IXL IXU PL PU CoL CoU
C
X
IX
P
Co

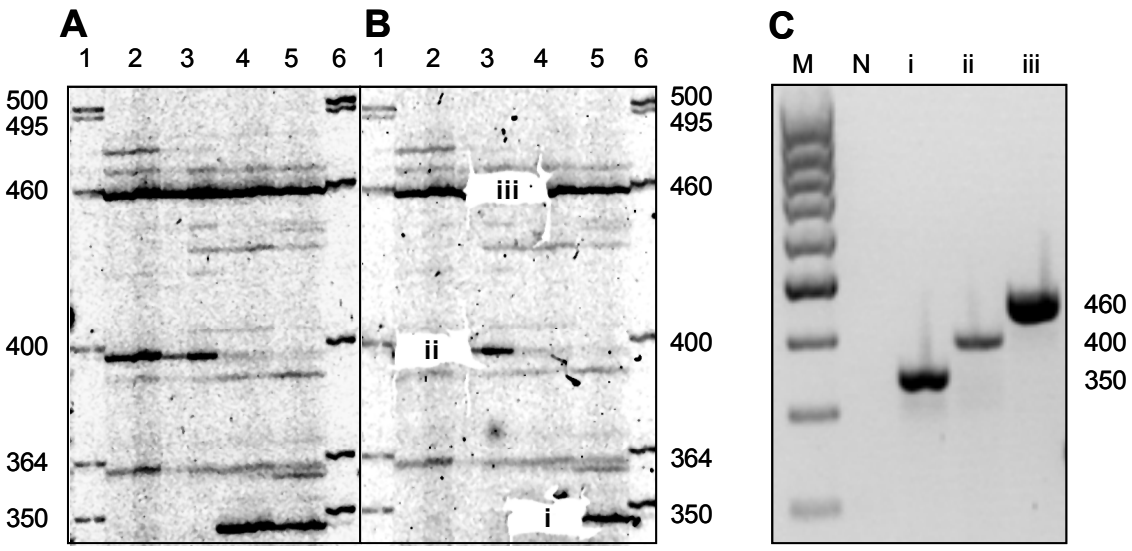


Figure 4

A. Section of a LI-COR gel image showing TDF profiles for a single primer combination across the 4 main tissues scanned on the Odyssey infrared scanner. Lanes 1,6: IRD700-labeled molecular size standard. Lanes 2, 3, 4, 5: mature xylem, immature xylem, phloem and cork respectively. **B**. Identical section of gel image as in A after band excision and re-scanning. i, ii, iii: excised fragments with approximate sizes of 350, 400 and 460 bp, respectively. **C**. Agarose gel electrophoresis of re-amplified fragments in lanes i, ii and iii. Lane M: 100 bp ladder (Fermentas). Lane N: negative PCR control.

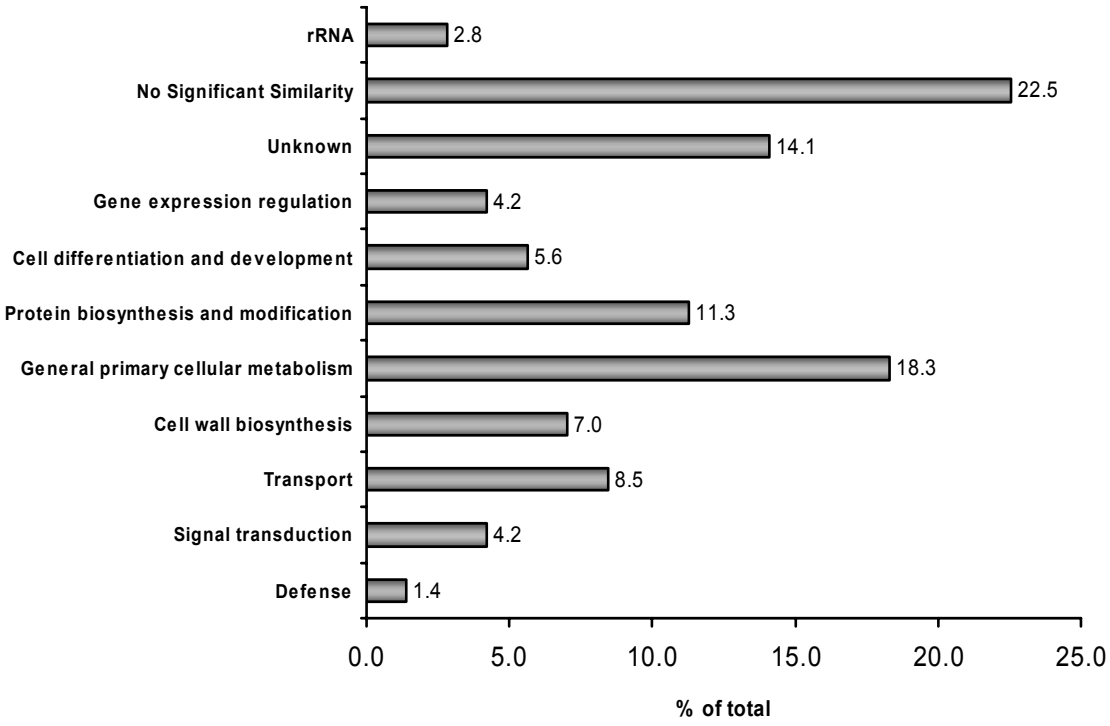


Figure 5

Broad classification of similarity-inferred TDF identities based on BLASTX results. "No significant similarity" entails E-values above $1e^{-10}$. "Unknown" denotes significant similarity to proteins of unknown function.

Genbank Accession	Sequence homology	E-Value	Expression pattern Lh X	Lh IX	Lh P	Lh C	Uh X	Uh IX	Uh P	Uh C
AY770746	ADP ribosylation factor, *Arabidopsis thaliana* (At3g62290)	5.E-42	■	■	■	■	■	■	■	■
DN596730	Putative aquaporin-1, *Phaseolus vulgaris*	1.E-63	■	■	■	■	■	■	■	■
DN596731	Unknown protein, *Oryza sativa*	1.E-11	■	■	■	■	■	■	■	■
DN596732	TCP-1 chaperonin-like protein, *Arabidopsis thaliana* (At5g16070)	5.E-27	■	■	■	■	■	■	■	
DN596733	Gigantea protein (GI), *Arabidopsis thaliana* (At1g22770)	3.E-20	■	■	■		■	■		
DN596766	Leucine rich repeat protein, *Arabidopsis thaliana* (At5g48740)	1.E-10	■	■	■		■	■	■	
DN596734	Putative chloroplast chaperonin, *Oryza sativa*	2.E-26	■	■	■		■	■	■	
DN596735	Putative receptor protein kinase ERECTA, *Arabidopsis thaliana*	3.E-19	■	■	■		■	■	■	
DN596736	Putative ribosomal protein L6, *Arabidopsis thaliana* (At2g18400)	5.E-19	■	■	■		■	■	■	
DN596737	Alpha tubulin, *Zea mays*	2.E-47		■		■		■		■
DN596738	Aspartic proteinase 3, *Nepenthes alata*	4.E-26	■	■			■	■		
DN596739	Caffeoyl-CoA O-methyltransferase, *Eucalyptus globulus*	3.E-56	■	■			■	■		
DN596740	Calmodulin, *Arabidopsis thaliana* (At1g66410)	1.E-23	■	■			■	■		
DN596741	Cytochrome b5, *Nicotiana tabacum*	3.E-19	■	■			■	■		
DN596742	Cytochrome b5, *Olea europaea*	2.E-34	■	■			■	■		
DN596743	D-3-phosphoglycerate dehydrogenase, *Arabidopsis thaliana* (At1g1774)	2.E-34	■	■			■	■		
DN596744	Expressed protein *Arabidopsis thaliana* (At1g14870)	2.E-16	■	■			■	■		
DN596767	Expressed protein, *Arabidopsis thaliana* (At5g62580)	7.E-25	■	■			■	■		
DN596745	HD Zip protein, *Zinnia elegans*	4.E-10	■	■			■	■		
DN596764	Hydroxymethyltransferase, *Arabidopsis thaliana* (At4g37930)	1.E-18	■	■			■	■		
DN596746	Actin-related protein, *Arabidopsis thaliana* (At2g33385)	2.E-17	■	■			■	■		
DN596747	Laccase precursor, *Nicotiana tabacum*	2.E-33	■	■			■	■		
DN596748	S-adenosylmethionine synthetase, *Arabidopsis thaliana* (At1g02500)	7.E-33	■	■			■	■		
DN596749	Secondary xylem cellulose synthase, *Populus tremuloides*	2.E-41	■	■			■	■		
DN596750	Thymidine diphospho glucose-4-6 dehydratase, *Prunus armeniaca*	5.E-41	■	■			■	■		
DN596751	Zinc transporter, *Eucalyptus grandis*	9.E-30	■	■			■	■		
DN596768	CDC2-like protein kinase, *Beta vulgaris*	1.E-25	■				■			
DN596752	Copper chaperone homolog, Glycine max	2.E-13	■	■						
DN596753	Alcohol dehydrogenase, *Vitis vinifera*	4.E-18		■	■			■	■	
DN596754	GATA-1 transcription factor, *Arabidopsis thaliana* (At3g24050)	4.E-19		■	■	■		■	■	■
DN596755	GTP binding protein beta chain, *Glycine max*	7.E-58		■	■			■	■	
DN596756	40S Ribosomal protein S26 *Arabidopsis thaliana* (At2g40510)	1.E-19			■	■			■	■
DN596769	Expressed protein, *Arabidopsis thaliana* (At4g16580)	1.E-27			■	■			■	■
DN596757	Membrane protein, *Oryza sativa*	6.E-37			■	■			■	■
DN596770	Putative plant defensin protein, *Arabidopsis thaliana* (At2g02100)	2.E-27			■	■			■	■
DN596771	Aquaporin, *Ricinus communis*	2.E-59				■				■
DN596772	Ariadne-like protein, *Arabidopsis thaliana* (At1g65430)	5.E-26				■				■
DN596758	Arsenical pump driving Atpase, *Salmonella typhimurium*	2.E-38				■				■
DN596773	IWS1 C-terminus family protein *Oryza sativa*	3.E-13				■				■
DN596759	Major intrinsic protein (MIP), *Arabidopsis thaliana* (At4g19030)	2.E-17				■				■
DN596760	Malate dehydrogenase NADP+ *Arabidopsis thaliana* (At2g19900)	1.E-25				■				■
DN596765	NADH -ubiquinone oxidoreductase subunit PSST, *Lupinus luteus*	1.E-14					■	■	■	■
DN596761	PEP carboxylase isoform I, *Euphorbia tirucalli*	2.E-56				■				■
DN596762	Protein phosphatase 2C (P2C-HA) *Arabidopsis thaliana* (At1g72770)	1.E-71					■	■	■	■
DN596763	T complex protein, *Cucumis sativus*	2.E-34				■				

Figure 6

Identities and tissue-specific expression patterns of 45 selected TDFs. The putative identities of the TDFs are displayed along with a representation of their expression patterns (black and white shading representing relative up or downregulation in the particular tissue). TDFs with no significant sequence similarities (Figure 5) are excluded, as well as TDFs with similarities to rRNA. X, IX, P, C: xylem, immature xylem, phloem and cork, from lower (Lh) or upper (Uh) half of the stem.

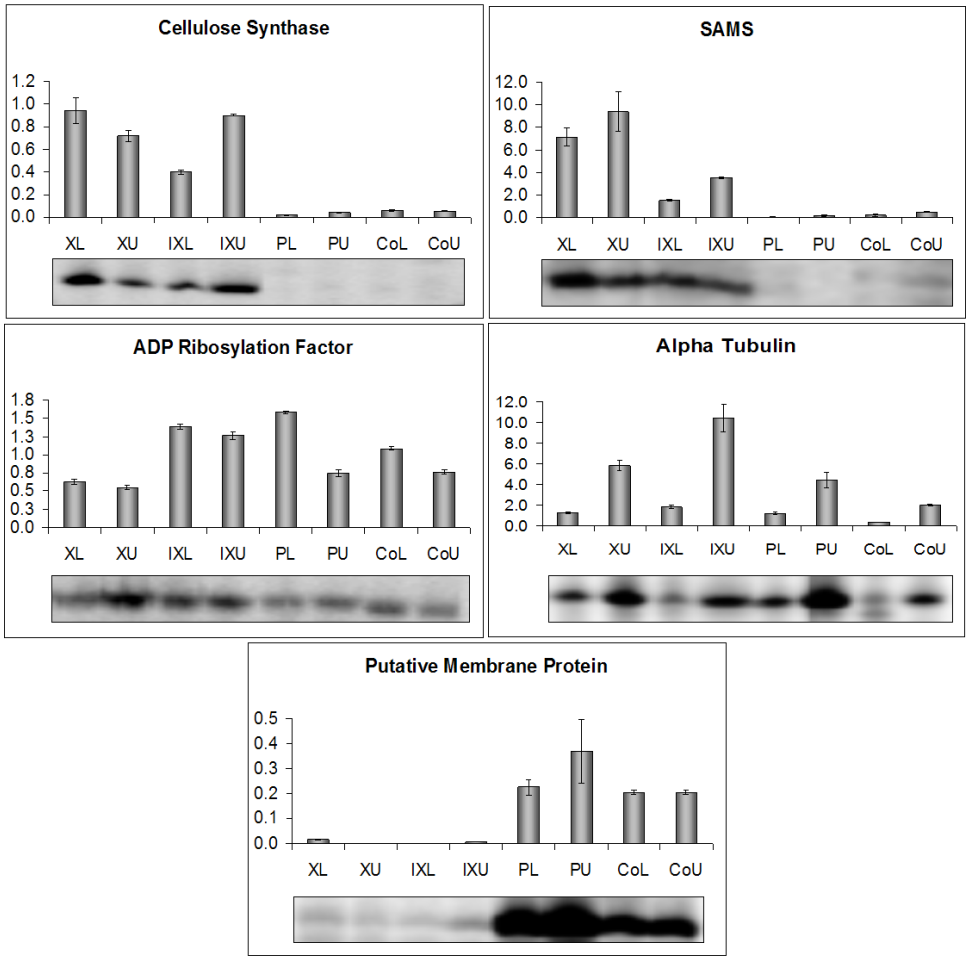


Figure 7

Confirmation of the expression patterns of five TDFs by relative quantification of transcript abundance using qRT-PCR. Vertical axes represent the measure of TDF abundance within the specific tissue, relative to the expression level of the constitutively expressed ADP ribosylation factor. Segments of the original cDNA-AFLP gel images containing the TDFs are shown below the horizontal axes for comparison. Relative abundance of ADP ribosylation factor is shown in standardised units. X, IX, P, Co: xylem, cambium, phloem and cork, from upper (U) or lower (L) half of the stem.

REFERENCES

Allona, I., Quinn, M., Shoop, E., Swope, K., St. Cyr, S., Carlis, J., Riedl, J., Retzel, E., Campbell, M.M., Sederoff, R. and Whetten, R.W. 1998. Analysis of xylem formation in pine by cDNA sequencing. Proc Natl Acad Sci U S A 95: 9693-9698.

Altschul, S.F., Gish, W., Miller, W., Meyers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. J Mol Biol 215: 403-410.

Bachem, C.W.B., Oomen, R.J.F.J. and Visser, R.G.F. 1998. Transcript imaging with cDNA-AFLP: a step-by-step protocol. Plant Mol Biol Rep 16: 157-173.

Bachem, C.W.B., van der Hoeven, R.S., de Bruijn, S.M., Vreugdenhil, D., Zabeau, M. and Visser, R.G.F. 1996. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. Plant J 9: 745-753.

Barnett, J. and Bonham, V. 2004. Cellulose microfibril angle in the cell wall of wood fibres. Biol Rev Camb Philos Soc 79: 461-472.

Boudet, A.M. 2000. Lignins and lignification: Selected issues. Plant Physiol Biochem 38: 81-96.

Boudet, A.M., Kajita, S., Grima-Pettenati, J. and Goffner, D. 2003. Lignins and lignocellulosics: a better control of synthesis for new and improved uses. Trends Plant Sci 8: 576-581.

Bowman, J.L., Eshed, Y. and Baum, S.F. 2002. Establishment of polarity in angiosperm lateral organs. Trends Genet 18: 134-141.

Breyne, P., Dreesen, R., Cannoot, B., Rombaut, D., Vandepoele, K., Rombauts, S., Vanderhaeghen, R., Inze, D. and Zabeau, M. 2003. Quantitative cDNA-AFLP analysis for genome-wide expression studies. Mol Genet Genomics 269: 173-179.

Chaffey, N., Cholewa, E., Regan, S. and Sundberg, B. 2002. Secondary xylem development in *Arabidopsis*: a model for wood formation. Physiol Plant 114: 594-600.

Chang, S., Puryear, J. and Cairney, J. 1993. A simple and efficient method for isolating RNA from pine trees. Plant Mol Biol Rep 11: 113-116.

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. and Fukuda, H. 2002. Visualization by comprehensive microarray analysis of gene expression programs during transdifferentiation of mesophyll cells into xylem cells. Proc Natl Acad Sci U S A 99: 15794-15799.

D'Hondt, K., Stack, S., Gutteridge, S., Vandekerckhove, J., Krebbers, E. and Gal, S. 1997. Aspartic proteinase genes in the *Brassicaceae Arabidopsis thaliana* and *Brassica napus*. Plant Mol Biol 33: 187-192.

Dubos, C. and Plomion, C. 2003. Identification of water-deficit responsive genes in maritime pine (*Pinus pinaster* Ait.) roots. Plant Mol Biol 51: 249-262.

Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95: 14863-14868.

Evans, R. and Ilic, J. 2001. Rapid prediction of wood stiffness from microfibril angle and density. Forest Products Journal 51: 53-57.

Fukuda, H. and Komamine, A. 1980. Establishment of an experimental system for the tracheary element differentiation from single cells isolated from the mesophyll of *Zinnia elegans*. Plant Physiol 65: 57-60.

Grima-Pettenati, J., Feuillet, C., Goffner, D., Borderies, G. and Boudet, A.M. 1993. Molecular cloning and expression of a *Eucalyptus gunnii* cDNA clone encoding cinnamyl alcohol dehydrogenase. Plant Mol Biol 21: 1085-1095.

Guo, M., Rupe, M.A., Zinselmeier, C., Habben, J., Bowen, B.A. and Smith, O.S. 2004. Allelic variation of gene expression in maize hybrids. Plant Cell 16: 1707-1716.

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

Higuchi, R., Fockler, C., Dollinger, G. and Watson, R. 1993. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Biotechnology 11: 1026-1030.

Israelsson, M., Eriksson, M.E., Hertzberg, M., Aspeborg, H., Nilsson, P. and Moritz, T. 2003. Changes in gene expression in the wood-forming tissue of transgenic hybrid aspen with increased secondary growth. Plant Mol Biol 52: 893-903.

Jensen, P.J., Rytter, J., Detwiler, E.A., Travis, J.W. and McNellis, T.W. 2003. Rootstock effects on gene expression patterns in apple tree scions. Plant Mol Biol 53: 493-511.

Ko, J.H., Chow, K.S. and Han, K.H. 2003. Transcriptome analysis reveals novel features of the molecular events occurring in the laticifers of *Hevea brasiliensis* (para rubber tree). Plant Mol Biol 53: 479-492.

Ko, J.H., Han, K.H., Park, S. and Yang, J.M. 2004. Plant body weight-induced secondary growth in *Arabidopsis* and its transcription phenotype revealed by whole-transcriptome profiling. Plant Physiol 135: 1069-1083.

Lev-Yadun, S. 1994. Induction of sclereid differentiation in the pith of *Arabidopsis thaliana* (L.) Heynh. J Exp Bot 45: 1845-1849.

Martz, F., Maury, S., Pincon, G. and Legrand, M. 1998. cDNA cloning, substrate specificity and expression study of tobacco caffeoyl-CoA 3-*O*-methyltransferase, a lignin biosynthetic enzyme. Plant Mol Biol 36: 427-437.

Milioni, D., Sado, P.E., Stacey, N.J., Domingo, C., Roberts, K. and McCann, M.C. 2001. Differential expression of cell-wall-related genes during the formation of tracheary elements in the *Zinnia* mesophyll cell system. Plant Mol Biol 47: 221-238.

Milioni, D., Sado, P.E., Stacey, N.J., Roberts, K. and McCann, M.C. 2002. Early gene expression associated with the commitment and differentiation of a plant tracheary element is revealed by cDNA-amplified fragment length polymorphism analysis. Plant Cell 14: 2813-2824.

Myburg, A.A., Remington, D.L., O'Malley, D.M., Sederoff, R.R. and Whetten, R.W. 2001. High-throughput AFLP analysis using infrared dye-labeled primers and an automated DNA sequencer. BioTechniques 30: 348-357.

Ohashi-Ito, K., Demura, T. and Fukuda, H. 2002. Promotion of transcript accumulation of novel *Zinnia* immature xylem-specific HD-Zip III homeobox genes by brassinosteroids. Plant Cell Physiol 43: 1146-1153.

Ohashi-Ito, K. and Fukuda, H. 2003. HD-Zip III homeobox genes that include a novel member, ZeHB-13 (*Zinnia*)/ATHB-15 (*Arabidopsis*), are involved in procambium and xylem cell differentiation. Plant Cell Physiol 44: 1350-1358.

Reijans, M., Lascaris, R., Groeneger, A.O., Wittenberg, A., Wesselink, E., van Oeveren, J., de Wit, E., Boorsma, A., Voetdijk, B., van der Spek, H., Grivell, L.A. and Simons, G. 2003. Quantitative comparison of cDNA-AFLP, microarrays, and genechip expression data in *Saccharomyces cerevisiae*. Genomics 82: 606-618.

Rudie, A.W. 1998. Wood and how it relates to paper products. Tappi J 81: 223-228.

Saldanha, A.J. 2004. Java Treeview--extensible visualization of microarray data. Bioinformatics 20: 3246-3248.

Sterky, F., Regan, S., Karlsson, J., Hertzberg, M., Rohde, A., Holmberg, A., Amini, B., Bhalerao, R., Larsson, M., Villarroel, R., Van Montagu, M., Sandberg, G., Olsson, O.,

Teeri, T.T., Boerjan, W., Gustafsson, P., Uhlen, M., Sundberg, B. and Lundeberg, J. 1998. Gene discovery in the wood-forming tissues of poplar: Analysis of 5,692 expressed sequence tags. Proc Natl Acad Sci U S A 95: 13330-13335.

Taylor, N.G., Laurie, S. and Turner, S.R. 2000. Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in *Arabidopsis*. Plant Cell 12: 2529-2539.

Torii, K.U., Mitsukawa, N., Oosumi, T., Matsuura, Y., Yokoyama, R., Whittier, R.F. and Komeda, Y. 1996. The *Arabidopsis* ERECTA gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. Plant Cell 8: 735-746.

Wu, L., Joshi, C.P. and Chiang, V.L. 2000. A xylem-specific cellulose synthase gene from aspen (*Populus tremuloides*) is responsive to mechanical stress. Plant J 22: 495-502.

Yang, J.M., Kamdem, D.P., Keathley, D.E. and Han, K.H. 2004a. Seasonal changes in gene expression at the sapwood-heartwood transition zone of black locust (*Robinia pseudoacacia*) revealed by cDNA microarray analysis. Tree Physiol 24: 461-474.

Yang, S.H., van Zyl, L., No, E.G. and Loopstra, C.A. 2004b. Microarray analysis of genes preferentially expressed in differentiating xylem of loblolly pine (*Pinus taeda*). Plant Science 166: 1185-1195.

Zhao, C.S., Johnson, B.J., Kositsup, B. and Beers, E.P. 2000. Exploiting secondary growth in *Arabidopsis*. Construction of xylem and bark cDNA libraries and cloning of three xylem endopeptidases. Plant Physiol 123: 1185-1196.

Zhong, R.Q., Morrison, W.H., Himmelsbach, D.S., Poole, F.L. and Ye, Z.H. 2000. Essential role of caffeoyl coenzyme A *O*-methyltransferase in lignin biosynthesis in woody poplar plants. Plant Physiol 124: 563-577.

CHAPTER 3

SIX NEW CELLULOSE SYNTHASE GENES FROM *EUCALYPTUS* ARE LINKED WITH PRIMARY AND SECONDARY CELL WALL BIOSYNTHESIS.

Martin Ranik and Alexander A. Myburg

Submitted: Plant Molecular Biology (2005)

ABSTRACT

Higher plants contain a family of cellulose synthase catalytic subunit genes (*CesA*) which encode components of the cellulose-synthesizing enzyme complex embedded in the cell membrane. Recent studies in monocot and dicot species demonstrated that two groups of *CesA*s exist, associated with either primary or secondary cell wall deposition. We cloned six full-length *CesA* cDNAs from *Eucalyptus grandis* (*EgCesA1* through *6*) and determined their expression patterns in a variety of organs from an adult tree. The six *EgCesA*s encode predicted proteins of 978 to 1097 amino acid residues, each of which contains all of the key regions and motifs characteristic of functional CESAs. The predicted EgCESA proteins share limited amino acid identity with each other, ranging from 61% to 70%. In contrast, the most similar CESAs from higher plant species exhibit 81% to 90% identity with the six EgCESAs. Gene expression analysis using quantitative reverse-transcription polymerase chain reaction (qRT-PCR) indicated that transcripts of *EgCesA1* through *3* were abundant in tissues enriched for cells laying down secondary cell walls (e.g. xylem), while being very weakly expressed in tissues undergoing primary growth (e.g. unfolding leaves). Expression of *EgCesA4* and *EgCesA5* was upregulated in tissues rich in rapidly dividing cells undergoing primary wall synthesis, while *EgCesA6* was very weakly expressed in all of the tissues analysed. The results suggest that *Eucalyptus*, like other higher plants, expresses two contrasting groups of apparently co-regulated *CesA*s involved in either primary or secondary cell wall biosynthesis.

INTRODUCTION

One of the defining features of plant cells is the cell wall, a significant proportion of which is composed of cellulose fibres. Consequently, cellulose is the most abundant biopolymer on

our planet and occurs as continuous, unbranched, chains of D-glucose monomers polymerised via (1→4)-β bonds. The vast variety of plant structures in existence demonstrates the versatility of cellulose in the determination of cell and consequently organ morphology. As the main component of wood, a natural product for which there is currently no substitute available, cellulose is the cornerstone of a multi-billion dollar worldwide industry (Fenning and Gershenzon, 2002).

Ultrastructural studies of plant cell walls have suggested the site of cellulose biosynthesis to be a large (greater than 500 kDa) rosette-shaped protein terminal complex (TC) associated with the cell membrane (Mueller and Brown, 1980). Recent efforts to synthesize cellulose *in vitro* from plant membrane preparations, indicated that intact TCs were required for the formation of crystalline cellulose (Kudlicka and Brown, 1997; Lai-Kee-Him *et al.*, 2002). Rosettes are composed of up to 36 individual catalytic subunits, each with cellulose synthase activity, and a number of models for rosette assembly and catalytic dynamics have been proposed to date (Brown and Saxena, 2000).

The first cellulose synthase gene characterised was identified as part of the cellulose synthase operon of *Acetobacter xylinum*, a cellulose-producing bacterium (Saxena *et al.*, 1990; Wong *et al.*, 1990) . It was not until the contribution of Pear *et al.* (1996), that the first plant cellulose synthases were identified. Two cellulose synthase genes were discovered in a sequenced cotton fiber cDNA library, by comparing amino acid motifs common to bacterial cellulose synthase proteins. Definitive evidence linking plant cellulose synthase genes to rosettes was provided by Arioli *et al.* (1998), who observed that a single point mutation in a cellulose synthase gene leads to the disassembly of rosettes in the *Arabidopsis thaliana* mutant *rsw1*. Immunogold labelling of rosettes showed that they contain subunits encoded by cellulose synthase genes (Kimura *et al.*, 1999). Cellulose synthase catalytic subunit genes,

abbreviated as [1]*CesA* (Delmer, 1999), have since been discovered in a number of higher plant species including *Arabidopsis*, rice, maize and poplar (Holland *et al.*, 2000; Richmond, 2000; Tanaka *et al.*, 2003; Joshi *et al.*, 2004). Recently, *CesA* genes were characterised from other genera including algae (Roberts *et al.*, 2002; Roberts and Roberts, 2004), social amoeba (Blanton *et al.*, 2000) and even animals (Matthysse *et al.*, 2004).

Expressed sequence tag (EST) and genome sequencing coupled to mutant studies have revealed that higher plant *CesA*s belong to a gene family, with functionally distinct family members. The *Arabidopsis thaliana* genome harbours 10 *CesA*s (Richmond, 2000), six of which have been identified as causative genes in cell wall mutant lines. *CesA* mutants in *Arabidopsis* form two distinct groups: those affecting cellulose deposition in primary cell walls: *rsw1*, *ixr1*, *ixr2, cev1, eli1, prc1* and those affecting the walls of cells undergoing secondary thickening: *irx1*, *irx3*, *irx5*, see Scheible and Pauly (2004). The three secondary cell wall related *CesA* mutants *irx1* (Taylor *et al.*, 2000), *irx3* (Taylor *et al.*, 1999) and *irx5* (Taylor *et al.*, 2003) are caused by mutations in *AtCesA8*, *AtCesA7* and *AtCesA4* respectively. Despite limited amino acid residue identity of the causative proteins (ca. 65%), the phenotypes of the three mutants are nearly identical. *AtCesA8*, *7* and *4* are expressed at similar levels in the same cell types at the same developmental stage (Taylor *et al.*, 2000; Taylor *et al.*, 2003). Additionally, the *CesA* proteins IRX1, IRX3 and IRX5 tend to co-purify (Taylor *et al.*, 2003), presumably due to their association as components of the rosette *in vivo*.

[1] In this study, we maintained the use of the convention for naming of the cellulose synthase genes, according to Delmer (1999) and Samuga and Joshi (2004). Cellulose synthase catalytic subunit genes are abbreviated as *XxCesA#* – with the first two letters signifying the organism of origin and the last character being a digit to differentiate between distinct family members from that particular species. Consequently, the *Arabidopsis thaliana* cellulose synthase catalytic subunit gene 1 would be abbreviated as *AtCesA1*. To differentiate between genes and their protein products, the cellulose synthase proteins are abbreviated similarly but without italicisation and with capital CESA e.g. AtCESA1.

Similar observations were made for the group of *AtCesA*s thought to act in primary cell wall development. Mutations in two relatively divergent *Arabidopsis* cellulose synthase genes, *AtCesA3* and *AtCesA6*, produce very similar *irregular xylem* phenotypes: *ixr1* and *ixr2* (Scheible *et al.*, 2001; Desprez *et al.*, 2002). Burn *et al.* (2002) showed via complementation studies that *AtCesA1* and *AtCesA3*, two distinct *CesA*s required for primary cell wall biosynthesis, are not functionally redundant. Several conclusions can therefore be drawn from the *Arabidopsis CesA* studies. Firstly, different *CesA*s are probably involved in the synthesis of cellulose in primary cell walls than are needed in the deposition of secondary cell walls. Secondly, a number (probably three) of distinct *CesA*s are required for the formation of a functional cellulose synthase complex (rosette). Lastly, the individual *CesA* components of the rosette are not functionally redundant.

CesA isolation and characterisation in other plant species has led to the reinforcement of the evidence gleaned from *Arabidopsis* work. In rice, mutations in any one of the three *CesA* genes orthologous to *AtCesA8*, *AtCesA7* and *AtCesA4* induce the identical brittle culm phenotype coupled to significantly reduced cellulose levels (Tanaka *et al.*, 2003). Detailed analysis of *CesA* expression patterns in barley showed that two groups of *CesA*s exist – expressed preferentially in cells laying down primary or secondary cell walls (Burton *et al.*, 2004). The only angiosperm forest tree species in which the *CesA* gene family has been well documented is poplar. Seven full-length *PtrCesA*s from *Populus tremuloides* have been isolated and comprehensively characterised, as reviewed by Joshi *et al.* (2004). Recently, Nairn and Haselkorn (2005) described three secondary cell wall-associated *CesA*s from the gymnosperm *Pinus taeda* and postulated that higher plant *CesA* family members have retained functional conservation since before the angiosperm-gymnosperm divergence.

CESA proteins contain certain defining regions which occur in all of the known higher plant CESAs (Figure 1A). The N-terminal of the CESAs contains a motif with

similarity to a RING finger domain which has been shown to bind Zn^{2+} and is likely to be involved in the dimerization of CESA proteins (Kurek *et al.*, 2002). Following the RING finger domain lies the first of two regions which show very limited conservation between CESA family members from the same species. Although originally described as hypervariable regions (HVRI and HVRII), these regions are conserved in CESA orthologs from different species and the term class specific regions (CSRI and CSRII) was suggested to define them (Vergara and Carpita, 2001). The CSRII amino acid residue identity between orthologs from two different species can be significantly higher than between two distinct CESA family members from the same species (Figure 1B). The cDNA regions of *CesA*s corresponding to CSRII are therefore very useful for distinguishing between individual family members (Liang and Joshi, 2004) and can potentially serve as starting points for full-length cDNA isolation. CSRI and CSRII are separated by a 300-400 amino acid residue-long region which is highly conserved in plant CESAs. Four motifs essential for the inverting glycosyl transferase activity have been identified in CESA proteins (Saxena *et al.*, 1995; Saxena *et al.*, 2001). Three of these are widely-spaced aspartate residues located in the conserved regions, the fourth being the "QXXRW" motif located approximately 35 residues after the third catalytic aspartate (Figure 1A). Additionally, all plant CESAs currently known are predicted to contain between eight and ten transmembrane domains, two of which are located at the start of the first conserved region with the others occurring near the C terminal.

Fast growing clones of *Eucalyptus* tree species are some of the most rapid producers of wood fibre (composed mostly of cellulose) known, designating *Eucalyptus* as an important system for studying the molecular biology of cellulose biosynthesis in forest trees. To date, the conservation and expression patterns of *CesA* genes in *Eucalyptus* remain unexplored. In this study, six previously unpublished full-length *Eucalyptus grandis CesA* cDNAs were isolated and their expression profiles determined in a variety of tissues.

MATERIALS AND METHODS

Plant materials

Tissues rich in cells that are actively depositing primary as well as secondary cell wall types were collected into liquid N_2 from a destructively sampled 4 year-old *Eucalyptus grandis* tree. The following tissues were collected from the debranched stem immediately after the removal of the bark. **Immature Xylem (IX)**: outer 1-2 mm glutinous layer covering the stem comprising xylem mother cells and early developing xylem tissue; **Xylem (X)**: 3-5 mm deep planing following the removal of the IX layer, encompassing xylem cells in stages of advanced maturity; **Phloem (P)**: 1-2 mm layer from the inner surface of the bark containing the vascular cambium and developing phloem cells; **Cork (C):** entire spongy bark material following the removal of the P layer, consisting of cork, cork cambium as well as some phloem tissue. In addition to the stem tissues, the following were sampled: **Young leaves (Yl)**: unfolding young leaves; **Flowers (F)**: stage 2 flowers; **Internodes (I)**: young shoot internodes undergoing secondary xylem deposition as well as rapid apical elongation.

Nucleic acid isolation and purification

Total RNA was extracted according to the method described by Chang *et al.* (1993). Each total RNA sample was incubated with 50U RNase-free DNaseI (Roche) for 30 min at 37°C in the presence of 10mM Tris-HCl (pH 7.5), 2.5mM $MgCl_2$ and 0.1mM $CaCl_2$ to remove co-extracted genomic DNA. RNA was then column-purified using the RNeasy kit (Qiagen, Valencia, CA) as per manufacturer's instructions. Poly(A)-enriched RNA was isolated from purified total RNA using the Oligotex mRNA kit (Qiagen, Valencia, CA). Genomic DNA (gDNA) was isolated from young leaves using the Dneasy kit (Qiagen, Valencia, CA).

Isolation of *EgCesA* class specific region II from cDNA

First-strand cDNA was synthesized from 1 µg total RNA using ImpromII reverse transcriptase (Promega, Madison, WI). Two-step reverse transcription polymerase chain reaction (RT-PCR) was performed with degenerate primers flanking the cDNA sequence encoding the class specific region II (CSRII) of the plant cellulose synthases, employing a modification of the approach described by Liang and Joshi (2004). Degenerate primers specific for the amino acid motifs CYVQFPQ and GWIYGS flanking the CSRII were designed using an *Eucalyptus grandis* codon usage table. The primers specific for the two motifs were: EgCesA-CYVQFPQ (forward primer, TGYTATGTKCARTTCCCWC) and EgCesA-GWIYGS-Rev (reverse primer, GANCCATARATCCANCC). Two-step RT-PCR was performed using the thermal cycling conditions exactly as described in Liang and Joshi (2004), but using cDNAs derived from the seven tissues as template. PCR products were cloned using the InsT/Aclone PCR product cloning kit (MBI Fermentas, Hanover, MD) and a number of cloned inserts (5-10) from each RT-PCR reaction were sequenced (60 clones in total). Cycle sequencing reactions were performed using the BigDye cycle sequencing kit (Applied Biosystems, Foster City, CA) and were analysed on an ABI3100 sequencer (Applied Biosystems).

Batches of sequences were analysed using the CAP3 sequence assembly program (Huang and Madan, 1999). Insert sequences were assembled into a number of non-redundant contiguous sequences (contigs), and were assigned provisory identities by similarity searches against the non-redundant protein database in GenBank by BLASTX (Altschul *et al.*, 1990). Phylogenetic analysis of the isolated CSRII sequences was performed by aligning the translated sequences with CSRII portions of 52 CESA proteins from plant species. CESA accession numbers are grouped in Table 1 with the all sequences available at the internet cellulose synthase database http://cellwall.stanford.edu. Multiple sequence alignments were

performed using CLUSTAL W (Thompson *et al.*, 1994). Neighbour-joining trees were constructed using the MEGA2 version 2.1 software (Kumar *et al.*, 2001).

Full-length cDNA isolation

The 3' ends of the *EgCesA* cDNAs were isolated using a rapid amplification of cDNA ends (3' RACE) technique adapted from Frohman *et al.* (1988). The isolated CSRII sequences were used to design sense PCR primers specific to the individual *EgCesA*s. First strand cDNA was synthesized from 1 µg of total RNA, extracted from young leaf or xylem tissue, using a polyT-anchor oligonucleotide (GACCACGCGTATCGATGGCTCAT$_{16}$V) as primer in the reverse transcription reaction. Approximately 50ng of this cDNA was used as template in a PCR reaction using the gene-specific sense primer (CSRII-F, Table 2) and adapter-specific antisense primer (GACCACGCGTATCGATGGCTCA). The following thermal cycling conditions were used for 3' RACE. Initial denaturation at 95°C for 1 minute followed by 12 cycles of 95°C for 30s, 67°C for 30s with a decrease in temperature of 1°C per cycle and 72°C for 120s. This was followed by 30 cycles of 95°C for 30s, 56°C for 30s, 72°C for 120s with an increase in elongation time of 1s per cycle. Final elongation of was performed at 72°C for 30 minutes. The 3' RACE products were cloned and sequenced as described above.

The 5' cDNA regions of the *EgCesA* cDNAs were isolated by 5' RACE using the FirstChoice RLM-RACE kit (Ambion, Austin, TX). 5' RACE-ready cDNA was produced from 250ng poly(A)-enriched RNA, from xylem or young leaf. Nested antisense primers 5R-1 and 5-R2 (Table 2), were designed (facing the 5' region of the cDNA). Primary and secondary 5' RACE PCRs were performed according to the manufacturer's instructions and the products were subsequently cloned and sequenced.

To confirm that the individual cDNA fragments (CSRII fragments, 5' and 3' RACE products) originated from single full-length cDNAs, PCR primers were designed on the 5' and 3' untranslated region (UTR) sequences (FL-F and FL-R, Table 2). These primers were

used to amplify full-length *EgCesA* cDNAs ranging between 3.3 and 3.8 kb, by RT-PCR. Full-length *EgCesA1* through *3* cDNAs were amplified from cDNA synthesised from xylem RNA, whereas *EgCesA4* to *6* cDNAs were obtained from young leaf cDNA. The full-length RT-PCR products were cloned as described above and inserts from at least 3 independent full-length clones per *EgCesA* were entirely sequenced by primer walking. The CESA-encoding open reading frames were translated and the resulting hypothetical protein sequences were aligned with 52 full-length CESA peptide sequences and analysed as described above.

Quantitative RT-PCR

To ascertain the gene expression levels of the *EgCesA*s in seven tissues, two-step quantitative reverse-transcription PCR (qRT-PCR) was performed using a LightCycler instrument (Version 1.2, Roche Diagnostics GmbH). PCR primers designed on the isolated CSRII regions, were used to amplify cDNA fragments (160-300 bp) from individual *EgCesA* family members (CSRII-F and CSRII-R, Table 2). One microgram of total DNaseI-treated and column-purified RNA extracted from the seven tissue types (X, IX, P, C, Yl, F and I) was reverse transcribed into first strand cDNA using ImpromII reverse transcriptase. The LightCycler FastStart DNA Master^{PLUS} SYBR Green I system (Roche) was used for real-time PCR starting with 10ng of first strand cDNAs as template in a standard 20µl reaction as recommended by the manufacturer. All PCR reactions were performed in triplicate. Relative quantification was performed with the LightCycler software (version 3.5.3, Roche) using the second derivative maximum method. For normalizing the *EgCesA* expression levels, we amplified a 220 bp cDNA fragment of a putative *Eucalyptus* ADP ribosylation factor (*EgArf*, GenBank accession AY770746) with primers Arf-F and Arf-R (Table 2). *EgArf* was previously found to be expressed constitutively in *Eucalyptus* woody tissues (Figure 7, Chapter 2). Amplification of *EgArf*, using a serial cDNA dilution series as template, was used

to create a standard curve, based on which the quantifications were performed. The gene expression levels of the individual *EgCesA*s are reported as fold expression level relative to the expression of the normalization gene in the particular tissue, thus allowing direct quantitative comparison between *EgCesA*s in different tissues. Melting curve analysis and agarose gel electrophoresis of the qRT-PCR products were performed to confirm that the individual qRT-PCR products corresponded to single homogenous DNA species. Additionally, qRT-PCR products of each *EgCesA* from young leaves, xylem and flowers were column-purified (QIAquick; Qiagen, Valencia, CA) and directly (i.e. without first cloning the products) cycle-sequenced to confirm that they represented the corresponding *EgCesA* CSRII sequence. As further quality control, we performed PCR with *E. grandis* gDNA using the same primers as used in qRT-PCR to ascertain whether any of the primer pairs spanned introns. Thus, during qRT-PCR any aberrant intron-containing products from gDNA could be distinguished from the shorter cDNA products by melting curve analysis and agarose gel electrophoresis.

RESULTS

Isolation of class-specific region II from seven tissue samples

Fifty six of the 60 sequenced CSRII clones showed significant similarity to CESA or cellulose synthase-like (CSL) protein sequences in GenBank. Seven unique non-redundant contigs were obtained following analysis with the CAP3 algorithm. Of these, six showed significant similarity to CESA proteins with a single contig being similar to CSL proteins. As all seven contigs represented uninterrupted coding regions, the corresponding hypothetical amino acid sequences were derived and aligned with the CSRII portions of 52 higher plant CESA protein sequences. Phylogenetic analysis of the aligned sequences (not shown) revealed a tree topology nearly identical to that resulting from the alignment of the full-length

CESA sequences as depicted in Figure 2, with a single *Eucalyptus* contig grouping with each of the six main clades of higher plant CESAs. Consequently, we concluded that the CSRII contigs originated from six new *Eucalyptus grandis* cellulose synthase catalytic subunit genes, which were designated *EgCesA1* through *EgCesA6*. The overall amino acid identity of the six EgCESA CSRII ranged from 32% to 53%.

The abundance of CSRII clones from specific *EgCesA*s in particular tissues provided a preliminary indication of the tissue-specific gene expression levels of these *EgCesA*s. Class-specific regions from *EgCesA1*, *EgCesA2* and *EgCesA3* were most abundant (together representing over 80% of the clones originating from that tissue) in the group of clones derived from xylem and immature xylem. In contrast, *EgCesA4* and *EgCesA5* CSRII sequences were most numerous in the groups of clones derived from young leaves and flowers. CSRII sequences from *EgCesA1* to *EgCesA5* were evenly distributed in the groups derived from cork, internode and phloem tissues. Only two clones (one from young leaves and one from phloem), out of a total of sixty, contained the CSRII sequence of *EgCesA6*, possibly signifying the low abundance of *EgCesA6* in the tissues analysed.

Sequence analysis of full-length *EgCesA* cDNAs

All six full-length *EgCesA* cDNAs contain an uninterrupted open reading frame (ORF) of over 2.9 kb, with 5' and 3' untranslated regions of variable length as summarised in Table 3 and Appendix A. The 5' and 3' UTR cDNA regions were obtained by RACE, which has been known to generate truncated products (Schaefer, 1995). A number of 5' and 3' RACE product clones per *EgCesA* were therefore sequenced prior to designing primers for full-length cDNA amplification. Despite selecting the longest RACE products, it is possible that the 5' and 3' UTR regions of some of the cDNAs are incomplete. However, the presence of long unbroken ORFs (spanning approximately 3000 bp) flanked on either side by non-coding

UTR sequence in all of the full-length cDNAs, indicates that the entire coding regions of the cDNAs were obtained.

The nucleotide sequence identities of the six full-length *EgCesA* cDNAs compared to each other ranged from 49% to 63%. By contrast, the most similar *Arabidopsis* full-length *CesA* cDNAs exhibited a 60% to 70% identity with the *EgCesA* cDNAs. Compared separately, the 5' and 3' UTR sequences of *EgCesA1* to *EgCesA6* showed no similarity to each other. Relatively low nucleotide sequence similarity of the six full-length cDNAs compared to each other indicated that distinct *EgCesA* family members, rather than allelic variants or gene copies were isolated. The predicted EgCESA protein sizes ranged from 978 residues (110 kilodaltons, kDa) for EgCESA1, to 1097 residues (123 kDa) for EgCESA6 (Table 3). A similar range of sizes (approximately 1000 amino acid residues) has been observed in higher plant CESA proteins (Richmond, 2000). Prediction of translation initiation codons was performed by comparing the cDNA sequences with known CESAs by BLASTX. Only *EgCesA4* contained any ATG triplets in the putative 5' UTR and both of these were in a suboptimal context compared to the predicted start of translation codon according to the ATG_EVALUATOR algorithm (Rogozin *et al.*, 2001). Table 4 illustrates the percentage amino acid identities and similarities of the six EgCESAs compared to each other as well as to the most similar poplar and *Arabidopsis* CESAs. EgCESAs share limited amino acid identity with each other (61-70%), but exhibit significantly greater homology with the most similar *Arabidopsis* (71-86% identity) and poplar (81-90% identity) CESAs.

The predicted EgCESA proteins were examined for the presence of key regions and amino acid motifs characteristic of functional cellulose synthases (Figure 3). The N-terminal section which shows limited conservation is followed by the conserved RING finger domain containing four "CXXC" motifs. The first CSR following the RING finger domain is clearly noticeable as a region of very limited similarity between the six EgCESAs (11-27% residue

identity). The first conserved region, located after the CSRI exhibits over 95% identity between all of the EgCESAs and contains the first two aspartate residues of the "D,D,D,35,QXXRW" glycosyl transferase signature sequence (Saxena *et al.*, 1995). Class-specific region II, is shorter than CSRI and the level of conservation between the six EgCESAs is slightly higher, ranging from 32% to 53% identity. The C-terminal region following the CSRII is highly conserved across all six EgCESAs (>90% identity). Analysis of the EgCESA sequences with the TMAP algorithm (Persson and Argos, 1994), indicated that eight transmembrane regions are predicted in all six putative proteins, two being located at the beginning of the first conserved region and six near the C terminal.

Phylogenetic distribution of *EgCesA*s within the plant *CesA* gene family

Fifty two full-length CESA amino acid sequences from 6 dicot, 4 monocot and a gymnosperm species, were contrasted with the six predicted EgCESA proteins (Figure 2). Similar previous comparisons have shown that higher plant CESAs group into 6 distinct clades containing members associated with either primary or secondary cell wall development (Holland *et al.*, 2000; Vergara and Carpita, 2001; Liang and Joshi, 2004; Samuga and Joshi, 2004; Nairn and Haselkorn, 2005). Our results were concordant with these studies as six distinct clades (each with 100% bootstrap support from 10,000 replicates) were evident: three containing CESAs known to act in secondary cell wall deposition and three with primary cell wall-associated CESAs. Each of the EgCESAs belonged to a different clade and all of the major clades contained a single *Eucalyptus* member (Figure 2). EgCESA1 through 3 belonged to clades containing CESAs known to function during secondary cell wall biosynthesis. The other three EgCESAs grouped with cellulose synthases associated with primary cell wall synthesis. CESAs from monocot species grouped separately from dicot CESA, within the major clades. With the exception of EgCESA6, all EgCESAs grouped within the dicot subgroups. Interestingly, EgCESA6 was most similar to PtrCESA6 described

by Samuga and Joshi (2004), and the *Eucalyptus* as well as *Poplar* proteins exhibited greater similarity to monocot CESAs than to any other dicot CESAs within their clade. This, to our knowledge, is only the second report of a monocot-like CESA from a dicot species, with both originating from forest tree species. The dicot subgroup within the clade contains the poplar PtrCESA7 (Samuga and Joshi, 2004), but we did not identify a *Eucalyptus* ortholog of *PtrCesA7*.

Expression of *EgCesA*s in diverse *Eucalyptus* tissues

The expression levels of the newly isolated *EgCesA*s were determined in seven *Eucalyptus grandis* woody as well as herbaceous tissues by qRT-PCR. In total, seven genes were assayed: the six newly isolated *EgCesA*s as well as a putative *Eucalyptus grandis* ADP ribosylation factor gene (*EgArf*) used for data normalization during analysis.

Melting curve analysis after each qRT-PCR reaction confirmed that single, homogenous PCR products were obtained in all cases (results not shown). Furthermore, direct sequencing of qRT-PCR products revealed that the PCR products originated from the expected cDNAs and no cross-amplification between the individual *EgCesA* family members was observed. Contamination of RNA samples with genomic DNA can lead to artefacts in qRT-PCR (Vandesompele *et al.*, 2002). Amplification of gDNA revealed that primers specific to *EgCesA1*, *EgCesA3*, *EgCesA4*, *EgCesA5* and *EgArf* all spanned intron-containing regions (results not shown). The resulting PCR products derived from gDNA were 160-600 bp longer than the cDNA-derived products. None of the gDNA-derived amplicons were observed after any of the qRT-PCR reactions, suggesting that gDNA contamination of RNA did not occur

Quantification of gene expression in seven diverse tissues revealed that the individual *EgCesA*s were expressed differentially in tissues enriched for cells depositing primary or secondary cell walls (Figure 4). Two groups of co-expressed *EgCesA*s could be

distinguished: the expression patterns of *EgCesA1* through *3* (group S) being distinctly different from those of *EgCesA4* and *EgCesA5* (group P). Tissue-specific transcript abundance of *EgCesA6* was distinctly different from both of these groups.

The highest expression levels observed in any of the tissues were that of group S *EgCesA*s in xylem. Group S transcripts were also abundant in immature xylem and internodes. By comparison, group S genes were expressed at considerably lower levels in the other two woody tissues: phloem and cork. For example, the expression level of *EgCesA3* was 50 times higher in xylem than in phloem and over 20 times higher in xylem than in cork. Group S transcripts were the least abundant in young leaves averaging less than 2% of the level of expression observed in xylem. The mRNA levels were similarly low in stage 2 flowers. Interestingly, the ratio of expression levels of *EgCesA1* : *EgCesA2* : *EgCesA3* remained approximately constant in all of the tissues at 1½ : 1 : 2. A constant expression ratio was also observed by Burton *et al.* (2004) for the three barley *CesA* genes associated with secondary cell wall deposition. Based on the relatively high level of expression in tissues enriched in cells actively depositing secondary cell walls coupled to extremely low transcript abundance in tissues lacking secondary xylem formation, it is very likely that *EgCesA1*, *EgCesA2* and *EgCesA3* are involved in cellulose biosynthesis in the secondary cell wall.

In contrast to the highly variable expression levels of group S, the transcript levels of Group P genes (*EgCesA4* and *EgCesA5*) were more uniform. In the four woody tissues, group P mRNA levels appeared to be constant and approximately half of that seen in young leaves, flowers and internodes (where the highest level of group P transcripts was observed). A certain base level of group P gene expression seems to exist in all tissues and is elevated where active cell division and hence primary cell wall deposition is taking place. In internodes, the expression levels of groups S and P were approximately equal. Internodes are undergoing secondary xylem deposition but are also rapidly elongating and contain cells

depositing walls of both types. Notably, even in internodes, the expression of group P genes never reached the transcript levels observed for group S genes in xylem. In young leaves, the abundance of group P mRNA was between 50 and 15 times greater than group S transcripts. It is evident that group P transcripts are upregulated in tissues enriched for cells laying down primary cell-walls but are expressed at significantly lower levels than group S *EgCesA*s overall.

Although the presence of *EgCesA6* transcripts could be detected in all tissues by qRT-PCR, the levels of gene expression were extremely low. In all tissues, the normalization transcript *EgArf*, was at least three orders of magnitude more abundant than *EgCesA6*. Similarly, *PtrCesA6* (the most similar gene to *EgCesA6*) was reported to have very low levels of expression in cells depositing primary and secondary cell walls in *Populus tremuloides* (Samuga and Joshi, 2004).

DISCUSSION

Thorough information about the *CesA* gene family is available from only a small number of plant species, including maize (Holland *et al.*, 2000; Appenzeller *et al.*, 2004), barley (Burton *et al.*, 2004), *Arabidopsis* (Richmond, 2000; Hamann *et al.*, 2004) and poplar (Djerbi *et al.*, 2004; Joshi *et al.*, 2004). Although certain key characteristics, such as the requirement of three different CESAs for the formation of a functional cellulose synthase complex, appear to be evolutionarily conserved in woody and herbaceous plants, details from additional forest tree species are required to further clarify the functional divergence of *CesA* family members. To this end, we focused on the isolation and comprehensive expression analysis of the *CesA* gene family from the commercially important hardwood *Eucalyptus grandis*. The approach described here, which utilizes relatively simple PCR techniques, permitted the rapid isolation of multiple *CesA*s and could be adapted to any system, especially when cDNA libraries or

genome sequence are not available.

Previous work by Vergara and Carpita (2001), suggested that the structure of the *CesA* phylogeny remains constant whether it is derived from full-length sequences or from the CSRII of the *CesA*s. This conclusion was also evident from our analysis of the full-length or CSRII amino acid sequences of the six EgCESAs compared to 52 plant CESAs. The class-defining nature of the CSRII allowed us to use this cDNA region for the design of family member-specific primers facilitating the isolation of the 5' and 3' cDNA regions as well as diagnostic qRT-PCR, and was vital in the isolation of the full-length *EgCesA* cDNAs. Although the CSR is clearly useful in discriminating partial *CesA* sequences, it is likely that it plays an important role in determining the unique functions of the *CesA* family members in plants (Vergara and Carpita, 2001).

All six *EgCesA* cDNAs isolated in this study are predicted to encode complete CESA proteins which exhibit all of the key characteristics of functional higher plant cellulose synthase catalytic subunits (Pear *et al.*, 1996). Sequence comparison of the *EgCesA*s to each other, as well as to other plant CESAs (Figure 2, Table 4), suggested that the cDNAs represent dissimilar family members. Based on sequence homology and comparison of expression data, it was possible to identify genes from other plant species orthologous to the six new *EgCesA*s on an individual basis, and thus infer the putative functions of the *EgCesA*s.

EgCesA1 is predicted to encode a protein of 978 residues and was found to be expressed mainly in tissues enriched in cells undergoing secondary cell wall deposition, especially xylem and immature xylem, whilst being weakly expressed in tissues lacking cells undergoing secondary cell wall biogenesis. The *Arabidopsis* gene most similar to *EgCesA1* is *AtCesA8*, a mutation of which causes the *irx1* phenotype (Taylor *et al.*, 2000). *AtCesA8* is one of the three *AtCesA*s required for cellulose biosynthesis in the secondary cell wall (Taylor *et al.*, 2003), and is expressed primarily in *Arabidopsis* stems (Hamann *et al.*, 2004). The most

similar poplar gene to *EgCesA1* is *PtrCesA1*, previously shown to be strongly expressed in secondary xylem of poplar (Wu *et al.*, 2000).

EgCesA2, was also found to be similar to *CesA*s known to be involved in the formation of secondary cell walls: *AtCesA4* and *PtrCesA3*. *AtCesA4* has been recently been described as the gene mutated in the *Arabidopsis* line *irx5*, which exhibits collapsed xylem elements (Taylor *et al.*, 2003). Like *AtCesA8*, *AtCesA4* is expressed mostly in stems but is weakly expressed in tissues lacking secondary xylem such as young leaves and flowers (Hamann *et al.*, 2004). Similar regulation of the poplar ortholog, *PtrCesA3*, was described by Kalluri and Joshi (2004), who found that *PtrCesA3* was upregulated in differentiating secondary xylem elements but was weakly expressed in primary-walled cells, such as ray parenchyma. We found *EgCesA2* to be expressed mostly in xylem, albeit at slightly lower levels than the two other secondary cell wall-related genes, *EgCesA1* and *EgCesA3*.

EgCesA3 transcript abundance in xylem was the highest observed for all of the genes assayed in any of the tissues. Accordingly, it is the third *Eucalyptus* cellulose synthase gene exhibiting similarity to other secondary cell wall-associated *CesA*s, namely *AtCesA7* and *PtrCesA2*. The irregular xylem phenotype of the mutant *irx3* is caused by a mutation in *AtCesA7* (Taylor *et al.*, 1999), which is the third *AtCesA* required for cellulose biosynthesis in the secondary wall (Taylor *et al.*, 2003). *PtrCesA2*, described by Samuga and Joshi (2002), was isolated from a xylem cDNA library and gene expression analysis revealed that the gene was highly upregulated in cells undergoing secondary xylem deposition.

When compared to sequences from plants, *EgCesA4* grouped with *CesA*s associated with primary cell wall biosynthesis, including the poplar *PtrCesA5* and *Arabidopsis AtCesA3*. *AtCesA3* was identified as the disrupted gene in three distinct mutant lines: *ixr1* (Scheible *et al.*, 2001), *cev1* (Ellis *et al.*, 2002) and *eli1* (Cano-Delgado *et al.*, 2003), and was shown to be one of the most strongly expressed *AtCesA*s, particularly in tissues undergoing rapid cell

division such as young leaves (Hamann *et al.*, 2004). Peculiarly, *PtrCesA5* (characterised by Kalluri and Joshi (2003)), was reported to be upregulated in developing xylem compared to young leaf tissue. We found that *EgCesA4* is approximately three-fold more abundant in unfolding leaves than it is immature xylem. Additionally, *EgCesA4* was expressed in young leaves at levels at least one order of magnitude greater than any of the *EgCesA*s associated with secondary cell wall deposition, strongly implicating *EgCesA* as being involved in primary cell wall biosynthesis. The discrepancy between the expression profiles of the putative poplar and *Eucalyptus* orthologs is intriguing, suggesting differences in the regulation of the two genes which encode proteins that are 85% identical.

EgCesA5 was similar to *CesA*s that play a role in the biosynthesis of cellulose in the primary wall - *AtCesA1* and *PtrCesA4*. A single amino-acid residue substitution in *AtCesA1* causes the temperature-sensitive mutant *rsw1*, which exhibits reduced cellulose synthesis and rosette disassembly (Arioli *et al.*, 1998). Morphological analysis of *rsw1* plants indicated that cells in organs undergoing primary cell wall deposition (including roots, hypocotyls and anthers) are severely affected (Williamson *et al.*, 2001). The expression level of *AtCesA1* in wild-type *Arabidopsis* was greatest in tissues undergoing rapid cell division and rich in primary walls (Hamann *et al.*, 2004). Recently, Kalluri and Joshi (2004) demonstrated by *in situ* hybridisation that *PtrCesA4* is expressed in woody as well as herbaceous poplar tissues and is most abundant in cells undergoing primary cell wall development. Preferential expression of *EgCesA5* in tissues that are rich in actively dividing and elongating cells, suggested that it also encodes a cellulose synthase involved in the synthesis of the primary wall.

The final full-length cDNA isolated, *EgCesA6*, was found to be most similar to *PtrCesA6* but shared significantly lower identity with the most similar *Arabidopsis* gene, *AtCesA2*. Expression levels of *PtrCesA6* were found to be very low in all poplar tissues

assayed by *in situ* hybridisation (Samuga and Joshi, 2004). Similarly, we found evidence of extremely low *EgCesA6* expression levels in all *Eucalyptus* tissues assayed. Possibly, these monocot-like *CesA*s play a minor role in cellulose biosynthesis or the expression observed is the result of a background level of promoter activity.

To summarise, we found two distinct groups of expressed *EgCesA*s in *Eucalyptus* - involved in either primary or secondary cell wall biosynthesis. Gene expression analysis of *EgCesA1* through *3* suggested that these three genes are co-regulated and are specifically expressed in tissues enriched for cells undergoing secondary cell wall biosynthesis (Figure 4). Additionally, they are all highly similar to *CesA*s associated with secondary wall formation from monocot and dicot species (Figure 2). However, the sequence similarity of *EgCesA1* through *3*, when compared to each other, is limited. It can therefore be concluded that *EgCesA1* to *3* are three *Eucalyptus CesA* family members associated with the synthesis of the secondary cell wall. The second group of *EgCesA*s, which includes *EgCesA4* and *EgCesA5*, is linked with the synthesis of the primary cell wall as concluded from homology and expression analyses. Expression of these two genes is significantly lower in any of the tissues than the expression of *EgCesA1* to *3* in xylem. This may reflect the fact that secondary cell walls can be approximately one hundred fold thicker than the primary cell wall and contain a significantly greater proportion of cellulose (Mellerowicz *et al.*, 2001; Plomion *et al.*, 2001), requiring a considerable investment of the cellular resources into cellulose production. Separate from these two groups lies *EgCesA6*, the function of which due to extremely low expression levels in all tissues analysed, remains to be reliably postulated.

Future analysis of the *Eucalyptus CesA* gene family will require that the expression profiles of the newly characterised *EgCesA*s be analysed by *in situ* techniques in order to determine the specific cell types in which they are expressed. Additionally, complementation of *Arabidopsis CesA* mutants by the heterologous expression of their putative *Eucalyptus*

orthologs, would provide further functional evidence in addition to sequence similarity and expression data.

ACKNOWLEDGEMENTS

The authors would like to thank Nicky Creux for technical assistance with qRT-PCR. This work was supported by funding provided by Mondi Business Paper South Africa and Sappi Forestry, through the Wood and Fibre Molecular Genetics Programme. Additional funding was provided by the Technology and Human Resources for Industry Programme (THRIP grant no 2546) and the National Research Foundation of South Africa (NRF grant no 5391).

FIGURES

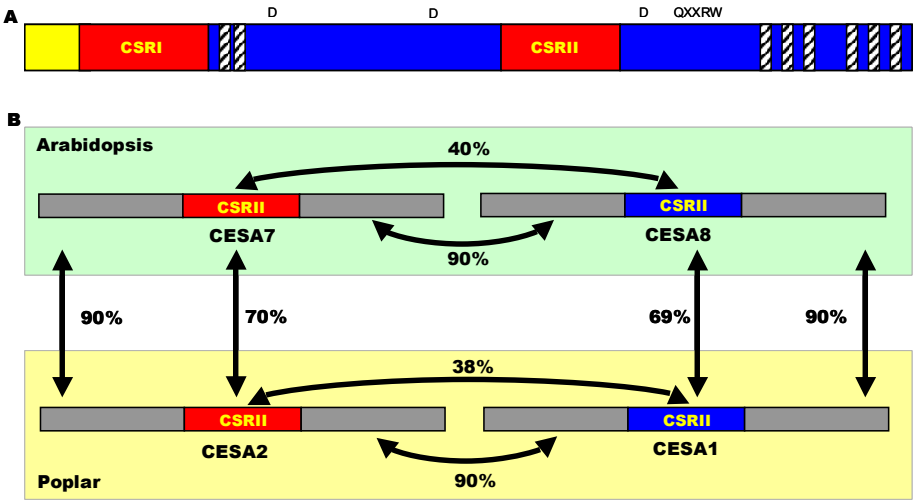


Figure 1.

A. Simplified schematic representation of the major regions of plant CESA proteins. N-terminal RING finger motif is shown in yellow, class-specific regions in red and conserved regions in blue. Putative transmembrane regions are shown as hatched rectangles. Conserved catalytic aspartate residues are marked "D" above the figure together with the motif "QXXRW". **B**. Approximate amino acid identities of conserved and class-specific regions of two *Arabidopsis* CESAs and their poplar orthologs. Two simplified CESA family members from *Arabidopsis* are shown on a green background with their poplar orthologs shown directly below on a yellow background. Grey blocks represent conserved regions while CSRII regions from orthologs are shown in blue or red. Curved arrows signify amino acid identity comparison between regions of two different CESAs from the same species. Vertical arrows indicate amino acid identity comparison between the corresponding regions of orthologs from the two species.

Figure 2

Unrooted neighbour-joining tree derived from the alignment of the deduced amino acid sequences of EgCESA1-6 with 52 full-length CESA protein sequences (Table 1). 10,000 bootstrap replicates were conducted and only branches with support of 80% or greater were considered for the development of the tree. Clades containing CESAs associated with primary cell wall synthesis are denoted on a green background while those linked to secondary cell wall synthesis are shown on a yellow background. CESAs from dicot species are labelled with a red circle while those from monocots are marked with a blue circle. The CESA from gymnosperm species *Pinus radiata* is marked with a black circle. The six new *Eucalyptus grandis* CESAs are highlighted with black ellipses. Species names were abbreviated – At: *Arabidopsis thaliana*, Eg: *Eucalyptus grandis*, Gh: *Gossypium hirsutum*, Hv: *Hordeum vulgare*, Mt: *Medicago truncatula*, Os: *Oryza sativa*, Pr: *Pinus radiata*, Ptr: *Populus tremuloides*, St: *Solanum tuberosum*, Ta: *Triticum aestivum*, Ze: *Zinnia elegans*, Zm: *Zea mays*.

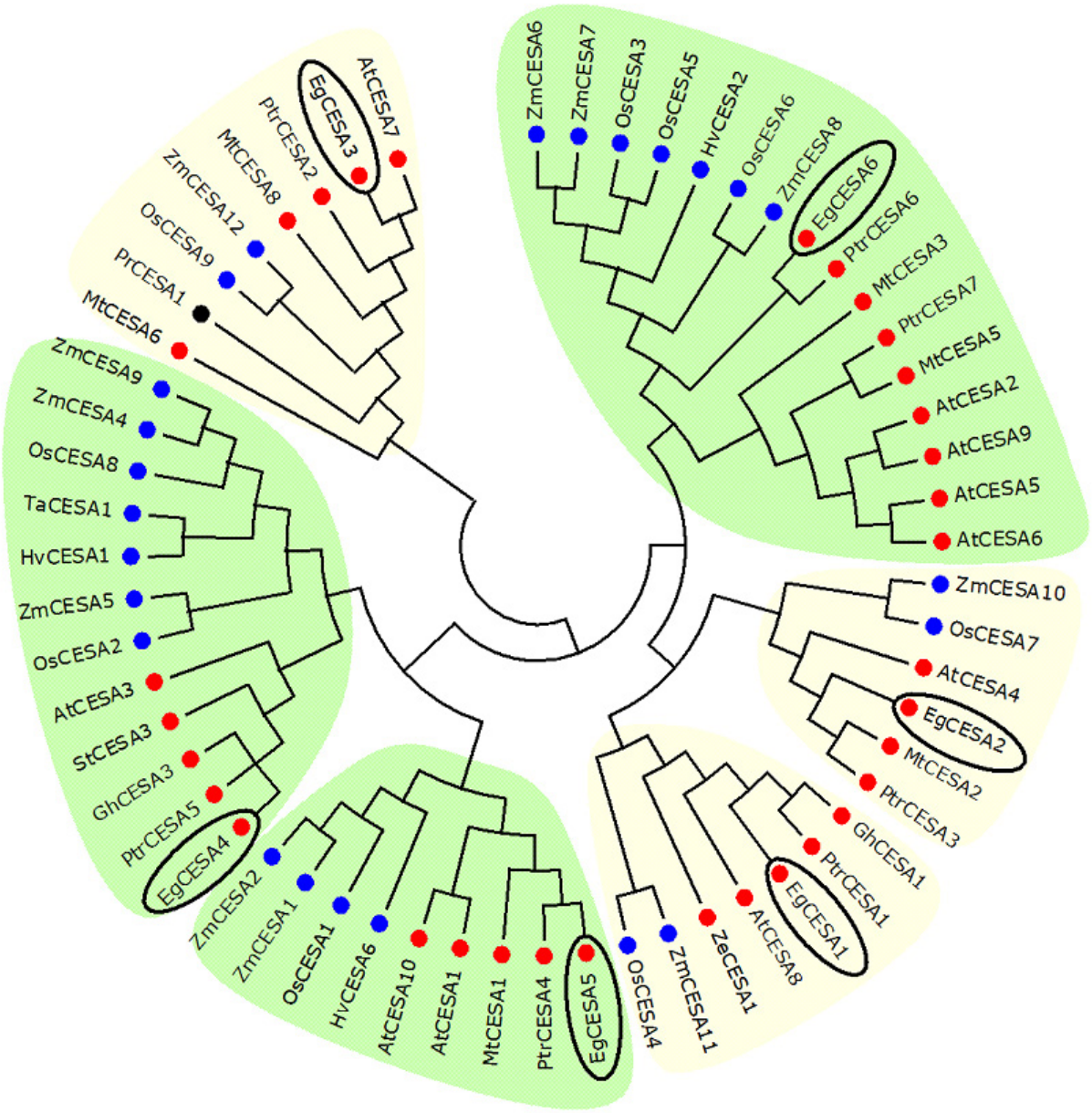
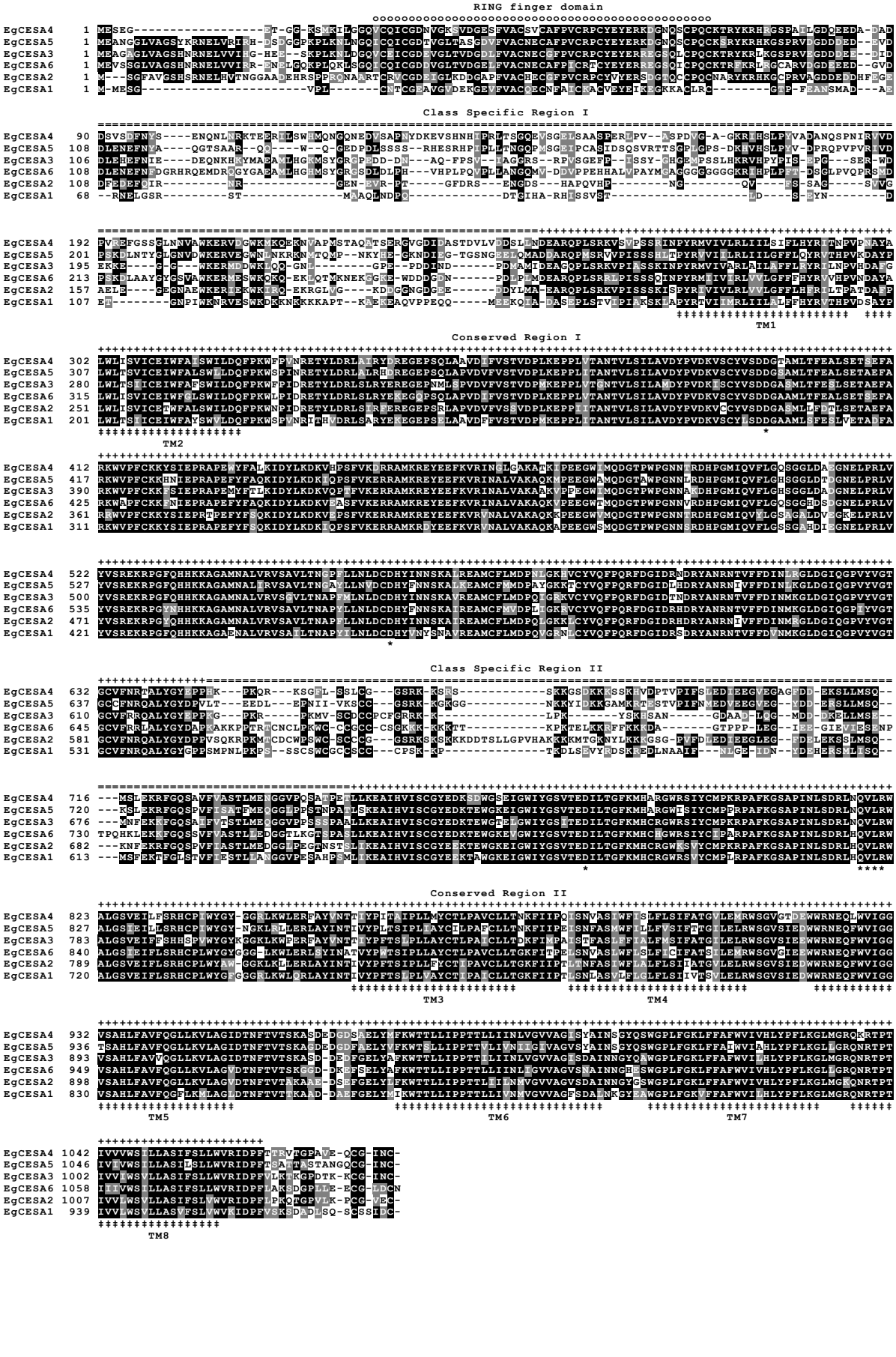

ZmCESA6
ZmCESA7
OsCESA3
OsCESA5
HvCESA2
OsCESA6
ZmCESA8
EgCESA6
PtrCESA6
MtCESA3
PtrCESA7
MtCESA5
AtCESA2
AtCESA9
AtCESA5
AtCESA6
AtCESA7
EgCESA3
PtrCESA2
MtCESA8
ZmCESA12
OsCESA9
PrCESA1
MtCESA6
ZmCESA9
ZmCESA4
OsCESA8
TaCESA1
HvCESA1
ZmCESA5
OsCESA2
AtCESA3
StCESA3
GhCESA3
PtrCESA5
EgCESA4
ZmCESA10
OsCESA7
AtCESA4
EgCESA2
MtCESA2
PtrCESA3
GhCESA1
PtrCESA1
EgCESA1
AtCESA8
ZeCESA1
ZmCESA11
OsCESA4
EgCESA5
PtrCESA4
MtCESA1
AtCESA1
AtCESA10
HvCESA6
OsCESA1
ZmCESA1
ZmCESA2

Figure 3

Amino acid residue alignment of the six predicted EgCESA protein sequences. Identical conserved residues at a particular site are shown on a black background. Non-identical conserved residues with similar properties are highlighted on a grey background. Important general CESA regions are highlighted above the sequence: RING finger domain, two class specific regions, two conserved regions. The four essential catalytic motifs (three aspartate residues and the "QXXRW" motif) are indicated below the sequence with asterisks. The eight predicted transmembrane regions (TM1-TM8) are indicated below the sequences.


RING finger domain
Class Specific Region I
TM1
Conserved Region I
TM2
Class Specific Region II
Conserved Region II
TM3
TM4
TM5
TM6
TM7
TM8
EgCESA4
EgCESA5
EgCESA3
EgCESA6
EgCESA2
EgCESA1

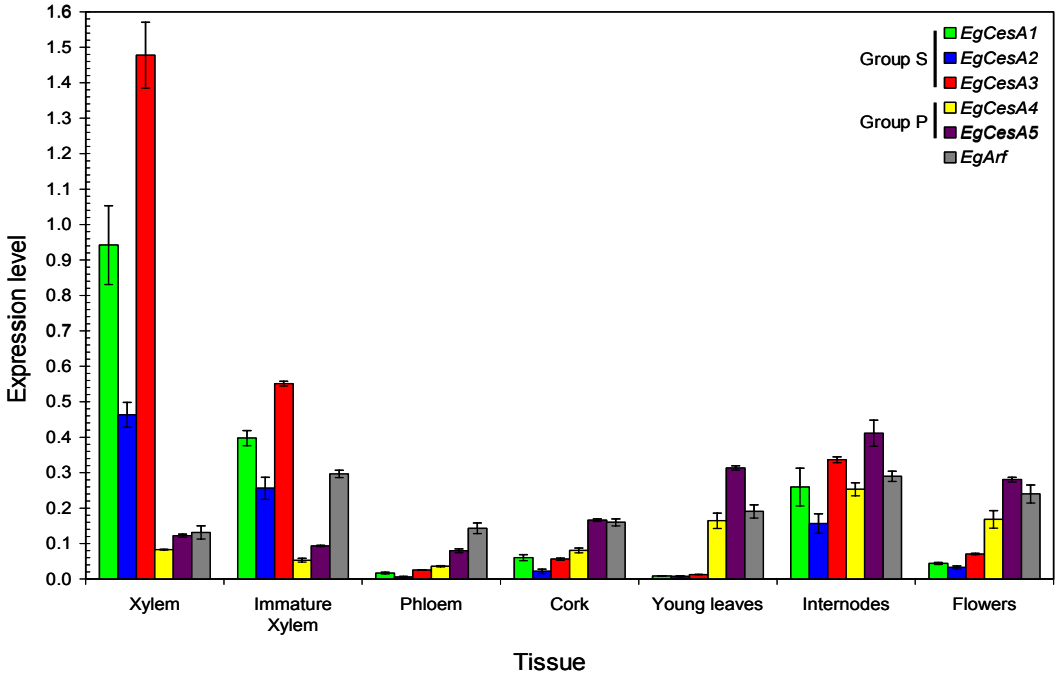


Figure 4

Tissue-specific gene expression levels of *EgCesA1* through *5* relative to the expression level of *EgArf*. Vertical axis represents units of fold expression level relative to the normalization gene *EgArf*. Expression level of *EgArf* is shown in arbitrary units for comparison of its abundance across the seven tissues. Although *EgCesA6* expression was detected in all tissues and was quantifiable in certain tissues, it could not be effectively displayed on this scale due the extremely low relative abundance (approximately 3 orders of magnitude lower than *EgArf*). Error bars denote the standard deviations of three replicate experiments.

TABLES

Table 1

GenBank accessions of full-length *CesA* sequences used in phylogenetic analyses

Species	*CesA*	Genbank accession/ source
Arabidopsis thaliana	*AtCesA1*	AF027172
	AtCesA2	AF027173
	AtCesA3	AF027174
	AtCesA4	AF458083
	AtCesA5	NM_121024
	AtCesA6	NM_125870
	AtCesA7	AF088917
	AtCesA8	AF267742
	AtCesA9	NM_127746
	AtCesA10	NM_128111
Gossipium hirsutum	*GhCesA1*	U58283
	GhCesA3	AF150630
Hordeum vulgare	*HvCesA1*	AY483150
	HvCesA2	AY483152
	HvCesA6	AY483155

[a]*Medicago truncatula*	*MtCesA1*	http://cellwall.stanford.edu
	MtCesA2	
	MtCesA3	
	MtCesA5	
	MtCesA6	
	MtCesA8	
Oryza sativa	*OsCesA1*	AAU44296
	OsCesA2	AAP21426
	OsCesA3	BAD30574
	OsCesA4	AK100475
	OsCesA5	AC104487
	OsCesA6	XM_477282
	OsCesA7	NM_196933
	OsCesA8	XM_477093
	OsCesA9	AK121170
Pinus radiata	*PrCesA1*	AY639654
Populus tremuloides	*PtrCesA1*	AF072131
	PtrCesA2	AY095297
	PtrCesA3	AF527387
	PtrCesA4	AA025536
	PtrCesA5	AY055724
	PtrCesA6	AY196961
	PtrCesA7	AY162180
Solanum tuberosum	*StCesA3*	AY221087
Triticum aestivum	*TaCesA1*	AB158407

Zinnia elegans	*ZeCesA1*	AF323039
Zea mays	*ZmCesA1*	AF200525
	ZmCesA2	AF200526
	ZmCesA4	AF200528
	ZmCesA5	AF200529
	ZmCesA6	AF200530
	ZmCesA7	AF200531
	ZmCesA8	AF200532
	ZmCesA9	AF200533
	ZmCesA10	AY372244
	ZmCesA11	AY372245
	ZmCesA12	AY372246

[a] Full-length *CesA* sequences from alfalfa (*Medicago truncatula*), were obtained from the cellulose synthase family database located at http://cellwall.stanford.edu

Table 2

PCR primers used in full-length cDNA isolation and expression profiling

Gene	Primer Name	Sequence (5'→3')
EgCesA1	CSRII-F	TCCAAGGTTGAAGATGGCGGCATT
	CSRII-R	CGATCGCAGTGATCGATATG
	FL-F	CTCATTGGGTCGCGAGAAGAT
	FL-R	ATTCCATGCATCGCACATTC
	5R-1	GTTGCACTCTTGACAAGCCACGAAGAC
	5R-2	CAGCCTCTCCGCAAGTGTTGCACAG
EgCesA2	CSRII-F	ACATGTGATTGCTGGCCTTC
	CSRII-R	TTCTGAGACATGAGCGATGA
	FL-F	TAGCAAGCACCTCTCTCGTA
	FL-R	GAGACGGCGTGTTGAATGAA
	5R-1	CAACGGCCAGGATTGAGAGGACAG
	5R-2	GCAGCATCGAAGCGCCATCATCAG
EgCesA3	CSRII-F	CTGTGATTGCTGCCCATGCT
	CSRII-R	CGCCACCCTGTTCCATCAAA
	FL-F	CTCCCATGGAAGCCGGAGCTG
	FL-R	GGACATTCTCGCCTTGTGATAC
	5R-1	CGTGCTCGAGATCATCAATGTCTT
	5R-2	TTGTGCTTGTTCTGCTCATCTTCA

EgCesA4	CSRII-F	AGCCAAAGCAGAGAAAGTCA
	CSRII-R	ATAAGCGTAGAGGCCACAAA
	FL-F	GTTCTATCCGGTCAAGATCG
	FL-R	CTCTCACAGCCTATCTATCC
	5R-1	CACTCACACTATCATCAGCATCAG
	5R-2	GATGCGCTCTTCAGTCTTCCGGTT
EgCesA5	CSRII-F	GGGAAGGGTGGCAATAAGAA
	CSRII-R	AACAGGAGACTGACCGAATC
	FL-F	GCGAAGAAGAACTCGGTCTC
	FL-R	CTTCCCGGACGAAAGTATTG
	5R-1	CCTCATCATCATCTCCGTCAACTC
	5R-2	ACTGTTGCCTTGCAGCACTGGTTC
EgCesA6	CSRII-F	CCACCACCTTTGGAAGGTAT
	CSRII-R	GGCTCGTGCCTTTCAGTGTT
	FL-F	TCGCTAAGAAGGGCTGAA
	FL-R	GTCAAATGGAGAGGCGGAGT
	5R-1	CAAGTCATCCACACCATCCTCTTC
	5R-2	TGGCGATCCATCTCTTGCCTATGC
EgArf	Arf-F	TTCTGGTGCCATGCTGAGAA
	Arf-R	GATGCTGTGTTGCTCGTCTT

Table 3

Comparison of *EgCesA* cDNA regions isolated

[a]cDNA	Total length (bp)	5' UTR (bp)	3' UTR (bp)	ORF (bp)	Predicted protein size (residues/ kDa)
EgCesA1	3341	112	295	2934	978/ 110
EgCesA2	3471	93	243	3135	1045/ 118
EgCesA3	3452	33	299	3120	1040/ 117
EgCesA4	3782	178	364	3240	1080/ 121
EgCesA5	3712	109	348	3225	1085/ 122
EgCesA6	3782	47	444	3291	1097/ 123

[a] The annotated *EgCesA* cDNA sequences are present in Appendix A, in GenBank format.

Table 4

Amino acid residue identity/ similarity of EgCESA proteins compared with each other and the most similar *Arabidopsis thaliana* and *Populus tremuloides* CESAs.

		EgCESA							
		1	2	3	4	5	6	[a]PtrCESA	[a]AtCESA
EgCESA	1		66/ 77	64/ 76	61/ 73	62/ 73	61/ 73	(1) 83/ 90	(8) 81/ 89
	2			67/ 77	66/ 76	65/ 76	66/ 77	(3) 81/ 89	(4) 80/ 87
	3				68/ 77	66/ 77	69/ 80	(2) 87/ 93	(7) 88/ 93
	4					70/ 81	68/ 79	(5) 88/ 91	(3) 85/ 91
	5						68/ 78	(4) 90/ 95	(1) 86/ 93
	6							(6) 87/92	(2) 71/82

[a] Numbers in brackets indicate the specific CESA from *Arabidopsis* or *Populus tremuloides* referred to e.g. PtrCESA1.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Meyers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. J Mol Biol 215: 403-410.

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

Arioli, T., Peng, L.C., Betzner, A.S., Burn, J., Wittke, W., Herth, W., Camilleri, C., Hofte, H., Plazinski, J., Birch, R., Cork, A., Glover, J., Redmond, J. and Williamson, R.E. 1998. Molecular analysis of cellulose biosynthesis in *Arabidopsis*. Science 279: 717-720.

Blanton, R.L., Fuller, D., Iranfar, N., Grimson, M.J. and Loomis, W.F. 2000. The cellulose synthase gene of *Dictyostelium*. Proc Natl Acad Sci U S A 97: 2391-2396.

Brown, R.M. Jr. and Saxena, I.M. 2000. Cellulose biosynthesis: A model for understanding the assembly of biopolymers. Plant Physiol Biochem 38: 57-67.

Burn, J.E., Hocart, C.H., Birch, R.J., Cork, A.C. and Williamson, R.E. 2002. Functional analysis of the cellulose synthase genes *CesA1*, *CesA2*, and *CesA3* in *Arabidopsis*. Plant Physiol 129: 797-807.

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

Cano-Delgado, A., Penfield, S., Smith, C., Catley, M. and Bevan, M. 2003. Reduced cellulose synthesis invokes lignification and defense responses in *Arabidopsis thaliana*. Plant J 34: 351-362.

Chang, S., Puryear, J. and Cairney, J. 1993. A simple and efficient method for isolating RNA from pine trees. Plant Mol Biol Rep 11: 113-116.

Delmer, D.P. 1999. Cellulose biosynthesis: exciting times for a difficult field of study. Annu Rev Plant Physiol Plant Mol Biol 50: 245-276.

Desprez, T., Vernhettes, S., Fagard, M., Refregier, G., Desnos, T., Aletti, E., Py, N., Pelletier, S. and Hofte, 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., Aspeborg, H., Nilsson, P., Sundberg, B., Mellerowicz, E., Blomqvist, K. and Teeri, T.T. 2004. Identification and expression analysis of genes encoding putative cellulose synthases (*CesA*) in the hybrid aspen, *Populus tremula* (L.) x *P. tremuloides* (Michx.). Cellulose 11: 301-312.

Ellis, C., Karafyllidis, I., Wasternack, C. and Turner, J.G. 2002. The *Arabidopsis* mutant *cev1* links cell wall signaling to jasmonate and ethylene responses. Plant Cell 14: 1557-1566.

Fenning, T.M. and Gershenzon, J. 2002. Where will the wood come from? Plantation forests and the role of biotechnology. Trends Biotechnol 20: 291-296.

Frohman, M.A., Dush, M.K. and Martin, G.R. 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci U S A 85: 8998-9002.

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

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

Huang, X. and Madan, A. 1999. CAP3: A DNA sequence assembly program. Genome Res 9: 868-877.

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

Kalluri, U.C. and Joshi, C.P. 2003. Isolation and characterization of a new, full-length cellulose synthase cDNA, *PtrCesA5* from developing xylem of aspen trees. J Exp Bot 54: 2187-2188.

Kalluri, U.C. and Joshi, C.P. 2004. Differential expression patterns of two cellulose synthase genes are associated with primary and secondary cell wall development in aspen trees. Planta 220: 47-55.

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

Kudlicka, K. and Brown, R.M. Jr. 1997. Cellulose and callose biosynthesis in higher plants I. Solubilization and separation of (1→3)- and (1→4)-*ß*-glucan synthase activities from mung bean. Plant Physiol 115: 643-656.

Kumar, S., Tamura, K., Jakobsen, I.B. and Nei, M. 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17: 1244-1245.

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

Lai-Kee-Him, J., Chanzy, H., Muller, M., Putaux, J.L., Imai, T. and Bulone, V. 2002. In vitro versus in vivo cellulose microfibrils from plant primary wall synthases: structural differences. J Biol Chem 277: 36931-36939.

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

Matthysse, A.G., Deschet, K., Williams, M., Marry, M., White, A.R. and Smith, W.C. 2004. A functional cellulose synthase from ascidian epidermis. Proc Natl Acad Sci U S A 101: 986-991.

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

Mueller, S.C. and Brown, R.M. Jr. 1980. Evidence for an intramembrane component associated with a cellulose microfibril-synthesizing complex in higher plants. J Cell Biol 84: 315-326.

Nairn, C.J. and Haselkorn, T. 2005. Three loblolly pine *CesA* genes expressed in developing xylem are orthologous to secondary cell wall *CesA* genes of angiosperms. New Phytol In press.

Pear, J.R., Kawagoe, Y., Schreckengost, W.E., Delmer, D.P. and Stalker, D.M. 1996. Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. Proc Natl Acad Sci U S A 93: 12637-12642.

Persson, B. and Argos, P. 1994. Prediction of transmembrane segments in proteins utilising multiple sequence alignments. J Mol Biol 237: 182-192.

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

Richmond, T. 2000. Higher plant cellulose synthases. Genome Biol 1: 1-6.

Roberts, A.W. and Roberts, E. 2004. Cellulose synthase (*CesA*) genes in algae and seedless plants. Cellulose 11: 419-435.

Roberts, A.W., Roberts, E.M. and Delmer, D.P. 2002. Cellulose synthase (*CesA*) genes in the green alga *Mesotaenium caldariorum*. Eukaryot Cell 1: 847-855.

Rogozin, I.B., Kochetov, A.V., Kondrashov, F.A., Koonin, E.V. and Milanesi, L. 2001. Presence of ATG triplets in 5' untranslated regions of eukaryotic cDNAs correlates with a 'weak' context of the start codon. Bioinformatics 17: 890-900.

Samuga, A. and Joshi, C.P. 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.

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

Saxena, I.M., Brown, R.M. Jr. and Dandekar, T. 2001. Structure-function characterization of cellulose synthase: relationship to other glycosyltransferases. Phytochemistry 57: 1135-1148.

Saxena, I.M., Brown, R.M. Jr., Fevre, M., Geremia, R.A. and Henrissat, B. 1995. Multidomain architecture of β-glycosyl transferases: implications for mechanism of action. J Bacteriol 177: 1419-1424.

Saxena, I.M., Lin, F.C. and Brown, R.M. Jr. 1990. Cloning and sequencing of the cellulose synthase catalytic subunit gene of *Acetobacter xylinum*. Plant Mol Biol 15: 673-83.

Schaefer, B.C. 1995. Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. Anal Biochem 227: 255-73.

Scheible, W.R., Eshed, R., Richmond, T., Delmer, D. and Somerville, C. 2001. Modifications of cellulose synthase confer resistance to isoxaben and thiazolidinone herbicides in *Arabidopsis Ixr1* mutants. Proc Natl Acad Sci U S A 98: 10079-10084.

Scheible, W.R. and Pauly, M. 2004. Glycosyltransferases and cell wall biosynthesis: novel players and insights. Curr Opin Plant Biol 7: 285-295.

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

Taylor, N.G., Howells, R.M., Huttly, A.K., Vickers, K. and Turner, S.R. 2003. Interactions among three distinct CesA proteins essential for cellulose synthesis. Proc Natl Acad Sci U S A 100: 1450-1455.

Taylor, N.G., Laurie, S. and Turner, S.R. 2000. Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in *Arabidopsis*. Plant Cell 12: 2529-2539.

Taylor, N.G., Scheible, W.R., Cutler, S., Somerville, C.R. and Turner, S.R. 1999. The *irregular xylem 3* locus of *Arabidopsis* encodes a cellulose synthase required for secondary cell wall synthesis. Plant Cell 11: 769-779.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. 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.

Vandesompele, J., De Paepe, A. and Speleman, F. 2002. Elimination of primer-dimer artifacts and genomic coamplification using a two-step SYBR green I real time RT-PCR. Anal Biochem 303: 95-98.

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

Williamson, R.E., Burn, J.E., Birch, R., Baskin, T.I., Arioli, T., Betzner, A.S. and Cork, A. 2001. Morphology of *rsw1*, a cellulose-deficient mutant of *Arabidopsis thaliana*. Protoplasma 215: 116-127.

Wong, H.C., Fear, A.L., Calhoon, R.D., Eichinger, G.H., Mayer, R., Amikam, D., Benziman, M., Gelfand, D.H., Meade, J.H., Emerick, A.W. and *et al.* 1990. Genetic

organization of the cellulose synthase operon in *Acetobacter xylinum*. Proc Natl Acad Sci U S A 87: 8130-8134.

Wu, L.G., Joshi, C.P. and Chiang, V.L. 2000. A xylem-specific cellulose synthase gene from aspen (*Populus tremuloides*) is responsive to mechanical stress. Plant J 22: 495-502.

SUMMARY

Summary

Expression profiling and characterization of wood formation genes in *Eucalyptus*

Martin Ranik

Supervised by ***Dr A.A. Myburg***

Submitted in partial fulfillment of the requirements for the degree ***Magister Scientiae***

Department of Genetics

University of Pretoria

Summary

Eucalyptus trees are capable of generating vast amounts of wood which is a product of the highly structured differentiation of cell layers centred around the vascular cambium. In order to understand xylogenesis (wood formation), it is necessary that the genes which underlie this complex process be isolated and characterised. Previously, very little was known about the types of genes that are expressed in the main woody tissue layers during xylogenesis in the *Eucalyptus* stem. The aim of this study was to develop a high-throughput transcript profiling platform and to use it in generating a global view of the woody stem transcriptome. Additionally, it was ventured that the newly established system be used for the isolation of gene fragments expressed during wood development and culminate in the characterisation of novel genes important in wood formation.

High-throughput cDNA-AFLP expression profiling utilizing a combination of infrared fragment detection and semi-automated fragment quantification was used to profile the expression of 6385 transcript derived fragments (TDFs) in four major woody tissues (immature xylem, mature xylem, phloem and cork). Clustering of gene expression data revealed that approximately a fifth of the TDFs were differentially expressed and several clusters of TDFs with shared tissue-specific expression profiles were identified. Gene

expression data derived from cDNA-AFLP band quantification was confirmed by quantitative RT-PCR. TDFs representing specific clusters were isolated, sequenced and assigned putative identities. A significant proportion of the 71 sequenced TDFs was found to be similar to genes known to play roles in processes associated with wood development – cell wall biosynthesis, cell fate and gene regulation. The technique of cDNA-AFLP performed using automated DNA analysers was found to be a powerful, fast and relatively inexpensive system for the analysis of gene expression coupled to gene discovery in forest trees. By quantifying and clustering TDFs with shared expression profiles, it was possible to isolate cDNA fragments of candidate genes for further characterization without the requirement for excessive sequencing.

One of the isolated TDFs, which was strongly upregulated in xylem, exhibited significant homology to cellulose synthase catalytic subunit genes (*CesA*) that are instrumental in the deposition of cellulose in the walls of all plant cells. In plants these genes comprise a small gene family of dissimilar members with distinct functions. Using degenerate RT-PCR as well as 5' and 3' RACE, six new full-length *Eucalyptus grandis CesA* cDNAs (*EgCesA1* through *6*) were isolated. The *EgCesA* cDNAs all span over 3.3kb and are predicted to encode proteins of 978 to 1097 amino acid residues which possess all of the motifs characteristic of functional *CesA*s. Sequence analysis and phylogenetic comparison to each other and to all currently known full-length plant *CesA* sequences suggested that *EgCesA1* to *6* represent distinct family members rather than allelic variants or paralogs. Expression profiling of the six *EgCesA*s using quantitative RT-PCR revealed that *EgCesA1* to *3* were very strongly expressed in tissues enriched for cells depositing secondary cell walls. By contrast, *EgCesA4* and *EgCesA5* were upregulated in tissues that contain mostly actively dividing cells, which are depositing primary walls, but include few cells that are undergoing secondary cell wall biogenesis. *EgCesA6* was found to be expressed at very low levels in all

tissues assayed. The six isolated *EgCesA*s represent a significant proportion of the *Eucalyptus CesA* gene family and are the first cellulose synthase catalytic subunit genes from *Eucalyptus* to be comprehensively characterised.

The non-isotopic cDNA-AFLP approach optimised during this study can now be applied to more specific questions pertaining to gene expression profiling in trees. With further characterisation, the six newly isolated *EgCesA* genes will significantly contribute towards the understanding of cellulose biosynthesis in forest trees.

APPENDICES

Appendix A: DNA sequence data

Annotated sequences of *EgCesA1* through *6* in GenBank format.

EgCesA1

```
LOCUS       DQ014505               3341 bp    mRNA    linear   PLN 06-JUN-2005
DEFINITION  Eucalyptus grandis cellulose synthase 1 (CesA1) mRNA, complete cds.
ACCESSION   DQ014505
VERSION     DQ014505
KEYWORDS    .
SOURCE      Eucalyptus grandis
  ORGANISM  Eucalyptus grandis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; Myrtales; Myrtaceae; Eucalyptus.
REFERENCE   1  (bases 1 to 3341)
  AUTHORS   Ranik,M. and Myburg,A.A.
  TITLE     Six new cellulose synthase genes from Eucalyptus are linked with
            primary and secondary cell wall biosynthesis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3341)
  AUTHORS   Ranik,M. and Myburg,A.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2005) Department of Genetics, University of
            Pretoria, Pretoria 0002, South Africa
FEATURES             Location/Qualifiers
     source          1..3341
                     /organism="Eucalyptus grandis"
                     /mol_type="mRNA"
                     /db_xref="taxon:71139"
     gene            1..3341
                     /gene="CesA1"
                     /note="EgCesA1"
     5'UTR           1..112
                     /gene="CesA1"
     CDS             113..3049
                     /gene="CesA1"
                     /codon_start=1
                     /product="cellulose synthase 1"
                     /protein_id="AAY60843"
                     /translation="MMESGVPLCNTCGEAVGVDEKGEVFVACQECNFAICKACVEYEI
                     KEGKKACLRCGTPFEANSMADAERNELGSRSTMAAQLNDPQDTGIHARHISSVSTLDS
                     EYNDETGNPIWKNRVESWKDKKNKKKKAPTKAEKEAQVPPEQQMEEKQIADASEPLST
                     VIPIAKSKLAPYRTVIIMRLIILALFFHYRVTHPVDSAYPLWLTSIICEIWFAYSWVL
                     DQFPKWSPVNRITHVDRLSARYEKEGEPSELAAVDFFVSTVDPMKEPPLITANTVLSI
                     LAVDYPVDKVSCYLSDDGAAMLSFESLVETADFARKWVPFCKKYSIEPRAPEFYFSQK
                     IDYLKDKIQPSFVKERRAMKRDYEEFKVRVNALVAKAQKAPEEGWSMQDGTPWPGNNS
                     RDHPGMIQVFLGSSGAHDIEGNELPRLVYVSREKRPGFQHHKKAGAENALVRVSAILT
                     NAPYILNLDCDHYVNYSNAVREAMCFLMDPQVGRNLCYVQFPQRFDGIDRSDRYANRN
                     TVFFDVNMKGLDGIQGPVYVGTGCVFNRQALYGYGPPSMPNLPKPSSSCSWCGCCSCC
                     CPSKKPTKDLSEVYRDSKREDLNAAIFNLGEIDNYDEHERSMLISQMSFEKTFGLSTV
                     FIESTLLANGGVPESAHPSMLIKEAIHVISCGYEEKTAWGKEIGWIYGSVTEDILTGF
                     KMHCRGWRSVYCMPLRPAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPLWYGFGG
                     GRLKWLQRLAYINTIVYPFTSLPLVAYCTIPAICLLTGKFIIPTLSNLASVLFLGLFL
                     SIIVTSVLELRWSGVSIEDWWRNEQFWVIGGVSAHLFAVFQGFLKMLAGLDTNFTVTT
                     KAADDAEFGELYMIKWTTLLIPPTTLLIVNMVGVVAGFSDALNKGYEAWGPLFGKVFF
                     AFWVILHLYPFLKGLMGRQNRTPTIVVLWSVLLASVFSLVWVKIDPFVSKSDADLSQS
                     CSSIDC"
     3'UTR           3050..3341
                     /gene="CesA1"
ORIGIN
```

Appendix A

```
   1 cccgctctcc ctcttctcgt cccgctttct catataaaga agtgaaagaa tacgaggata
  61 ctccacttgg gtatcgccaa gaactcattg ggtcgcgaga agattggcca acatgatgga
 121 atccggggtt cccctgtgca acacttgcgg agaggctgtt ggggttgatg agaaaggcga
 181 ggtcttcgtg gcttgtcaag agtgcaactt cgccatttgc aaggcttgtg tcgaatatga
 241 gattaaggaa ggaaaaaaag cgtgcttgcg ctgtggcact ccatttgaag cgaactcaat
 301 ggctgatgct gagagaaatg aattgggaag tcgatcgaca atggcagctc aactcaatga
 361 tcctcaggac acagggattc atgctagaca catcagcagt gtttctacgt tggatagtga
 421 atacaatgat gagactggga accctatctg gaagaataga gtggagagct ggaaggacaa
 481 aaagaataag aagaagaagg ccccgacgaa ggctgagaaa gaggctcaag ttccaccaga
 541 gcagcagatg gaagagaagc aaattgctga tgcttcagag ccactctcga ccgttattcc
 601 cattgccaaa agcaaactcg caccataccg aaccgtaata ataatgcgat tgatcatttt
 661 ggcacttttc ttccattacc gagtgacaca tccagtcgac agcgcctatc cattgtggct
 721 tacatccatc atatgtgaga tttggttcgc atactcatgg gtgttggatc agttcccgaa
 781 gtggtctccc gtgaatagga ttactcatgt tgacagactg tctgcaaggt atgaaaaaga
 841 aggtgaacct tcagagcttg ctgctgtcga cttcttcgtg agtacagtcg atccaatgaa
 901 agaaccacct ttgatcactg ccaatacagt gctttccatc cttgctgtgg actatcctgt
 961 tgataaagtc tcttgttatc tttctgatga tggtgctgcg atgctctcat tcgagtctct
1021 tgttgaaacg gccgactttg caagaaagtg ggtacctttc tgcaaaaagt attcgattga
1081 gccccgtgcg cctgagttct acttttccca gaagattgac tacttgaagg ataaaattca
1141 gccttctttt gtgaaggaac gaagagcaat gaaaagagac tatgaggagt tcaaagttcg
1201 agttaatgct ttggtagcaa aagcgcagaa agcgccggag gaaggttggt ctatgcagga
1261 tgggacacct tggcctggaa ataactctcg tgaccaccct ggaatgattc aggttttcct
1321 cggaagtagt ggagcccacg acatcgaggg aaatgaactt cctcggctgg tgtatgtttc
1381 cagagagaag agacccggtt tccaacatca taaaaaggcc ggagcggaaa atgctctggt
1441 gagagtatct gcaatcctga caaatgctcc gtacatcctc aaccttgact gcgatcatta
1501 tgtcaactat agtaatgctg ttcgtgaagc aatgtgcttc ctgatggatc ctcaagttgg
1561 tagaaatttg tgctatgtgc agttccctca gagatttgat ggtatcgatc gcagtgatcg
1621 atatgccaac cgaaataccg ttttctttga tgttaacatg aaagggctgg atggaatcca
1681 aggaccagtt tatgtcggga caggatgtgt tttcaatagg caagcacttt acggatacgg
1741 acccccttct atgccgaatt taccaaagcc ctcctcatca tgctcctggt gtggctgctg
1801 ctcttgttgc tgcccttcaa agaagcccac taaagatctc tcagaagttt accgagactc
1861 aaaaagggaa gacctaaatg ccgccatctt caaccttgga gagattgaca attatgatga
1921 gcatgagaga tcaatgctga tctctcagat gagctttgag aaaacttttg gcttgtccac
1981 tgtgtttatt gagtcgacgc tgttggcaaa tggaggagtt cctgaatctg cccatccgtc
2041 tatgctgatc aaagaagcca ttcatgtcat tagctgtggc tatgaagaga agaccgcatg
2101 gggcaaagag atcggttgga tatatggctc ggttactgag gatatcttaa cagggttcaa
2161 gatgcactgc agaggatgga ggtcggtcta ttgcatgccc ttaaggcctg cattcaaggg
2221 atcagcgccc attaatttgt cagacagact gcaccaagtt ctccgatggg ctctgggttc
2281 cgtggaaata ttcttgagta gacattgtcc actttggtat ggatttggcg gaggccgcct
2341 caagtggctt caaagattgg cgtacataaa caccatcgtt tatccattca catccctccc
2401 tcttgtcgct tattgcacaa ttcctgcgat ctgcctcctt actggaaagt tcatcattcc
2461 aacgctttcc aacttggcaa gcgtgctatt tcttggtctt ttcctctcca tcatcgtcac
2521 aagtgtgctc gagctgcgat ggagtggcgt gagcattgag gactggtggc gtaacgagca
2581 gttctgggtg attggaggtg tctcagccca tctctttgcc gtcttccaag gattcctgaa
2641 gatgttagct ggcctagaca ccaacttcac cgtcactacc aaagcagccg acgacgcgga
2701 gtttggtgag ctctacatga tcaagtggac gacgctgctg atacccccga ccactcttct
2761 catcgtgaac atggtgggtg tcgttgccgg gttctccgat gcgctgaaca aaggttatga
2821 ggcgtgggga cccctcttcg gcaaggtctt ctttgcattc tgggtgattc ttcatctcta
2881 tccattcctg aaaggtctca tgggtaggca gaacaggact ccgaccattg tggttctctg
2941 gtcggtgctt ctggcttctg tcttctctct cgtctgggtg aagatcgatc cattcgtgag
3001 caaatccgat gctgacctct cccagagctg cagttctata gattgttgaa ttgcgccagt
3061 gttggcttgt gtttatcaag gatctcaagc tgtttttgca gttttgcgcc tcttgaagat
3121 tgggaaatac cgagtttatg atgttggaaa tttgctaaag aaagggtgat tattgtattc
3181 atgaattgat atgtagaggt gagacatttt ttctcattat cccaaaaatt cctggctttg
3241 tgtacttgta aaatggtacg agaaacaaag agttttccct gtatttcttg aacaattaac
3301 attagaatgt gcgatgcatg gaatcgaaat ctatttttct c
```

EgCesA2

```
LOCUS       DQ014506               3471 bp    mRNA    linear   PLN 06-JUN-2005
DEFINITION  Eucalyptus grandis cellulose synthase 2 (CesA2) mRNA, complete cds.
ACCESSION   DQ014506
VERSION     DQ014506
KEYWORDS    .
SOURCE      Eucalyptus grandis
  ORGANISM  Eucalyptus grandis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; Myrtales; Myrtaceae; Eucalyptus.
REFERENCE   1  (bases 1 to 3471)
  AUTHORS   Ranik,M. and Myburg,A.A.
  TITLE     Six new cellulose synthase genes from Eucalyptus are linked with
            primary and secondary cell wall biosynthesis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3471)
  AUTHORS   Ranik,M. and Myburg,A.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2005) Department of Genetics, University of
            Pretoria, Pretoria 0002, South Africa
FEATURES             Location/Qualifiers
     source          1..3471
                     /organism="Eucalyptus grandis"
                     /mol_type="mRNA"
                     /db_xref="taxon:71139"
     gene            1..3471
                     /gene="CesA2"
                     /note="EgCesA2"
     5'UTR           1..93
                     /gene="CesA2"
     CDS             94..3231
                     /gene="CesA2"
                     /codon_start=1
                     /product="cellulose synthase 2"
                     /protein_id="AAY60844"
                     /translation="MSGFAVGSHSRNELHVTNGGAADEHRSPPRQNAARTCRVCGDEI
                     GLKDDGAPFVACHECGFPVCRPCYVYERSDGTQCCPQCNARYKRHKGCPRVAGDDEDD
                     HFEGEDFEDEFQIRNRGENEVRPTGFDRSENGDSHAPQVHPNGQVFSSAGSVVGAELE
                     GEGNAEWKERIEKWKIRQEKRGLVGKDDGGNGDGEEDDYLMAEARQPLSRKVPISSSK
                     ISPYRIVIVLRLVVLGFFLHFRILTPATDAFPLWLISVICETWFALSWILDQFPKWNP
                     IDRETYLDRLSIRFEREGEPSRLAPVDVFVSSVDPLKEPPIITANTVLSILAVDYPVD
                     KVCCYVSDDGASMLLFDTLSETAEFARRWVPFCKKYSIEPRTPEFYFSQKIDYLKDKV
                     EPSFVKERRAMKREYEEFKVRVNALVAKAQKKPEEGWVMQDGTPWPGNNTRDHPGMIQ
                     VYLGSAGALDVEGKELPRLVYVSREKRPGYQHHKKAGAMNALVRVSAVLTNAPFLLNL
                     DCDHYINNSKAIREAMCFLMDPQLGKKLCYVQFPQRFDGIDRHDRYANRNIVFFDINM
                     RGLDGIQGPVYVGTGCVFNRQALYGYDPPVSQKRPKMTCDCWPSWCSCCCGGSRKSKS
                     KKKDDTSLLGPVHAKKKKMTGKNYLKKKGSGPVFDLEDIEEGLEGFDELEKSSLMSQK
                     NFEKRFGQSPVFIASTLMEDGGLPEGTNSTSLIKEAIHVISCGYEEKTEWGKEIGWIY
                     GSVTEDILTGFKMHCRGWKSVYCMPKRPAFKGSAPINLSDRLHQVLRWALGSVEIFLS
                     RHCPLWYAWGGKLKLLERLAYINTIVYPFTSIPLLFYCTIPAVCLLTGKFIIPTLTNF
                     ASIWFLALFLSIIATGVLELRWSGVSIEDWWRNEQFWVIGGVSAHLFAVFQGLLKVLA
                     GVDTNFTVTAKAAEDSEFGELYLFKWTTLLIPPTTLIILNMVGVVAGVSDAINNGYGS
                     WGPLFGKLFFAFWVIVHLYPFLKGLMGKQNRTPTIVVLWSVLLASIFSLVWVRIDPFL
                     PKQTGPVLKPCGVEC"
     3'UTR           3232..3471
                     /gene="CesA2"
ORIGIN
        1 cagccctagc aagcacctct ctcgtaccgt tcttctcctc ggcgcctccc cctcgcgatc
       61 gtttcccgct cggcccgtgg cctccccgac accatgtccg gcttcgccgt gggctctcac
      121 tcccggaacg agctccatgt cacgaatggt ggcgctgctg acgaacaccg ctctcctccc
      181 cgccaaaacg cggccagaac ctgccgcgtc tgcggcgacg agatcggcct gaaggacgac
      241 ggcgctccgt tcgtcgcctg ccacgagtgc ggcttccccg tctgccgccc ctgctacgtc
      301 tacgagcgca gcgacggcac ccagtgctgc ccccagtgca acgcccgcta caagcgccac
      361 aaagggtgcc cccgggtcgc gggagacgac gaggacgacc acttcgaagg cgaggatttc
      421 gaggacgagt ttcagatcag gaaccgcggc gagaatgaag ttcgccccac cggtttcgat
```

Appendix A

```
 481 cgttcggaga atggggacag tcacgcgccg caagtccatc cgaacggtca ggttttctct
 541 tcggccggaa gcgtcgtcgg cgcggagttg gaaggagaag gcaatgcgga gtggaaggag
 601 aggatcgaga agtggaaaat caggcaagaa aagaggggct tagtgggcaa ggacgatggc
 661 gggaacggcg atggagagga agatgactac ctgatggctg aagctcggca accactttcg
 721 agaaaagtac cgatttcttc gagcaaaata agcccatacc gaattgtcat cgtcctgcgc
 781 ctcgtagtcc taggcttttt cctccatttc cgtatcttaa cccctgcaac tgatgcattc
 841 cctctatggc ttatctcagt tatatgtgaa acatggtttg ccttgtcgtg gattcttgat
 901 caattcccta agtggaaccc gatagacaga gaaacttatt tggatagatt atccataagg
 961 tttgagaggg agggtgagcc cagtcgctta gctcctgtgg atgtgttcgt cagttctgtg
1021 gaccctctta aggaaccacc aataatcact gcaaatactg tcctctcaat cctggccgtt
1081 gattacccgg tggacaaagt ttgttgctat gtatccgatg atggcgcttc gatgctgctt
1141 tttgacactc tctctgaaac tgctgagttt gcgaggaggt gggtcccatt ctgcaagaag
1201 tatagcatcg agccgaggac tccagagttt tacttttctc aaaagattga ttacctgaaa
1261 gataaggtgg agcccagctt tgtgaaggaa cgtagagcca tgaaaagaga gtatgaagag
1321 ttcaaagtga gggtcaatgc attggtggca aaagctcaga aaaaacctga agaaggatgg
1381 gtaatgcaag atggtacccc ctggcctgga aataatacgc gcgatcatcc tggcatgatc
1441 caggtttatt tgggaagtgc tggagcattg gacgtggaag gtaaggagtt gcctcgactt
1501 gtatatgtgt cccgtgagaa gcgacctggt taccagcacc acaagaaggc tggtgcaatg
1561 aatgctctgg ttcgagtgtc ggcagtgcta acaaacgcac ccttcttgtt gaacttggat
1621 tgtgaccact acatcaacaa cagtaaggct atcagggaag ctatgtgttt tctaatggat
1681 ccccaacttg gaaagaagct ttgctatgtt caatttcctc agaggttcga tggcattgat
1741 cgacatgaca gatatgctaa taggaacata gttttctttg atatcaacat gagagggctt
1801 gatgggatac aaggaccagt gtatgttgga actggatgtg tgttcaatcg gcaggcattg
1861 tatgggtatg atcctccagt gtcccaaaag cggccaaaga tgacatgtga ttgctggcct
1921 tcatggtgct cttgttgctg cggtggttca aggaagtcaa agtcaaagaa gaaggatgat
1981 acgagtttgc ttgggcctgt tcatgcgaag aagaaaaaga tgacaggaaa gaactacttg
2041 aagaagaaag ggtctggacc tgtctttgat ctagaagaca ttgaagaagg acttgagggt
2101 tttgatgagc tagaaaaatc atcgctcatg tctcagaaga attttgagaa gcggtttgga
2161 cagtcacctg tattcattgc ctccacacta atggaagatg gtggcttgcc agaagggact
2221 aactccactt cacttattaa ggaagctatc catgtcataa gttgtggcta tgaagagaaa
2281 acagaatggg gcaaagagat tggatggatt tatggctccg ttacagaaga tatcttgaca
2341 ggcttcaaga tgcattgtag aggatggaag tctgtatatt gcatgcccaa aagaccagct
2401 ttcaagggat cagcacctat aaatctgtca gatcgactcc atcaagttct gagatgggct
2461 cttggctccg ttgagatttt cctcagtcgt cattgtcctt tgtggtatgc ttggggagga
2521 aaactcaaac tgcttgagag gcttgcctat atcaacacca ttgtctaccc tttcacttcc
2581 attcctttgc ttttctactg tacaatacct gccgtttgcc ttctcactgg gaaattcatt
2641 atccccacgc tcactaactt tgcgagcata tggttcttgg ccctttttcct atccatcata
2701 gccactggcg tgcttgaact acggtggagt ggtgtcagca tcgaggactg gtggcgtaat
2761 gaacaattct gggtcattgg tggagtatct gcacacctct tcgctgtatt ccaaggcctc
2821 ctcaaggtgc ttgccggagt tgatactaac ttcactgtta cagcaaaggc agccgaggac
2881 agtgagtttg gtgaactcta ccttttcaag tggactaccc ttctcatacc accaaccact
2941 ctaataatct tgaacatggt cggtgtcgtc gccggtgttt cggatgccat aaacaatgga
3001 tacggatcgt ggggccctct gttcgggaag ctcttcttcg ccttttgggt gatcgtccat
3061 ctctaccctt tcctcaaagg tctgatggga aaacagaaca ggacacccac gatcgtggtc
3121 ctttggtccg tacttctcgc ctctattttc tcattggtct gggtccggat cgatccgttc
3181 ctgccgaagc aaaccggtcc agttctcaaa ccgtgtgggg tggagtgctg attctggcgt
3241 cggatttcat tcaacacgcc gtctctccga cctgattaga tgtgtcgctt tacggagctg
3301 tttctttctg tctcttactt gggacatatt gtaatgcact aggggaaatc ttcccgattg
3361 aaatctcttg atcagcatag gttttgcttg aagagtgtgg aactgaaatg tgcaaagtcc
3421 tggttttgaa ctttttgcaa tatattctgc tcaagattaa gcacttgcca a
```

Appendix A

EgCesA3

```
LOCUS       DQ014507               3452 bp    mRNA    linear   PLN 06-JUN-2005
DEFINITION  Eucalyptus grandis cellulose synthase 3 (CesA3) mRNA, complete cds.
ACCESSION   DQ014507
VERSION     DQ014507
KEYWORDS    .
SOURCE      Eucalyptus grandis
  ORGANISM  Eucalyptus grandis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; Myrtales; Myrtaceae; Eucalyptus.
REFERENCE   1  (bases 1 to 3452)
  AUTHORS   Ranik,M. and Myburg,A.A.
  TITLE     Six new cellulose synthase genes from Eucalyptus are linked with
            primary and secondary cell wall biosynthesis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3452)
  AUTHORS   Ranik,M. and Myburg,A.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2005) Department of Genetics, University of
            Pretoria, Pretoria 0002, South Africa
FEATURES             Location/Qualifiers
     source          1..3452
                     /organism="Eucalyptus grandis"
                     /mol_type="mRNA"
                     /db_xref="taxon:71139"
     gene            1..3452
                     /gene="CesA3"
                     /note="EgCesA3"
     5'UTR           1..33
                     /gene="CesA3"
     CDS             34..3156
                     /gene="CesA3"
                     /codon_start=1
                     /product="cellulose synthase 3"
                     /protein_id="AAY60845"
                     /translation="MEAGAGLVAGSHNRNELVVIHGHEESKPLKNLDGQVCEICGDEV
                     GLTVDGDLFVACNECGFPVCRPCYEYERREGSQLCPQCKTRYKRLKGSPRVEGDDDEE
                     DIDDLEHEFNIEDEQNKHKYMAEAMLHGKMSYGRGPEDDDNAQFPSVIAGGRSRPVSG
                     EFPISSYGHGEMPSSLHKRVHPYPISEPGSERWDEKKEGGWKERMDDWKLQQGNLGPE
                     PDDINDPDMAMIDEAGQPLSRKVPIASSKINPYRMVIVARLAILAFFLRYRILNPVHD
                     AFGLWLTSIICEIWFAFSWILDQFPKWFPIDRETYLDRLSLRYEREGEPNMLSPVDVF
                     VSTVDPMKEPPLVTGNTVLSILAMDYPVDKISCYVSDDGASMLTFESLSETAEFARKW
                     VPFCKKFSIEPRAPEMYFTLKIDYLKDKVQPTFVKERRAMKREYEEFKVRINALVAKA
                     AKVPPEGWIMQDGTPWPGNNAKDHPGMIQVFLGHSGGLDADGNELPRLVYVSREKRPG
                     FQHHKKAGAMNALVRVSGVLTNAPFMLNLDCDHYINNSKAVREAMCFLMDPQIGRKVC
                     YVQFPQRFDGIDTNDRYANRNTVFFDINMKGLDGIQGPVYVGTGCVFRRQALYGYEPP
                     KGPKRPKMVSCDCCPCFGRRKKLPKYSKHSANGDAADLQGMDDDKELLMSEMNFEKKF
                     GQSAIFVTSTLMEQGGVPPSSSPAALLKEAIHVISCGYEDKTEWGTELGWIYGSITED
                     ILTGFKMHCRGWRSIYCMPKRPAFKGSAPINLSDRLNQVLRWALGSVEIFFSHHSPVW
                     YGYKGGKLKWPERFAYVNTTIYPFTSLPLLAYCTLPAICLLTDKFIMPAISTFASLFF
                     IALFMSIFATGILELRWSGVSIEEWWRNEQFWVIGGVSAHLFAVVQGLLKVLAGIDTN
                     FTVTSKASDDEDFGELYAFKWTTLLIPPTTILIINLVGVVAGISDAINNGYQAWGPLF
                     GKLFFAFWVILHLYPFLKGLMGRQNRTPTIVVIWSVLLASIFSLLWVRIDPFVLKTKG
                     PDTKKCGINC"
     3'UTR           3157..3452
                     /gene="CesA3"
ORIGIN
        1 ccacctcctc ctcctcctcc tcctcctcct cccatggaag ccggagctgg acttgtcgcc
       61 ggttctcaca accgcaacga gctcgtcgtg attcacggcc atgaggagtc gaagcctttg
      121 aagaacttgg atgggcaagt gtgtgagatc tgtggggatg aggttgggct cacggttgat
      181 ggagatttgt tcgtggcatg caatgagtgc ggatttccgg tttgtcggcc ttgctatgag
      241 tatgagagga gagaagggag ccagttgtgc cctcagtgca agactcgata caagcgtctc
      301 aaagggagcc caagagtgga gggtgatgat gatgaagaag acattgatga tctcgagcac
      361 gaattcaaca ttgaagatga gcagaacaag cacaagtaca tggcagaagc tatgcttcat
      421 gggaagatga gctatggaag aggtcctgag gatgacgata acgctcaatt tccatcagtt
```

Appendix A

```
 481 attgctggtg gcagatcccg acctgttagt ggcgagttcc caatatcatc ttatggtcac
 541 ggagagatgc cctcttccct tcacaaacga gttcatccat atccaatttc tgaacccgga
 601 agtgaaagat gggatgaaaa gaaagaggga gggtggaaag aaagaatgga cgactggaag
 661 ctgcagcagg gcaacctcgg ccctgaacct gatgacatca atgacccgga catggcaatg
 721 atagatgagg cagggcagcc actctccagg aaagtaccaa ttgcatcgag caagatcaac
 781 ccataccgga tggtgatagt tgctcggctt gccatattgg ctttcttcct tcgatacagg
 841 atattgaacc cagtacatga tgcatttggt ctttggttaa catccatcat ctgtgagata
 901 tggttcgctt tctcctggat cctggatcag tttcccaaat ggttccctat tgatcgtgag
 961 acctatcttg atcgcctctc tctcagatat gaaagggaag gtgaacccaa tatgctttct
1021 ccagtagatg tctttgtcag tacagtggat cctatgaaag agcccccact tgtgacagga
1081 aacacagttc tgtcaatatt ggctatggac tatccagtcg ataagatttc ctgctacgtt
1141 tctgatgatg gagcttcaat gctcaccttt gaatctttgt ccgaaaccgc cgaatttgct
1201 cgaaaatggg taccttctg caaaaagttc tctatagagc cccgagcccc tgaaatgtac
1261 ttcactttga agattgatta tctcaaagac aaagtccagc caaccttcgt caaagagcgg
1321 cgagcaatga agagagaata tgaagaattc aaagtgcgga tcaatgcgtt ggttgctaaa
1381 gctgcgaagg tacctccaga gggatggatc atgcaagacg gtacaccatg gccaggaaac
1441 aacgcaaagg atcatccagg catgattcaa gtctttcttg gacacagcgg aggtcttgat
1501 gcagacggaa atgagctccc tcgccttgtt tatgtatctc gtgagaagag gcctggcttc
1561 caacaccaca agaaagctgg tgccatgaat gccttggttc gtgtctctgg tgtgcttaca
1621 aatgctcctt tcatgctgaa cttggattgt gaccactaca taaataacag caaggccgtc
1681 cgcgaggcca tgtgcttctt gatggaccct cagataggaa gaaaggtttg ttacgtccaa
1741 ttccctcaaa gatttgatgg cattgatacg aatgatcggt acgccaacag gaacacagtt
1801 ttctttgata tcaacatgaa aggtctagat ggaatccagg gcccagtata tgtgggcact
1861 ggatgtgtct tccggagaca ggcattatat ggctatgaac cccccaaggg tcctaagcgt
1921 cctaagatgg tcagctgtga ttgctgccca tgctttggcc gccgcaaaaa actcccgaag
1981 tattctaagc atagtgccaa tggggatgct gcagatctgc aaggaatgga tgatgacaag
2041 gagctgttga tgtccgagat gaattttgag aagaagtttg ggcagtcggc aatttttgtt
2101 acctcaactt tgatggaaca gggtggcgta cccccttctt caagtccagc agcactgctg
2161 aaagaagcca tccatgttat cagttgtggg tatgaagata aaactgaatg gggcaccgag
2221 ttgggctgga tttatggttc tatcacagag gatatcctaa caggctttaa gatgcattgc
2281 cgtggttgga ggtccattta ttgtatgcca aagagaccag cattcaaggg atctgctcca
2341 attaatcttt cagataggct caatcaagtg cttcgttggg cacttggttc tgttgagatc
2401 ttttttagcc accacagtcc agtctggtat ggatacaaag gggggaagct taaatggccc
2461 gaacggtttg catatgtcaa cacaaccatc tatcctttta cctccttacc actccttgcc
2521 tactgtactc ttccggccat ttgcttgctg acagacaagt tcatcatgcc agcaattagc
2581 acttttgcca gtcttttctt cattgctctt tttatgtcaa tctttgcaac cggcattctt
2641 gagctgaggt ggagtggagt gagcattgag gaatggtgga gaaatgagca attctgggtc
2701 attggtggtg tgtcagctca cttgtttgct gttgtacaag gtctcctgaa ggtattggct
2761 ggaatagaca ctaatttcac agtgacatcc aaagcgtctg atgacgaaga ctttggcgaa
2821 ctgtacgcat tcaaatggac aacgttgtta atccctccaa ctactatatt gatcatcaac
2881 ctagttggag tagttgccgg aatatctgat gctattaata atggatacca agcatgggga
2941 cctctgtttg ggaaattgtt cttcgccttc tgggtgattc tccatctcta tcccttcctg
3001 aaagggctga tggggagaca gaaccgtaca ccaacaatag tggtcatatg gtcagtactc
3061 ctggcttcca tattctcctt gctgtgggtc cgaattgatc cgttcgtgct aaagaccaag
3121 gggcctgata ccaagaaatg tggaatcaat tgttaattct caaatttcct gcttcagtgt
3181 gtgcatctcc tcaagtacaa ggatatttcg tccagatgaa tttagtgtac agatatgcag
3241 cacagaagtg cttgagattg aataaggata gagtgcaatt ttttgtgtat cacaaggcaa
3301 gaatgtccat ttttgtaatt gaatgatatg tcctctgata agattggaga aatcccaagt
3361 ggatcatact aggctataag cgttttcttc cgagaaatta agtatttaca atatcctatt
3421 tcatactaaa aaaaatcaat cctacatttc tc
```

Appendix A

EgCesA4

```
LOCUS       DQ014508               3782 bp    mRNA    linear   PLN 06-JUN-2005
DEFINITION  Eucalyptus grandis cellulose synthase 4 (CesA4) mRNA, complete cds.
ACCESSION   DQ014508
VERSION     DQ014508
KEYWORDS    .
SOURCE      Eucalyptus grandis
  ORGANISM  Eucalyptus grandis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; Myrtales; Myrtaceae; Eucalyptus.
REFERENCE   1  (bases 1 to 3782)
  AUTHORS   Ranik,M. and Myburg,A.A.
  TITLE     Six new cellulose synthase genes from Eucalyptus are linked with
            primary and secondary cell wall biosynthesis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3782)
  AUTHORS   Ranik,M. and Myburg,A.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2005) Department of Genetics, University of
            Pretoria, Pretoria 0002, South Africa
FEATURES             Location/Qualifiers
     source          1..3782
                     /organism="Eucalyptus grandis"
                     /mol_type="mRNA"
                     /db_xref="taxon:71139"
     gene            1..3782
                     /gene="CesA4"
                     /note="EgCesA4"
     5'UTR           1..178
                     /gene="CesA4"
     CDS             179..3421
                     /gene="CesA4"
                     /codon_start=1
                     /product="cellulose synthase 4"
                     /protein_id="AAY60846"
                     /translation="MESEGETGGKSMKILGGQVCQICGDNVGKSVDGESFVACSVCAF
                     PVCRPCYEYERKDGNQSCPQCKTRYKRHRGSPAILGDQEEDADADDSVSDFNYSENQN
                     LNRKTEERILSWHMQNGQNEDVSAPNYDKEVSHNHIPRLTSGQEVSGELSAASPERLP
                     VASPDVGAGKRIHSLPYVADANQSPNIRVVDPVREFGSSGLNNVAWKERVDGWKMKQE
                     KNVAPMSTAQATSERGVGDIDASTDVLVDDSLLNDEARQPLSRKVSVPSSRINPYRMV
                     IVLRLIILSIFLHYRITNPVPNAYALWLISVICEIWFAISWILDQFPKWFPVNRETYL
                     DRLAIRYDREGEPSQLAAVDIFVSTVDPLKEPPLVTANTVLSILAVDYPVDKVSCYVS
                     DDGTAMLTFEALSETSEFARKWVPFCKKYSIEPRAPEWYFALKIDYLKDKVHPSFVKD
                     RRAMKREYEEFKVRINGLGAKATKIPEEGWIMQDGTPWPGNNTRDHPGMIQVFLGQSG
                     GLDAEGNELPRLVYVSREKRPGFQHHKKAGAMNALVRVSAVLTNGPFLLNLDCDHYIN
                     NSKALREAMCFLMDPNLGKHVCYVQFPQRFDGIDRNDRYANRNTVFFDINLRGLDGIQ
                     GPVYVGTGCVFNRTALYGYEPPHKPKQRKSGFLSSLCGGSRKKSRSSKKGSDKKKSSK
                     HVDPTVPIFSLEDIEEGVEGAGFDDEKSLLMSQMSLEKRFGQSAVFVASTLMENGGVP
                     QSATPETLLKEAIHVISCGYEDKSDWGSEIGWIYGSVTEDILTGFKMHARGWRSIYCM
                     PKRPAFKGSAPINLSDRLNQVLRWALGSVEILFSRHCPIWYGYGGRLKWLERFAYVNT
                     TIYPITAIPLLMYCTLPAVCLLTNKFIIPQISNVASIWFISLFLSIFATGVLEMRWSG
                     VGTDEWWRNEQLWVIGGVSAHLFAVFQGLLKVLAGIDTNFTVTSKASDEDGDSAELYM
                     FKWTTLLIPPTTLLIINLVGVVAGISYAINSGYQSWGPLFGKLFFAFWVIVHLYPFLK
                     GLMGRQKRTPTIVVVWSILLASIFSLLWVRIDPFTTRVTGPAVEQCGINC"
     3'UTR           3422..3782
                     /gene="CesA4"
ORIGIN
        1 tctatccggt caagatcgga ttttgaggaa gttactctcg gatctgtgtt tttactggaa
       61 aacaagttgc ttctaaatgc aacactagag atctctacag cttctgctaa tgccacatca
      121 agttcggaat cagtgaagtc atcctctctt agcatccgag ccaggaggag ctattgcgat
      181 ggagtcggaa ggagaaactg ggggaaagtc aatgaaaatt ctgggtggtc aagtctgcca
      241 gatttgtggt gataacgttg gcaaaagtgt tgatggcgag tcatttgttg cttgcagtgt
      301 ctgtgcattt cctgtctgta ggccatgcta tgagtatgag aggaaagacg ggaatcagtc
      361 atgtcctcaa tgcaaaacca gatacaagag gcacagagga agtccggcta ttcttggtga
      421 ccaagaagaa gatgctgatg ctgatgatag tgtgagcgat ttcaattact cagaaaatca
```

Appendix A

```
 481 aaatctaaac cggaagactg aagaacgcat cttgagttgg cacatgcaga atggacagaa
 541 tgaggatgtg agtgcaccaa actacgataa ggaggtttct cacaaccata ttcctcgact
 601 tacgagtggc caagaggttt ctggggagtt atctgctgct tcacctgaac gcctccctgt
 661 ggcatctcct gatgttggtg ctgggaagcg catccattct ctaccttatg tagccgatgc
 721 taatcaatca cctaacatca gggtggtgga cccagtgcgg gaatttggtt catcaggact
 781 gaacaacgtt gcctggaagg agagagttga tggctggaaa atgaagcaag agaagaatgt
 841 agctccaatg agcactgccc aggctacttc cgaaagagga gttggtgata tcgatgccag
 901 cactgatgta cttgtggatg actctttact gaatgacgaa gccaggcagc ccctttcgag
 961 gaaggtctct gttccatcat caaggataaa tccatataga atggttatag tcctgcggct
1021 tatcattctt tccatctttt tgcactaccg aattacaaat cctgtgccca atgcctatgc
1081 tctgtggtta atatctgtga tatgtgagat ttggtttgca atatcttgga tattggatca
1141 gtttccgaag tggtttccag taaatcgtga aacgtatctc gacagactag ccattaggta
1201 tgacagagaa ggagaaccat cacagttagc tgctgttgac atctttgtca gtactgtcga
1261 cccactcaaa gaacctcctc ttgtgacagc aaataccgta ctctccattc ttgcagttga
1321 ctacccagtt gacaaggtct catgctatgt gtccgatgat ggaactgcaa tgttaacttt
1381 tgaggcccct tctgaaacgt cagagtttgc aaggaaatgg gttccttttct gcaagaagta
1441 tagcattgag cctcgagcac ctgaatggta ctttgctctg aaaattgact acttgaagga
1501 caaagttcat ccatcatttg tcaaagatcg cagagctatg aagagagaat atgaggaatt
1561 taaagtccgt ataaatggac ttggtgccaa ggcaacaaag attcccgagg aaggatggat
1621 catgcaagat gggacaccat ggcctggaaa taacacaaga gaccaccctg gaatgattca
1681 ggttttctta gggcaaagtg gaggtcttga tgctgagggt aatgagctgc cacgtttggt
1741 ctatgtttct cgtgaaaaac gtccgggctt ccaacatcac aagaaagccg gtgctatgaa
1801 tgctcttgtg cgagtctcag ctgtcctcac taatggacct ttcttgttga atcttgattg
1861 tgatcactac ataaacaaca gtaaggcatt gagagaagcc atgtgtttct tgatggaccc
1921 taaccttgga aaacatgtgt gttacgtaca atttccacag agatttgacg gtattgatag
1981 aaatgacaga tacgccaatc gcaatactgt gttctttgat ataaacttga gaggtttgga
2041 tggcatccaa ggtcctgtgt atgtcggaac tgggtgcgtg ttcaatagaa cagccttata
2101 tggttatgaa cctcctcaca agccaaagca gagaaagtca ggattcctgt cttctctctg
2161 tggtggatca cgaaagaaat caagatcaag taaaaaaggc tctgacaaga agaagtctag
2221 caaacatgta gatccaactg tgcctatctt cagcctggaa gatatagaag agggagtgga
2281 aggtgctgga tttgatgatg aaaaatcgtt gctcatgtcg caaatgagcc tagagaagag
2341 gtttggtcag tctgctgtct ttgtggcctc tacgcttatg gagaatggtg gtgttcctca
2401 atctgcaact ccagaaaccc ttcttaagga ggctattcat gtcatcagct gtggttatga
2461 ggataaatca gactggggaa gtgagattgg atggatctat ggatctgtca cagaagatat
2521 tcttacagga ttcaagatgc atgcgcgtgg ttggcggtca atctactgca tgccaaaacg
2581 tcctgcattt aaagggtctg ctcctattaa tctttctgat cggttgaacc aggtgcttcg
2641 atgggctcta ggttctgttg aaattctttt tagccgccat tgccctatct ggtatggata
2701 tggcggtagg cttaaatggc ttgagaggtt tgcatatgtc aacactacca tttatcccat
2761 cactgccata cctcttctga tgtattgtac cctgccagct gtctgcctgc tgacaaacaa
2821 gtttattatt ccacagatta gtaacgttgc cagtatttgg ttcatttcgc tcttcctttc
2881 catttttgcc actggtgtct tggagatgag gtggagtggt gttggaactg atgaatggtg
2941 gaggaatgag caactttggg tgattggtgg tgtgtctgcc catctttttg ccgtcttcca
3001 gggcctgctc aaagtcttgg ctggaattga caccaacttt actgttacat ccaaagcatc
3061 agatgaagat ggggattctg ccgaactcta catgttcaag tggacaactc ttctcatccc
3121 tcccactacg ctcctcatca taaacttggt cggggtcgtg gctgggattt cgtatgccat
3181 caatagtgga tatcagtcgt ggggtccact ttttggtaag ttgttctttg ctttttgggt
3241 gatcgttcac ctgtacccct tcctcaaagg tctgatggga cgccagaaac gcacaccaac
3301 cattgtagtt gtttggtcga tcctcctagc atccattttc tcacttctgt gggtgcgtat
3361 tgatcccttc acgacccgag ttactggacc tgccgtcgag cagtgtggaa tcaactgcta
3421 ataattggac acagatgggg gaactgtaag atccgaaatt ggttactccc cactaagcta
3481 aagttgtgaa ggtgagggta acctttgtct gtcgttagat gtacccacct ggctggctgc
3541 aaagtgcaat cttgtttgtt gtcgattata agatttattt atttgtggat agataggctg
3601 tgagagtttt tgtttcaaag gtcataagtt aatgtgcatt agttggagga tctgtatttt
3661 ggcttccctt ttgactgctg ccaaaacagt taatgtattg ttctcatgaa cagctaatac
3721 attgtatttg ctcgttcaaa tgcattatac ctatcgcatg gtggtgatcc taagcttatt
3781 gg
```

Appendix A

EgCesA5

```
LOCUS       DQ014509               3712 bp    mRNA    linear   PLN 06-JUN-2005
DEFINITION  Eucalyptus grandis cellulose synthase 5 (CesA5) mRNA, complete cds.
ACCESSION   DQ014509
VERSION     DQ014509
KEYWORDS    .
SOURCE      Eucalyptus grandis
  ORGANISM  Eucalyptus grandis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; Myrtales; Myrtaceae; Eucalyptus.
REFERENCE   1  (bases 1 to 3712)
  AUTHORS   Ranik,M. and Myburg,A.A.
  TITLE     Six new cellulose synthase genes from Eucalyptus are linked with
            primary and secondary cell wall biosynthesis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3712)
  AUTHORS   Ranik,M. and Myburg,A.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2005) Department of Genetics, University of
            Pretoria, Pretoria 0002, South Africa
FEATURES             Location/Qualifiers
     source          1..3712
                     /organism="Eucalyptus grandis"
                     /mol_type="mRNA"
                     /db_xref="taxon:71139"
     gene            1..3712
                     /gene="CesA5"
                     /note="EgCesA5"
     5'UTR           1..109
                     /gene="CesA5"
     CDS             110..3367
                     /gene="CesA5"
                     /codon_start=1
                     /product="cellulose synthase 5"
                     /protein_id="AAY60847"
                     /translation="MEANGGLVAGSYKRNELVRIRHDSDGGPKPLKNLNGQICQICGD
                     TVGLTASGDVFVACNECAFPVCRPCYEYERKDGNQSCPQCKSRYKRHKGSPRVDGDDD
                     EDEVDDLENEFNYAQGTSAARQQWQGEDPDLSSSSRHESRHPIPLLTNGQPMSGEIPC
                     ASIDSQSVRTTSGPLGPSDKHVHSLPYVDPRQPVPVRIVDPSKDLNTYGLGNVDWKER
                     VEGWNLNKRKNMTQMPNKYHEGKNDIEGTGSNGEELQMADDARQPMSRVVPISSSHLT
                     PYRVVIILRLIILGFFLQYRVTHPVKDAYPLWLTSVICEIWFALSWLLDQFPKWSPIN
                     RETYLDRLALRHDREGEPSQLAPVDVFVSTVDPLKEPPLITANTVLSILAVDYPVDKV
                     SCYVSDDGSAMLTFEALSETAEFARKWVPFCKKHNIEPRAPEFYFAQKIDYLKDKIQP
                     SFVKERRAMKREYEEFKVRINALVAKAQKMPEEGWAMQDGTAWPGNNLRDHPGMIQVF
                     LGHSGGLDTDGNELPRLVYVSREKRPGFQHHKKAGAMNALIRVSAVLTNGAYLLNVDC
                     DHYFNNSKALKEAMCFMMDPAYGKKTCYVQFPQRFDGIDLHDRYANRNIVFFDINLKG
                     LDGIQGPVYVGTGCCFNRQALYGYDPVLTEEDLEPNIIVKSCCGSRKKGKGGNKKYID
                     KKGAMKRTESTVPIFNMEDVEEGVEGYDDERSLLMSQKSLEKRFGQSPVFISATFMEQ
                     GGLPPSTNPATLSKEAIHVISCGYEDKTEWGKEIGWIYGSVTEDILTGFKMHARGWIS
                     IYCMPPRPAFKGSAPINLSDRLNQVLRWALGSIEILLSRHCPIWYGYNGKLRLLERLA
                     YINTIVYPLTSIPLIAYCILPAFCLLTNKFIIPEISNFASMWFILLFVSIFTTGILEL
                     RWSGVSIEDWWRNEQFWVIGGTSAHLFAVFQGLLKVLAGIDTNFTVTSKAGDEDGDFA
                     ELYVFKWTSLLIPPTTVLIVNIIGIVAGVSYAINSGYQSWGPLFGKLFFAIWVIAHLY
                     PFLKGLLGRQNRTPTIVIVWSILLASILSLLWVRIDPFTSATTASTANGQCGINC"
     3'UTR           3368..3712
                     /gene="CesA5"
ORIGIN
        1 gcgaagaaga actcggtctc tcctctctcc tctcctctct cctccgccgc cagatcctct
       61 cgcttccgcc ttcgatctcg gggagaagga aggaaggaag aggacgacga tggaggccaa
      121 tggcggcctg gtcgccggat cttacaagag gaacgagctg gtccggattc gccacgactc
      181 ggacggcgga cccaaacccc tgaagaattt gaatggccag atttgtcaga tatgtggcga
      241 taccgttgga cttacggcca gcggcgatgt ttttgttgct tgcaatgagt gtgcattccc
      301 tgtgtgccgt ccctgttatg agtacgagag gaaagatggt aaccaatcat gtcctcagtg
      361 caagtctcga tataagaggc acaaaggtag tcctcgagtt gacggagatg atgatgagga
      421 tgaggttgat gacctggaga atgagttcaa ttatgcccag ggaaccagtg ctgcaaggca
```

Appendix A

```
 481 acagtggcag ggagaagatc cagatctttc ttcttcttct agacatgaat ctcgacatcc
 541 aatccctctt ctaaccaatg ggcagccgat gtctggtgaa atcccttgtg ctagtattga
 601 cagccaatct gtgaggacta catctggacc tctgggtcct tctgataaac atgtgcactc
 661 gcttccctat gttgatccca gacagccagt tcctgtgcgg attgtggatc catcaaagga
 721 tttgaatact tatggcctcg gaaatgttga ctggaaggaa agggttgaag gatggaactt
 781 aaacaagaga aaaaacatga cgcagatgcc aaacaaatat catgaaggga agaacgacat
 841 agagggcact ggctctaatg gagaagaact tcaaatggct gatgatgcac gtcaacctat
 901 gagtcgtgtg gtgcctatat cgtcgtctca cctcactccg taccgtgttg taattatcct
 961 tcggttgatt attttggggt ttttcctaca gtatcgtgta acacacccag tgaaagatgc
1021 ttatcctttg tggcttacat cggttatttg tgagatttgg tttgctttat cctggctttt
1081 ggatcagttc ccaaaatggt cgcccatcaa tcgggaaaca taccttgata ggcttgcttt
1141 gagacatgat cgtgaagggg agccttcaca gttggctcct gttgatgtct ttgtgagtac
1201 tgtggatcct ttgaaagaac ctcctctcat aacagcgaac actgtcttat ccatacttgc
1261 cgtggattac cccgtcgaca aagtgtcgtg ttatgtctct gatgatgggt ctgcgatgtt
1321 gacttttgaa gccctctctg aaactgctga gtttgcaagg aagtgggtgc ccttctgcaa
1381 aaagcacaac attgagccac gagctcctga attctacttt gcccagaaaa tagattactt
1441 gaaggacaag atacagcctt cttttgtaaa agaacgcagg gcaatgaaga gagaatatga
1501 ggagttcaag gtgcgaatca atgctttggt tgcaaaggca caaaagatgc cagaggaagg
1561 gtgggcaatg caggatggca ctgcctggcc tggaaataac ctcagggatc accctggaat
1621 gatacaggtt ttcctgggcc acagtggggg acttgatact gatggaaatg agctacctcg
1681 acttgtttat gtttctcgtg aaaagcgacc tggtttccaa catcacaaga aagctggagc
1741 catgaatgct ttgatccggg tctcggctgt cctaaccaat ggagcgtatc ttttgaatgt
1801 tgactgtgat cattacttta ataatagtaa agcattgaaa gaagcaatgt gtttcatgat
1861 ggatcccgct tatggaaaga agacgtgcta tgtgcagttc ccacaacgtt ttgatgggat
1921 tgacttgcat gatcgatatg ctaaccgcaa catcgtcttc tttgatatta acttgaaagg
1981 gcttgacggc atccaaggtc ctgtctatgt cggaactgga tgttgtttca acaggcaagc
2041 cctttatgga tatgaccctg tattgactga ggaagatctg gaaccaaata ttattgtaaa
2101 gagttgttgt ggttcaagaa agaaggggaa gggtggcaat aagaagtaca ttgacaagaa
2161 aggagcaatg aaaagaactg aatccactgt tccaattttc aatatggaag atgttgagga
2221 gggggttgaa ggatatgatg atgagaggtc gctcctgatg tctcagaaaa gcctagagaa
2281 aagattcggt cagtctcctg ttttcatttc ggctactttc atggaacagg gtggcctacc
2341 accatctacc aatcctgcaa cactttcgaa agaagccatc catgttatta gctgtggcta
2401 tgaggacaag actgaatggg gcaaagagat cggatggata tatggttctg ttacagaaga
2461 tattttgact ggatttaaga tgcatgctcg gggctggatt tcaatctatt gtatgcctcc
2521 tcgcccagca ttcaaggggt ctgctcccat caatctttct gatcgtttaa accaggttct
2581 caggtgggcc ttggggtcaa ttgagatttt gctaagcagg cattgtccta tatggtatgg
2641 atataatggg aaattgaggc tcttggagag attggcgtac atcaatacaa ttgtgtatcc
2701 cctcacctca ataccgctga ttgcgtattg tattctccct gcattttgtc ttctcactaa
2761 caaatttatt atacctgaga taagcaactt tgctagcatg tggttcattc tccttttcgt
2821 ctcaattttc acgactggta ttcttgagct ccgttggagt ggtgtcagca ttgaggattg
2881 gtggaggaat gagcagttct gggtcattgg tggtacatcc gctcaccttt tcgctgtgtt
2941 ccaagggctt cttaaagtgc ttgctggcat tgataccaac ttcacagtca cttcgaaagc
3001 aggcgatgag gatggagact ttgctgagct ttatgtgttc aaatggacat cgcttctcat
3061 tcctccgacc acagttctta tcgtgaacat cattggcatt gtagcaggtg tgtcttatgc
3121 gattaacagc gggtatcagt catggggtcc actgtttggc aagttgttct tcgccatatg
3181 ggttattgct caccttttacc cattcctgaa gggtctgttg ggtcggcaga atcgtactcc
3241 tacgattgtt attgtgtggt ctatccttct tgcttccata ctctcattgc tctgggtacg
3301 gattgaccct ttcacctcgg ccaccacagc atctactgca aatggtcaat gcggcatcaa
3361 ctgctaattg tatcatgctc gcctgtgggg ggagccatca tcttcaagtg cttattgttc
3421 ttttatgtat ataacaagat tctcattgcc acccaaagtc acttttcatc attgaggcaa
3481 tactttcgtc cgggaagaag ctgggtggaa aacagaaaaa gcagagaatg tcatttttg
3541 tcatgtcctt tttacagttg cccgtgttgg atcaaaacag gcatctcgaa cctttgattc
3601 gtctgctgtt atttattgtg cagtattttg ttgtttgatc acaagagtgt agcactttcg
3661 aacttaattc agcacttaaa ggttgaatac agaacaagtt cattggtcaa gc
```

EgCesA6

```
LOCUS       DQ014510               3782 bp    mRNA    linear   PLN 06-JUN-2005
DEFINITION  Eucalyptus grandis cellulose synthase 6 (CesA6) mRNA, complete cds.
ACCESSION   DQ014510
VERSION     DQ014510
KEYWORDS    .
SOURCE      Eucalyptus grandis
  ORGANISM  Eucalyptus grandis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; Myrtales; Myrtaceae; Eucalyptus.
REFERENCE   1  (bases 1 to 3782)
  AUTHORS   Ranik,M. and Myburg,A.A.
  TITLE     Six new cellulose synthase genes from Eucalyptus are linked with
            primary and secondary cell wall biosynthesis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3782)
  AUTHORS   Ranik,M. and Myburg,A.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2005) Department of Genetics, University of
            Pretoria, Pretoria 0002, South Africa
FEATURES             Location/Qualifiers
     source          1..3782
                     /organism="Eucalyptus grandis"
                     /mol_type="mRNA"
                     /db_xref="taxon:71139"
     gene            1..3782
                     /gene="CesA6"
                     /note="EgCesA6"
     5'UTR           1..47
                     /gene="CesA6"
     CDS             48..3341
                     /gene="CesA6"
                     /codon_start=1
                     /product="cellulose synthase 6"
                     /protein_id="AAY60848"
                     /translation="MEVSSGLVAGSHNRNELVVIRRENELGQKPLQKLSGQICQICGD
                     DVGLTVDGELFVACNECAFPICRTCYEYERREGSQICPQCKTRFKRLRGCARVDGDEE
                     EDGVDDLENEFNFDGRHRQEMDRQGYGAEAMLHGHMSYGRGSDLDLPHVHPLPQVPLL
                     ANGQMVDDVPPEHHALVPAYMGAGGGGGGGGKRIHPLPFTDSGLPVQPRSMDPSKDLA
                     AYGYGSVAWKERMESWKQKQEKLQTMKNEKGGKEWDDDGDNPDLPLMDEARQPLSRRL
                     PISSSQINPYRMIIVIRLVVLGFFFHYRVVHPVNDAYALWLISVICEIWFGLSWILDQ
                     FPKWLPIDRETYLDRLSLRYEKEGQPSQLAPVDIFVSTVDPLKEPPLVTANTVLSILA
                     VDYPVDKVSCYVSDDGAAMLTFEALSETSEFARKWAPFCKKFNIEPRAPEFYFAQKID
                     YLKDKVEASFVKERRAMKREYEEFKVRINALVAKAQKVPEEGWTMQDGTPWPGNNVRD
                     HPGMIQVFLGQSGGHDSDGNELPRLVYVSREKRPGYNHHKKAGAMNALVRVSAVLTNA
                     PYLLNLDCDHYFNNSKAIREAMCFMVDPLIGKRVCYVQFPQRFDGIDRHDRYANRNTV
                     FFDINMKGLDGIQGPIYVGTGCVFRRLALYGYDAPKAKKPPTRTCNCLPKWCCCGCCC
                     SGKKKKKKTTKPKTELKKRFFKKKDAGTPPPLEGIEEGIEVIESENPTPQHKLEKKFG
                     QSSVFVASTLLEDGGTLKGTSPASLLKEAIHVISCGYEDKTEWGKEVGWIYGSVTEDI
                     LTGFKMHCHGWRSIYCIPARPAFKGSAPINLSDRLHQVLRWALGSIEIFLSRHCPLWY
                     GYGGGLKWLERLSYINATVYPWTSIPLLAYCTLPAVCLLTGKFITPELSNVASLWFLS
                     LFICIFATSILEMRWSGVGIEEWWRNEQFWVIGGVSAHLFAVFQGLLKVLAGVDTNFT
                     VTSKGGDDKEFSELYAFKWTTLLIPPTTLLIINLIGVVAGVSNAINNGHESWGPLFGK
                     LFFAFWVIVHLYPFLKGLLGRQNRTPTIIIVWSILLASIFSLLWVRIDPFLAKSDGPL
                     LEECGLDCN"
     3'UTR           3342..3782
                     /gene="CesA6"
ORIGIN
        1 cgctaagaag ggctgaaaat ccagaacggt agtaagagca aagagaaatg gaggtgagtt
       61 ctggtttagt agcgggctct cacaacagga acgagctggt tgtcatccgc cgcgagaatg
      121 aactcggaca aaagccgttg cagaagttga gcgggcaaat ttgccagatt tgcggcgacg
      181 acgttggatt gaccgtggac ggcgagctat tcgtcgcctg caatgagtgt gcgttcccca
      241 tttgcaggac ttgctatgag tacgaacggc gcgagggaag ccaaatttgt cctcagtgca
      301 aaaccagatt caagcgctta agggggtgtg caagagtgga tggagatgag gaagaggatg
      361 gtgtggatga cttggagaac gagttcaact ttgatgggag gcataggcaa gagatggatc
```

Appendix A

```
 421 gccagggata tggtgcagag gcaatgcttc atggccatat gagctatggc cgtggctcgg
 481 atttggatct gcctcacgtt catccactgc cccaagtccc actcctcgcc aatggtcaaa
 541 tggttgatga tgttcctccg gagcaccatg ctttggtgcc agcctacatg ggagctggag
 601 gcggcggtgg cggaggtggc aaaaggattc acccacttcc tttcactgat tctggtcttc
 661 cagtgcaacc tcgatccatg gatccttcaa aggacttggc tgcttatgga tatggaagcg
 721 ttgcttggaa agagaggatg gagagttgga aacaaaagca agagaaacta cagacgatga
 781 agaacgagaa aggtggcaag gaatgggacg atgatgggga caatccagat ctaccactaa
 841 tggatgaggc gagacagccg ctgtcaagaa ggttgcctat atcctccagc caaatcaatc
 901 cctacagaat gatcattgtc atccggcttg tagtgcttgg attcttttttc cattacagag
 961 tagtgcatcc tgtgaacgat gcatacgcat tgtggctcat atctgttatc tgtgagatct
1021 ggtttggtct ctcgtggatt ctggatcaat tcccgaaatg gctccctatt gaccgggaga
1081 catacctcga cagactgtcc ctgaggtatg agaaggaagg ccaaccttct caacttgctc
1141 cagtagatat atttgtgagt actgttgacc cgctgaagga acccccccctg gtgacggcaa
1201 atactgtctt atcgattctc gctgtggact accctgttga taaggtgtcc tgttatgttt
1261 ctgatgacgg tgctgctatg ctgacattcg aagcactctc agaaacatca gaatttgcaa
1321 gaaaatgggc ccctttctgt aagaaattca acattgagcc acgggcacct gagttttatt
1381 ttgctcagaa gatagattat ctcaaagata aggttgaggc atcttttgtg aaggagcgaa
1441 gagcaatgaa gagagagtat gaggagttta aggttcgtat aaatgctttg gtggccaaag
1501 cccagaaggt tcctgaagaa gggtggacga tgcaggacgg aacgccatgg cctggaaata
1561 acgtccgtga tcatcctggc atgattcagg tttttcttgg ccaaagtgga ggacatgact
1621 ctgacgggaa tgaattacct cgccttgtct atgtttccag agagaagagg cctggctata
1681 atcaccacaa aaaggcagga gccatgaatg ctttggtcag ggtctccgct gtgcttacaa
1741 atgcgcctta tcttttgaat ctggattgtg atcactactt caataatagt aaggctatta
1801 gggaggcaat gtgcttcatg gtggacccac ttattggaaa aagagtgtgc tacgtccagt
1861 tccctcaaag attcgacggt atagataggc acgatcgata tgccaatcgg aacaccgtgt
1921 ttttcgatat taacatgaaa ggtctggatg gcattcaagg tcctatttat gttgggactg
1981 gttgtgtatt cagaaggctg gcactctatg gttatgatgc tccaaaagca aagaagccac
2041 cgaccagaac ttgcaactgc ttgccaaagt ggtgctgttg tgggtgctgt tgctctggga
2101 agaagaagaa gaagaagact accaagccta agaccgaact gaagaagaga ttctttaaaa
2161 aaaaagatgc aggtactcca ccacctttgg aaggtatcga agagggcatt gaagtaattg
2221 aatctgaaaa tcccacgcct cagcataagc ttgaaaagaa gtttgggcaa tcctcagtct
2281 tcgtcgcatc gacattgcta gaggatggag gaacactgaa aggcacgagc cctgcatctt
2341 tgctaaaaga agctatccat gtcatcagct gcggctacga agataaaact gagtggggaa
2401 aagaggtcgg atggatatat ggttcagtta cagaagatat attgacaggc tttaaaatgc
2461 actgccatgg atggcgttcc atctactgca tccctgcaag acccgctttt aaaggatctg
2521 ctcccattaa tctttccgac cgtctccacc aggtccttcg ttgggctctt ggttccattg
2581 agatcttctt aagcaggcat tgtccccttt ggtatggtta tggaggcggt ttgaaatggt
2641 tagagcgatt atcttacata aacgccacag tatacccctg gacatcaatc cctctgttgg
2701 catactgtac tttgcccgct gtgtgcttgc ttactgggaa attcatcacg ccagagctga
2761 gcaatgtggc tagtctgtgg tttctgtcac ttttcatttg cattttcgcg acaagtatcc
2821 tggaaatgcg atggagtggg gtcggaatag aggaatggtg gaggaacgaa cagttctggg
2881 tcatcggagg agtttcagca catctctttg cagtttttca ggggcttttg aaggttctgg
2941 ctggtgttga cacgaacttc actgtgacat caaagggtgg ggatgacaag gagttctctg
3001 aactttatgc gttcaaatgg accaccttgc tcattccgcc aaccacattg cttataataa
3061 acctgattgg tgttgtagct ggcgtatcaa acgcaataaa caacggtcac gagtcctggg
3121 gtcctctctt cgggaagcta ttctttgcct tctgggtgat tgtccacttg tatcctttcc
3181 taaagggtct gcttggacgg caaaacagga ctcctacaat catcatcgtg tggtcaatct
3241 tgctcgcttc catcttctcc cttctgtggg tacgtatcga tccattcttg gccaagtcgg
3301 acggaccgct tttggaagaa tgcgggttgg actgtaatta gcagagtgtc tgattcctgg
3361 aacatttggt cagtaggatt tgtttatggg ttatttatcg taaggataca gaaagaatac
3421 atttggtgtg gtgccgtctg caatgcaaat aaagggcatg tttggacgtc actgtttccg
3481 caactccgcc tctccatttg accgcaacac tctgaagggc tacaaaaacc aactttgctg
3541 ctcttttgca acagctggga gattgtcgga aaggactgtg acactgcttc ttgattcttt
3601 atttggttca taatttttt ctttttctagg atgaggaggt ggcgtgaacc acagtataat
3661 gtgtattatc atctcataat ctctctctct ccctccctct ctctcccct gtgtaagtgt
3721 aactcagccg aaattgtgag catgtattat atacggggca tcgtccctca gttgatgctg
3781 ct
```